Reshaping the Economy? Local Reallocation Effects of Place-Based Policies

Sarah Fritz[1] and Catherine van der List[2]

[1]Halle Institute for Economic Research (IWH)
[2]University of Essex

July, 2025

Click here for the latest version

Abstract

We study the effects of place-based policies on aggregate productivity using administrative data on projects co-financed by the EU in Italy linked to balance sheet data. We exploit quasi-experimental variation in funding for a large place-based policy stemming from measurement error in regional GDP estimates. Results show that the policy likely decreases productivity. Decompositions reveal that aggregate declines are driven by reallocation of labor to low-productivity firms. Mechanism analysis using firm-level event studies reveals that negative reallocation effects are caused by high-productivity firms taking up the funds and subsequently becoming more liquidity constrained, leading to slowdowns in employment growth.

Keywords: Place-Based Policy, Productivity, EU Cohesion Policy
JEL-classification: R11, R58, J23, Z18

Sarah Fritz: sarah.fritz@iwh-halle.de
Catherine van der List: catherine.vanderlist@essex.ac.uk

We thank Michel van der List for his assistance with querying REST APIs using Python. We thank Richard Bräuer, Ada Checola, Anand Chopra, Riccardo Crescenzi, Mirja Hälbig, Max Löffler, Elena Maffi, Javier Miranda, Steffen Müller, Brunella Reverberi, Susann Rosipal, Raffaele Saggio, Michel Serafinelli, Sebastian Siegloch, Martina Uccioli, Simon Wiederhold, seminar participants at Bates College, the Halle Institute for Economic Research, the London School of Economics, the UK Competition and Markets Authority, the University of Cologne, the University of Essex, and numerous conference and workshop participants for their helpful comments. Sarah Fritz would like to thank her advisors Christoph Wunder and Mirko Titze for their guidance, support, and feedback.

1 Introduction

Influencing the shape of the economy is back in vogue. From industrial policy to tariffs, governments across the world are passing large policy packages designed to reshape the structure of their economies to pursue public prosperity. Place-based policies are a commonly used form of such interventions, with the goal of revitalizing regions that have been economically lagging. The United States spends approximately $50 billion annually on place-based policies (Bartik, 2020). The European Union (EU) spent roughly €530 billion, on EU Cohesion policy (henceforth cohesion policy) during the 2014-2020 funding period - one-third of the EU budget.[1] The policies are often justified on redistributive and equality grounds given the profound negative social and political effects of pockets of persistent disadvantage (Autor et al., 2024; Heddesheimer et al., 2024), and that gaps have stopped closing or are growing over time (Austin et al., 2018; von Ehrlich and Overman, 2020). However, skeptics of these policies worry that the policies may have unintended negative consequences due to subsidizing unproductive firms or shifting production from more- to less-productive locales (Glaeser and Gottlieb, 2008; Neumark and Simpson, 2015).

To better understand the impact of these policies, we study the effects of cohesion policy on aggregate productivity changes over seven-year funding periods (2007-2013 and 2014-2020) in Italian local labor markets. We decompose these productivity changes in order to understand if competing effects are at play. On one hand, place-based policies may foster productivity gains of all firms in a local labor market by improving the provision of public goods, investing in training programs for the local labor force, or fostering agglomeration economies. Programs may also promote labor reallocation by helping productive firms in expanding industries grow and/or enter relative to unproductive firms. On the other hand, policies could also have the effect of slowing structural change by, for example, subsidies contributing to inefficient allocation of resources. Given the broad scope of cohesion policy – encompassing investments across a wide variety of sectors and objectives – understanding which potential effect dominates is key to understanding if place-based policies can create long-term gross domestic product (GDP) growth – the ultimate goal of cohesion policy.

To overcome the endogeneity challenges presented by relating cohesion policy funding, which by design is targeted towards less developed geographic areas, and aggregate productivity growth we develop a novel identification strategy building on Borusyak and Hull (2023). Borusyak and Hull (2023) demonstrate a methodology to identify the effects of a treatment involving endogenous exposure to exogenous shocks. The methodology hinges on

[1] The €530 billion includes national government co-financing. The EU financing was approximately €400 billion. See `https://cohesiondata.ec.europa.eu`.

the ability to construct counterfactual treatments – in our case counterfactual funding levels – using exogenous variation. To construct such counterfactuals, we leverage measurement error in regional GDP per capita, the key determinant for funding levels, and the known funding formulas used by the European Commission – simulating funding amounts which would have occurred under different empirical realizations of measurement error. Identification then comes from comparing commuting zones that *expected* to receive similar levels of funding but experienced different *realizations* of funding. Our empirical findings relate aggregate changes in commuting-zone-level labor productivity over the seven-year funding periods – decomposed into uniform changes across firms, reallocation of labor to more productive firms, as well as firm entry and exit – to deviations in per-capita funding levels stemming from exogenous measurement error.

We begin by using newly-digitized historical regional GDP estimates to construct empirical distributions of GDP measurement error. Though measurement error is typically under 10% and centered around zero, error can be up to +/-30%, leading to substantial changes in funding envelopes in-practice. This is amplified by the fact that the funding formulas used by the EU are relative, and funding levels are sticky over the seven-year funding period. So, measurement error affects not only the region's own shock, but its relative position in the distribution – which affects every EU region for the entirety of the funding period. The exposure to these shocks varies *within* region in different commuting zones. The Borusyak and Hull (2023) methodology is specifically designed to deliver causal identification in such cases, in our application demonstrating the causal effect of additional funding on commuting-zone level labor productivity.

We find that cohesion policy has a negative impact on aggregate productivity changes in local labor markets in Italy – with a 1 percent increase in cohesion policy funding leading to a 0.0636 percent decrease in aggregate productivity growth over the 7-year funding period due to reduced allocative efficiency. While we do not find any effect of productivity growth within firms on average, we show that there is some evidence of positive effects on aggregate productivity stemming from establishment entry. Meaning, our evidence suggests that multiple effects are at play – and the negative labor reallocation effects dominate the positive entry effects over the periods we study. Analyzing the effects for specific industries, we find that the negative labor reallocation effects are driven by the manufacturing sector, while the positive entry effects are driven by the construction and services sectors.

We subsequently delve into the mechanisms underlying the negative labor reallocation. Negative labor reallocation could stem from two sources: less productive firms growing more, or more productive firms growing less. We provide two sets of complementary evidence that the mechanism at play is the latter. First, we repeat the aggregate analysis relating changes in unemployment rates of commuting zones over the seven year funding periods to the deviations in funding from the measurement error, and we find that unemployment rates grew faster in local labor markets that received more funds conditional on the expected funding level. If the mechanism was less productive firms growing more as a result of cohesion policy, we would expect the opposite. Second, we utilize rich administrative data on funding recipients to undertake firm-level event studies around the start of projects undertaken by firms. We find that while beneficiary firms tend to be more productive, such projects have little impact on firm-level productivity growth, but strong negative employment effects. Further examination reveals that beneficiary firms are more liquidity constrained after undertaking the project, likely leading to their slowdown in employment growth.

We contribute to the literature on place-based policies in several ways. Our findings relate most directly to a strand of literature studying the aggregate region-level effects of cohesion policy. These studies have typically found positive GDP per capita effects stemming from cohesion policy (Becker et al., 2010; Mohl and Hagen, 2010; Pellegrini et al., 2013). Subsequent studies, some using the same identification strategies, have drawn a more nuanced picture, showing that despite the EU's objective to induce long-term growth, positive growth effects are not long-lived upon cessation of the funding (Barone et al., 2016; Di Cataldo, 2017; Becker et al., 2018). Our study helps reconcile these findings. Our results using our rich firm-level data show that the gains in GDP per capita from other studies are not underpinned by productivity-enhancing developments in (recipient) firms or reallocation to more-productive firms. This provides a plausible explanation for why income growth effects tend to vanish after the funding stops: without improvements in productivity – the key driver of long-run economic growth (Jones, 2016) – such income gains are unlikely to persist.

Moreover, our study is naturally related to a large literature on declining productivity growth, which has been documented for many developed countries (Adler et al., 2017; Fernald et al., 2025). Decker et al. (2017) show that declining business dynamism, observed both in the US (De Ridder, 2024) and Europe (Biondi et al., 2025), is a driving force of the productivity slowdown through reduced allocative efficiency. In particular, the productivity slowdown in Europe – of which Italy is a particularly severe example – has been driven by increased misallocation to less-productive firms (Adler et al., 2017; Fernald et al., 2025). Our results are consistent with this literature, and further demonstrate in the reduced-form a potential contributor to the productivity slowdown in the Italian case: cohesion policy.

To our knowledge, our paper is the first to identify place-based policies as a source of reduced allocative efficiency in the reduced form, though it has been emphasized as a potential drawback to place-based policies in theory (Glaeser and Gottlieb, 2008; Neumark and Simpson, 2015). Several papers have investigated the firm-level productivity impacts of country-specific place-based policies, and – consistent with our firm-level results – do not find evidence for any effects (Cerqua and Pellegrini, 2014; Criscuolo et al., 2019). However, as shown in our paper, firm-level impacts do not tell the entire story. In our results, neutral firm-level productivity impacts still lead to productivity slowdown in the aggregate through labor reallocation. While aggregate productivity impacts of cohesion policy in Italy have been investigated descriptively by Albanese et al. (2021) (whose results are consistent with our aggregate effects) to our knowledge our paper is the first to identify causal impacts of place-based policies on aggregate productivity growth, and also the first to decompose the effect into components to identify specific underlying mechanisms.

Our results also speak to the importance of careful design and targeting of place-based policies, recently emphasized in the literature (Austin et al., 2018; von Ehrlich and Overman, 2020). Cohesion policy is characterized by being very broad in scope and goals. This leads to many potentially competing effects, and different types of funding have been shown in the structural literature to have differential impacts (Blouri and von Ehrlich, 2020; Canova and Pappa, 2025). Our results quantify these competing effects directly in the reduced-form, demonstrating the negative labor reallocation effect dominates a competing positive firm-entry effect. We are also able to link these aggregate results to their underlying mechanism using our firm-level event studies.

Program design likely also explains why in our firm-level event studies we find evidence of a slowdown in employment growth, different from previous literature studying country-specific regional policies. Italy's national regional policy – Law 488/92 – has been shown to have positive employment effects (Cerqua and Pellegrini, 2014). Studies of place-based employment incentives in other countries such as Norway (Ku et al., 2020), the UK (Criscuolo et al., 2019), and Germany (Siegloch et al., 2025; Grunau et al., 2025) also find positive employment effects. These country-specific place-based policies were specifically designed to "create or safeguard" employment in structurally weak or less densely populated areas. In contrast, the main objective of cohesion policy is not to create employment. Employment indicators are only one of many factors taken into account when designing operational programs, and cohesion policy has a much broader impact on local economies.

Only a few studies have analyzed the impact of cohesion policy on employment in Italy. Our findings are in contrast with studies by Giua (2017) and Cerqua and Pellegrini (2022) who find positive employment effects. However, these studies differ from ours along various dimensions. Both studies utilize spatial regression discontinuity (RD) designs at the municipal level – estimating the effect of cohesion policy on aggregate municipal employment along borders between Italian regions where funding eligibility differed. The first key difference between these papers and ours is that we study firm-level impacts of cohesion policy of employment using event studies, while previous studies utilized aggregate employment data. The second key difference is that a spatial RD design estimates a Local Average Treatment Effect (LATE). Our strategies utilize different identification assumptions – with our aggregate analysis identifying an Average Treatment Effect (ATE) across Italy, while our firm-level event studies identify an Average Treatment Effect on the Treated (ATT). It may be the case that, for example, estimated LATEs along borders capture in-part displacement of economic activity within narrow geographic areas or concurrent unobserved region-level shocks.

Place-based and industrial policies have been shown to promote long-term growth in some cases by creating self-sustaining agglomeration economies in countries such as the United States (Kline and Moretti, 2013; Garin and Rothbaum, 2025), Finland (Mitrunen, 2025), and Italy (Incoronato and Lattanzio, 2024). A goal of place-neutral industrial policies more generally is to create self-sustaining industries by subsidizing the most productive (Aghion et al., 2015; Juhász et al., 2023). However, in some cases, long-term growth has been shown to be hampered by regional policies that subsidize particular industries by promoting over-specialization (Heblich et al., 2022) or more generally by policies subsidizing unproductive industries or firms (Glaeser and Gottlieb, 2008). Our results complement this literature by suggesting that in the absence of careful policy design, even if more-productive firms take up the funds they may utilize subsidies in a way that slows their longer-term growth and hurts workers in the invested areas.

The remainder of this paper is organized as follows. In Section 2 we describe the relevant institutional aspects of cohesion policy, in Section 3 we describe our data sources, Section 4 describes the aggregate productivity decomposition, Section 5 describes our empirical strategy for the aggregate productivity change analysis, Section 6 discusses our aggregate results, Section 7 explores the mechanisms driving the main results using firm-level event studies, and Section 8 concludes.

2 Institutional Background

The goal, from the perspective of the EU, of cohesion policy is to minimize gaps in economic outcomes across space within the member states, particularly in GDP per capita, which is the primary eligibility criterion. Cohesion policy has become the largest budget item of the EU since its kick off in 1989. The financial scope of cohesion policy (roughly €530 billion Euros over the 2014-2020 funding period) makes it one of the largest place-based policies in the world. The EU funds represent *additional* resources available to the regions and cannot substitute national public funds (see Article 15 of Council of the European Union (2006) and Article 95 of Council of the European Union (2013b)).

In terms of types of investment and beneficiaries the policy is very wide-ranging. For each funding period, the European Commission defines thematic objectives such as *Promoting sustainable transport and improving network infrastructures* or *Investing in education, training and lifelong learning*[2] to achieve their overarching goal. Expenses on cohesion policy and its thematic objectives currently mainly stem from two so-called Structural Funds, the European Regional Development Fund (ERDF) and the European Social Fund (ESF). The ERDF co-finances R&D or capital investment subsidies for firms as well as infrastructure projects in the broad sense, i.e. transport, social, environmental or telecommunication infrastructure projects such as roads, hospitals and schools, hardware and software for computer service centers or water purification plants. In contrast, the ESF invests in the human capital endowments in the targeted regions through for instance scholarships, female laborforce participation or re-training programs. As managing authorities of the funds, regions design programs that would fit those goals, subject to EU minimums and targets surrounding particular funding themes.[3] Actors with a potential project apply to the regional fund for project funding through a variety of processes, and receive the money if the project is determined to fit the goals of cohesion policy.[4] Beneficiaries from cohesion policy can therefore range from individuals, to companies and NGOs, to municipalities and regions.

Cohesion policy is designed as a place-based policy, with funding heavily skewed toward low-income regions. For instance, during the 2014-2020 funding period, Italian regions were allocated approximately €32.5 billion in cohesion funds, out of which €17.4 billion (more than half) was allocated to the five poorest regions: Calabria, Campania, Puglia, Sicily, and Basilicata.[5] These regions had a GDP per capita less than 75% of the EU average were therefore classified as "less developed". In that funding period, regions with a GDP per capita measured in purchasing power standards (PPS) between 75 and 90% of the EU average were classified as "transition" regions, while those above the 90% threshold were

[2]See `https://ec.europa.eu/regional_policy/policy/how/priorities/2014-2020_en`.

[3]For example, the EU mandated at least 20% of the more developed regions funding from the ERDF in the 2014-2020 funding period be dedicated to the thematic objective *Supporting the shift towards a low-carbon economy* (Council of the European Union, 2013a).

[4]See Article 125(3) of Council of the European Union (2013b) for details on the requirements of processes. In short, the rules require that the project is determined to fit the thematic priorities of cohesion policy, and the process to be non-discriminatory and transparent, *not* necessarily competitive or based on quality. The most competitive allocation method used by Italian regions is a "bando di gara" – essentially a public call analogous to a procurement auction. Projects selected by such processes are responsible for only approximately 30% of payments in our data, and it is impossible to know how competitive this process truly is since we do not observe denied applications and therefore cannot calculate award rates.

[5]These numbers are for region-specific Programma Operativo Regionale (PORs) for the ERDF and ESF, including national co-financing.

classified as "more developed". Crossing a threshold significantly reduced funding: regions above the 75% cut-off received at most 40% of what they would have received as a less developed region and regions above the 90% threshold received even lower funding levels than transition regions. In the 2007-2013 funding period, there were two types of transitional funding schemes: "Phasing in" and "phasing out" funding.[6] For the remainder of this paper, "phasing out" and "phasing in" regions will also be referred to as "transition" regions for simplicity.

Moreover, in both funding periods, there were also "intensive" margins in the funding formulas for all categories of regions – i.e. how far the region was from the various thresholds matters. For example in the 2014-2020 funding period, the funding formula for the less developed regions included the following clause: "determination of an absolute amount (in EUR) obtained by multiplying the population of the region concerned by the difference between that region's GDP per capita, measured in PPS, and the EU-27 average GDP per capita (in PPS)" (Council of the European Union, 2013b). See Appendix B.4 for further details on the funding formulas.

Funding eligibility is determined at the beginning of (currently) seven-year-long periods based on a three-year average of the GDP per capita of a NUTS2 territorial unit (hereafter referred to as regions) compared to the EU average calculated over the same years. The three-year average is based on data available to the European Commission two years prior to the beginning of the funding period (i.e. as of 2012 for the 2014-2020 funding period), and using the latest available data, ending three years prior to that (i.e. GDP per capita data for 2007-2009 for the 2014-2020 funding period). After the European Commission determines the recommended financial envelope of the regions, the national governments receive separate funding "pots" for each eligibility category. Then, the regions within a particular category negotiate an exact split of the funds with the state. After the split is agreed upon, the regional governments administer their funds over the seven-year funding period.

Within regions, certain local labor markets (henceforth commuting zones) attract higher shares of the funding available to their region than others.[7] There are two main reasons for such variation in funding levels across commuting zones. First, the thematic objectives set by the EU, while broad, inevitably favor commuting zones with a particular industry composition. For example, the EU allocated funds to support projects in the tourism sector,

[6] "Phasing out" funding refers to regions which would have fallen under the 75% threshold based on the EU-15 countries, but above the 75% threshold of the EU-25 due to the EU enlargement in 2004 where mainly poorer countries joined the EU. "Phasing in" funding refers to regions which are above the 75% threshold of the EU-15 but were less developed regions in 2006, the last year of the previous funding period.

[7] Italy is composed of 20 regions and 610 commuting zones with an average of 13 municipalities. Most commuting zones in Italy are contained entirely within a single region. Commuting zones that span multiple regions are excluded from the aggregate analysis, but included in the firm-level event studies.

making it more likely that areas with a strong tourism industry would apply for and receive funding. Second, cohesion policy is a demand-driven program. Hence, it is also more likely that areas with better actors (e.g. firms and local governments) will attract more funding.

3 Data

In this section, we outline our main data sources and describe their key features. Specifically, we combine detailed administrative records on projects co-financed by the EU with firm-level balance-sheet data that allow us to measure productivity and track market entry and exit. The resulting dataset is aggregated to the commuting-zone–funding-period level for the aggregate analysis presented in Section 6, while we keep the unbalanced yearly firm panel for the firm-level results presented in Section 7. Detailed documentation of our sample construction for both project and firm data is provided in Appendix B.

3.1 Project Data

Our main dataset consists of projects financed by a single region's funding allocation under the ERDF or ESF. Data on projects co-financed by the EU comes from the OpenCoesione database. The data contain information on the universe of projects co-financed by any EU fund in Italy for the 2007-2013 and 2014-2020 funding periods. The data include information about the type and goals of the project, the municipality in which the project takes place, detailed information on project payments as well as their source, and dates at which various stages of the project began and ended. The data also have the codice fiscale – the Value Added Tax (VAT) tax identifier – of actors involved in the projects in a variety of roles. The availability of tax identifiers allows us to link the funding data to firm-level balance sheet data (see below), enabling us to study both selection into the program and the direct effects of funding on beneficiary firms (see Section 7).

Throughout this study, projects taking place across multiple commuting zones are assumed to have equal spending in each location. Moreover, we consolidate individual projects into single initiatives when it is evident that they are part of a broader program. For instance, scholarship programs often appear in the dataset as multiple entries, with each entry corresponding to an individual student application. To ensure comparability across initiatives, particularly with respect to their scale, we aggregate these entries. See Appendix B.1 for further details on this process.

Figure 1 maps the actual allocation of cohesion policy funding across Italian commuting zones during the 2007-2013 and the 2014-2020 funding periods, respectively. As expected, given the place-based nature of cohesion policy, less developed regions in the South of Italy receive the lion's share of fiscal transfers (including the national co-financing). Notably, there is also substantial variation in the distribution of funding within regions due to the factors discussed above in Section 2.

To (further) illustrate the defining characteristics of cohesion policy – its place-based nature, its broad thematic scope and the wide range of beneficiaries – Table 1 shows summary statistics of projects (taking place within specific commuting zones) by the regional eligibility classification. Panel A shows financial and implementation characteristics of the projects. In terms of per capita funding, commuting zones in less developed regions receive more than twice as much as those in more developed regions. While there are fewer projects per capita compared to more developed regions (four projects per 1000 inhabitants compared to eight in the 2007-2013 funding period), the typical project within a less developed region is more than four times as large on average, with €195,815 in commitments and €154,267 in payments during the 2007-2013 funding period against €37,897 in commitments and €35,394 in payments for more developed regions. Overall, the difference between funding commitments and actual payments is moderate, indicating that most of the planned projects are actually carried out.[8]

Panel B highlights the broad thematic scope of cohesion policy, showing how funding is distributed across the 11 distinct thematic objectives, both in terms of project counts and financial allocations. In both funding periods and across less developed and transition regions, no single objective dominates: roughly half of the thematic objectives each account for at least 10% of total payments, and none exceed 21%. This dispersion reflects the wide range of investment areas covered by cohesion policy. At the same time, the data also reveal the presence of very large projects in certain sectors. For example, in the 2007-2013 period, projects classified under the thematic objective *Transportation* made up less than 1% of all projects in less developed regions, yet they accounted for almost 17% of total payments -— illustrating how a small number of large-scale projects can absorb a substantial share of available funds.

Panel C shows the characteristics of project implementers.[9] Unsurprisingly, the most common form of implementer is a government actor, but approximately 30-50% of payments

[8]See Fritz, Incoronato, and van der List (2025) for detailed descriptions concerning cohesion policy project completion in Italy.

[9]The OpenCoesione database distinguishes between the project planner (*programmatore*) and three implementation roles: *attuatore*, *beneficiario*, and *realizzatore*. Most projects do not list a *realizzatore* as it's a role for actors holding procurement contracts. In Panel C we therefore only classify *attuatore* and *beneficiario*. The shares can exceed 100% because (i) a project may list multiple implementers, and (ii) the roles can be filled by different types of actors (e.g., a government body and a private firm). OpenCoesione does not indicate the degree of involvement or payment shares per actor – only the total project payments.

go to projects which involve private firms in implementation roles, depending on the funding period and regional classification.

3.2 Firm Data

To measure productivity, we use balance sheet data from the Bureau van Dijk's historical Orbis database. This dataset contains general information on companies operating in Italy, their legal form and industry classifications, as well as their financial statements. The general information typically includes the date and place of incorporation as well as their current and past status (i.e. active, dissolved, bankrupt etc.). Since our data includes the historical vintages, we observe both the time of exit and information about defunct firms.[10] We can also link the actors from the funding project data to their balance-sheet data via their tax identifiers.

We assess the quality of Orbis coverage for Italy in detail in Appendix B.3. We show, using three separate aggregated data sources, that at least 50% of all firms (of any incorporation form and sector) in Italy are observed in Orbis during our analysis period, and these firms account for approximately 60% of Italian employment between 2007 and 2017.

Our unbalanced panel covers roughly 1.2 million firms. We exclude the agricultural sector, the financial sector as well as non-business sectors.[11,12] Furthermore, we exclude sole proprietorships and branches as well as other firms that never report value-added. The vast majority of firms in our sample (90%) are private limited companies. With only four employees, the median active firm in 2014 is small. More than half of the firms in our sample entered the market after 2006, and about 15% of those have already exited the market. Among the other half that were founded before 2007, about one third have exited the market by the end of 2021.

As mentioned above, in our mechanism analysis we leverage the fact that we can match project beneficiaries from the funding data to our firm sample. However, we lose a number of firms due to Orbis coverage, missing information and sample restrictions. For instance, for the ERDF, we identified approximately 77.5 thousand beneficiaries (for both funding periods). Of these, around 60% can be matched to firms in our analysis dataset. Among

[10]Historical vintages are crucial to track firms with a *known* exit. Orbis does not always report exits reliably – some firms have an "unknown" or outdated status. See Appendix B.2.3 for details on identifying entry and exit.

[11]Specifically, non-business sectors are the NACE industries *Public administration and defense* (sector O), *Education* (sector P), *Human health and social work activities* (sector Q), *Activities of households as employers* (sector T) and *Activities of extraterritorial organizations and bodies* (sector U).

[12]The manufacturing sector also includes the mining, electricity, and water supply industries (NACE Codes B, C, D and E).

the 30.9 thousand unmatched beneficiaries, roughly 53% are not listed in Orbis at all. The remaining 47% either lack value-added data for any year or are excluded due to sample restrictions detailed above.

3.3 Additional Region- and Commuting-zone-level data

Data for counterfactual funding simulations Historical data on regional GDP and GDP per capita come from Eurostat press releases. We digitized yearly historical GDP estimates for the years 1998-2009.[13] Revised figures are from the ARDECO database, the Annual Regional Database of the European Commission's Directorate General for Regional and Urban Policy. Other data at the regional level needed to apply the funding formulas were obtained via Eurostat and the World Bank. These variables include Gross National Income (GNI) per capita, number of unemployed persons, unemployment rates, population, employed persons and employment rates, tertiary education rates, employees with low education, early leavers from education, and average population densities of provinces (NUTS3 territorial units within the regions).

Additional regression controls We also extracted commuting-zone-level statistics from the Italian censuses of 2001 and 2011 as well as detailed population information from the Italian National Institute of Statistics (Istituto Nazionale di Statistica, Istat).

4 Productivity Measure and Decomposition

Productivity gains are the major driver of long-term economic growth (Jones, 2016). Therefore, understanding if cohesion policy has a positive impact on aggregate productivity is key to our understanding of the policy's efficacy and longer-term effects. Before diving into aggregate productivity and its components, we briefly discuss what productivity is and how it can be measured.

[13] 1998 is the earliest year for which data is available for the majority of EU regions, and 2009 is the final relevant year of data for our purposes, being used to calculate 2014-2020 funding period eligibility.

4.1 Measuring productivity

In simple terms, productivity is how much output is obtained from a particular set of inputs.[14] Measures of productivity can either be single-factor (e.g., labor productivity) or multi-factor (e.g., total factor productivity, TFP). Because labor productivity, defined here as value-added per worker, can directly be measured from the Orbis data, we consider it our preferred measure for productivity – despite it having the notable limitation that it does not account for the use of other production factors. In contrast, TFP needs to be estimated – which requires additional assumptions and is more data hungry. In particular, it requires determining or estimating the appropriate weights of the various inputs in the production function to accurately aggregate them into the productivity measure. In a robustness check, we repeat our main specifications using TFP as the dependent variable, and find qualitatively similar results.[15,16]

While there is substantial research on productivity at the firm-level, the broader question addressed in this paper is how aggregate productivity growth in particular commuting zones is affected by place-based investments. Decompositions of aggregate productivity allow us to bridge the gap between firm-level dynamics and market-level outcomes to gain insights about the mechanisms at play.

4.2 Aggregate Productivity Decomposition

Aggregate productivity can be understood as a combination of two processes: improvements *within* firms in terms of productivity and reallocation of market share *between* firms. Productivity in a commuting zone increases either when firms themselves become more productive or when firm-level productivity remains constant but resources – such as labor and capital – shift from less productive firms to more productive ones.

Formally, we can express aggregate productivity, Φ, as:

$$\Phi_t = \sum_i s_{it}\phi_{it} \tag{1}$$

[14] In practice, most datasets such as the Orbis dataset we use provide information that is not quantity-specific. To measure output, we only observe revenue rather than physical quantities of units produced. If price differences reflect quality (and not market power, for instance), using revenue data to measure productivity is even desirable. Because we only observe revenue, we also have to abstract from the fact that certain firms produce multiple outputs.

[15] We use the estimator of Wooldridge (2009). At the firm level, the correlation between worker value added and TFP is .90, and we lose 9% of firms (5% of sample observations) due to missing financial information required for TFP estimation.

[16] To exclude outliers and since productivity is hard to measure, we trim the bottom and top 5 % of both productivity measures. See Appendix B.2.4 for more details.

where Φ_t represents the weighted average of firm-level productivity ϕ_{it} at time t. The weights, s_{it}, typically reflect firms' market shares, with $\sum_i s_{it} = 1$. In our analysis, we measure productivity as the logarithm of value-added per worker[17], using employment shares as weights[18], and calculate aggregate productivity at the level of commuting zones.[19]

Olley and Pakes (1996) formalize the idea that aggregate productivity reflects two key mechanisms: firm-level productivity and allocative efficiency. Specifically, they show that aggregate productivity can be decomposed into two components: the unweighted average of firm-level productivity and a term that is proportional to the covariance[20] between firms' employment shares and their productivity:

$$\begin{aligned} \Phi_t &= \bar{\phi}_t + \sum_i (s_{it} - \bar{s}_t)(\phi_{it} - \bar{\phi}_t) \\ &= \bar{\phi}_t + cov(s_{it}, \phi_{it}) \end{aligned} \tag{2}$$

Changes in productivity can then result from productivity increases stemming from most or all firms ($\Delta\bar{\phi}_t$), and from changes in the joint distribution of firm productivity and size ($\Delta cov(s_{it}, \phi_{it})$).[21] The novelty of the approach taken in Olley and Pakes (1996) is that it does not follow firms over time. Rather the decomposition is based on the comparison of the cross-sectional distributions of firm productivity taken at two different points in time. No need for panel data, two "snapshots" of the economy suffice.

Melitz and Polanec (2015) extend the Olley-Pakes decomposition to account for the contribution of entering and exiting firms to changes in aggregate productivity. Intuitively, aggregate productivity in period 1 corresponds to the sum of the aggregate productivity of surviving and exiting firms, weighted by their respective aggregate employment shares. In period 2, exiting firms will by definition have no aggregate employment share ($s_{X2} = 0$), but some firms will enter the market ($s_{E2} > 0$). Therefore, aggregate productivity in period 2 corresponds to the sum the aggregate productivity of surviving and entering firms, again weighted by their respective aggregate employment shares in that period.

[17] We apply value-added deflators at the NACE 2-digit sector level and express value-added in 2005 prices.

[18] As is standard in the literature, in the case of labor productivity market shares are employment shares, while in the case of TFP, nominal value-added shares are used (Melitz and Polanec, 2015).

[19] Cells with less than ten firms in the initial year are excluded from our analysis.

[20] In Equation 2 and throughout the paper, we refer to the covariance term, though strictly speaking, this is a slight abuse of notation, as we are not dividing by (n-1).

[21] Technically, Kehrig and Vincent (2021) show that $\Delta cov(s_{it}, \phi_{it})$ captures both labor reallocation as well as productivity changes at different positions in the firm size distribution. In other words, the covariance term would increase over time, both if more productive firms grow while less productive firms shed workers (keeping productivity constant), and if positive within-firm productivity changes are observed among the biggest firms only.

Formally Melitz and Polanec (2015) show that changes in aggregate productivity $\Delta\Phi$ between the two periods can be decomposed into four terms:

$$\Delta\Phi = \Delta\bar{\phi}_S + \Delta cov_S + s_{E2}(\Phi_{E2} - \Phi_{S2}) + s_{X1}(\Phi_{S1} - \Phi_{X1}) \quad (3)$$

The first two terms capture the change in aggregate productivity of surviving firms, following Olley and Pakes (1996).[22] Entering firms can contribute to positive changes in aggregate productivity if their aggregate productivity is higher than that of surviving firms in period 2. Overall aggregate productivity will also rise if aggregate productivity of exiting firms is inferior to that of surviving firms in period 1.

5 Empirical Strategy

The nature of place-based policies leads to obvious identification challenges. If we simply compared commuting zones in Italy that receive high levels of financial support to those that receive lower funding levels, we would likely underestimate the impact of cohesion policy on productivity growth since by design, structurally weak regions receive more support. Looking at changes in treatment intensity over time would not solve the problem, as areas that see a bigger increase are typically those that are on a downward trend. Using variation across commuting zones within regions would also be problematic since cohesion policy is a demand-driven program. Within a region, it might be economic agents in the areas with relatively better economic conditions or developments that systematically apply for more or bigger projects and thus attract more funds. In this case, we would likely overestimate the treatment effect.

In this section, we discuss the identification strategy that we developed to isolate the causal effect of funding levels from the ERDF and ESF on aggregate productivity growth in Italian commuting zones over a funding period. We begin by providing a general overview of our novel strategy – an application of Borusyak and Hull (2023) – in Section 5.1. The Borusyak and Hull (2023) methodology is particularly well-suited for this estimation, as it is specifically designed to isolate the causal effects of a treatment combining endogenous exposure to broader exogenous shocks. In our case we leverage the role of exogenous measurement error shocks at the *regional* level in the treatment assignment process – calculating deviations in actual financial support from the expected financial envelope available to *commuting zones* that are due to this measurement error, which is independent from a region's (true) GDP per capita or demand for cohesion policy by commuting-zones. Then, identification

[22] We define surviving firms as firms that are not identified as an entering or exiting firm.

comes by comparing commuting zones which would be expected to receive the same funding level, but in reality received different levels due to measurement error. We then detail implementation in Section 5.2, concurrently discussing potential violations of the assumption of exogenenity of measurement error shocks and how we account for them. Finally, in Section 5.3 we demonstrate that measurement error leads to meaningful deviations in funding envelopes available to commuting zones in-practice – the variation used in our aggregate analysis in Section 6.

5.1 Overview: Identification With Partially Exogenous Treatment

Borusyak and Hull (2023) develop a method to estimate causal effects of a treatment consisting of nonrandom exposure to exogenous shocks. The method may be used in cases where the treatment combines multiple sources of variation with a known formula. Intuitively, the econometrician observes a single realization of this treatment process combining both the endogenous exposure and exogenous shocks. Borusyak and Hull (2023) show that by constructing counterfactual treatments using exogenous variation over many simulations and averaging them to an expected treatment, the omitted variable bias from the endogenous exposure to the treatment is purged when controlling for the expected treatment in the regression – isolating the causal effect by comparing two units that expected to receive the same treatment, but did not due to the exogenous shocks.[23]

Weighting each observation by the start-of-period commuting zone population, our empirical specification is therefore:

$$y_{ct} = \beta_0 + \beta_1 \log\left(FUNDSpc\right)_{ct} + \beta_2 \log\left(\mathbb{E}\left[FUNDSpc\right]\right)_{ct} + \beta_3 x_{ct} + \eta_t + \epsilon_{ct} \qquad (4)$$

Where $FUNDSpc_{ct}$ and $\mathbb{E}\left[FUNDSpc\right]_{ct}$ are the actual and averaged counterfactual per capita funding in the commuting zone c during funding period t, respectively (pooling funding from the ERDF and ESF) and x_{ct} are control variables capturing a commuting zone's initial characteristics such as start-of-period[24] demographic characteristics, which may independently affect productivity. All specifications control for initial productivity level and include funding period fixed effects, η_t (to account for time-specific shocks such as the Great

[23] Borusyak and Hull (2023) show that the researcher may either recenter and instrument for the treatment by subtracting the expected treatment from the realization, or control for the expected treatment as in a control function approach. As recommended in their paper's conclusion in the case of natural experiments compared to true random experiments such as RCTs, we include the expected counterfactual treatment as a control in the regression.

[24] To be more precise, we use census data which is only available for 2001 and 2011. The 2001 controls are 6 years before the start of period and the 2011 controls are 3 years before the start of period.

Financial Crisis). Standard errors are clustered at the regional level to account for treatment assignment at the regional level. Our outcome variables y_{ct} are the commuting-zone-level change in aggregate productivity over the funding period, $\Delta\Phi_{ct}$, and its four components discussed in Equation 3 of Section 4.2.[25] Our coefficient of interest is β_1, representing the effect of more funding conditional on the expected funding.

5.2 Construction of Counterfactual Shocks

In our setting, we observe non-random exposure of commuting zones to exogenous regional measurement error shocks to GDP per capita. The endogenous portion of the treatment combines both the overall level of development of the region and differences in demand for the policy within the region.

In this section, we first discuss the role of measurement error in treatment assignment. We then provide details about our construction of counterfactual measurement error shocks. We jointly simulate measurement error in regional GDP both in levels and per capita because the funding formulas utilize both statistics.[26] Detailed information about the funding formulas and the data used to construct the counterfactual levels of funding can be found in Appendix B.4. Finally, we also take into account potential violations of the exogeneity of measurement error in our counterfactual generation process. Specifically, we account for serial autocorrelation of GDP estimates as well as the potential relationship between measurement error and a region's economic characteristics, which will be discussed below.[27]

5.2.1 The Role of Measurement Error

Imagine that EU GDP per capita was €24,000, and both Sicily and Campania had the same true GDP per capita of €18,000 – 75% of the EU average, which would make them eligible for less-developed status funding as discussed previously in Section 2. Within Sicily and Campania, commuting zones are differently *exposed* to measurement error shocks, depending on how much funding they attract. However, further imagine that there are two commuting zones in Sicily and Campania with the exact same demand for cohesion funding and industry composition. In the absence of measurement error, those two commuting zones

[25]Specifically, we look at change over a period of seven years – 2006-2013 and 2014-2021, respectively.

[26]The formulas also require population statistics. Historical population data are not readily available in much of the EU, so we back out historical population estimates from the GDP and GDP per capita data. The counterfactual population errors are similarly calculated using the ratios of the counterfactual GDP and GDP per capita estimates.

[27]The only aspect of the funding assignment process that is impossible for us to precisely replicate is the three-year-average of GDP and GDP per capita discussed in Section 2 due to limitations in the historical GDP data releases. See Appendix B.4 for further discussion.

would receive the same level of funding. Now imagine Sicily had a "good" measurement error shock, with GDP per capita measured as €17,000. However, imagine Campania had a "bad" measurement error shock, and their GDP per capita was measured as €19,000 – they would be classified as a transition region and receive at most 40% of what they would as a less developed region. In this simplified example, the identifying variation would be how much more funding the commuting zone in Sicily got than expected vs. how much less the commuting zone in Campania got than expected specifically due to these measurement error shocks.

More formally, we can expand the formula for $FUNDS$:

$$FUNDS_{ct} = s_{ct} \cdot FUNDS_{rt} = s_{ct} \cdot g(GDPpc_{rt}, GDPpc_{-rt}, ME_{rt}, ME_{-rt}) \tag{5}$$

where $FUNDS_{rt}$ is the funding pot for region r, and s_{ct} is the share of region r's total funding which flows to commuting zone c. The share of the region's funding flowing to a given commuting zone captures endogenous exposure within region. Variation in s_{ct} across commuting zones may reflect differences in local demand for funds or varying capacities to attract funding — which could be correlated with commuting-zone-level trends in aggregate productivity. In addition, differences in funding shares may arise from pre-existing industry compositions: for instance, commuting zones with a high employment share in targeted sectors (such as tourism) may be more likely to receive cohesion policy funding.

$FUNDS_{rt}$ is a function of both the true regional GDP per capita (also endogenous), $GDPpc_{rt}$, and the measurement error, ME_{rt}, associated with the regional GDP per capita (an exogenous shock). As discussed in Section 2, funding for the ERDF and ESF depends on a measurement of regional GDP per capita at a specific time, and funding is sticky throughout the seven-year period. Also recall that there are two channels through which measurement error affects the regional funding pots. First, different realizations of measurement error may move regions above or below the funding thresholds. Second, the funding formulas take into account how far from the threshold a region is conditional on being, for example, classified as a less-developed region. Furthermore, since the funding rules are based on *relative* levels of prosperity across the EU, the level of funding is also a function of the true GDP per capita of all other EU regions, $GDPpc_{-rt}$, as well as their measurement error shocks, ME_{-rt}. In other words, even if GDP of all Italian regions had been precisely estimated ($ME_{rt} = 0$), the expected funding envelope could differ from the actual funding allocated if the measurement error of other regions in Europe moves the EU average.

A discussed in Section 3, we digitized historical estimates of regional GDP and regional GDP per capita for all European regions from 1998-2009. Using this data, we calculate empirical distributions of measurement error by comparing historical estimates to revised estimates. Empirical probability density functions (PDF) of the measurement error in GDP and GDP per capita are shown in Figure 2. For both, the measurement error is centered around zero, so there is not systematic underestimation of GDP per capita.[28] There are also fat tails on both ends of the distribution and extreme values at roughly +/-30 % error – representing very large shocks.[29]

5.2.2 Serial Correlation

The key identifying assumption – that measurement error represents an exogenous shock – would be violated if a region's past measurement error fully predicted its current measurement error. In this section, we show that although measurement error in both GDP and GDP per capita is serially correlated, a substantial random component remains. It is this unpredictable variation that we use in constructing our counterfactual measurement error shocks.

GDP estimates are released by statistical offices year after year and are later subject to multiple revisions. It is not unlikely that if a team is making an error for the latest GDP estimate, it's similar to an error they made the previous year. Due to this, we would expect serial correlation across time in the measurement error within a region. However, over time we would also expect the teams to correct their mistakes, then make different ones, but be right on average. Concretely, we would expect that in two adjacent years the team might be overestimating (underestimating) a region's GDP, but ten years later they are much less likely to be making the same mistake.

First, we test these hypotheses – that measurement error is serially correlated but that serial correlation diminishes over time. Figure 3 shows the correlation of the percent error in GDP in 2009 with the values for 2008, 2002, and 1998 in Panels A, B, and C, respectively. The figures demonstrate our two hypotheses to hold empirically as the correlation declines over time, to a very weak correlation after eleven years have passed. The line graph in Panel D shows the correlation coefficient of the 1998 GDP estimate with subsequent years' estimates, which decays steadily year-over-year.

We incorporate this serial correlation in counterfactual construction in all specifications. Specifically, we estimate an Autoregressive (AR) model of first order using the measurement

[28]Measurement error would not be exogenous if regions purposely misreported their GDP per capita in order to qualify for more generous cohesion funding. Previous studies on cohesion policy exploiting the discontinuity in funding intensity around the 75% cutoff show that there is no evidence of bunching around the threshold (Becker et al., 2010).

[29]Note that it is slightly more common to severely overestimate rather than severely underestimate.

error panel data to back out the distribution of the remaining random component, which we use for the simulations of counterfactual regional funding envelopes.

$$\%Error_{GDP,rt} = \gamma_0 + \gamma_1 \%Error_{GDP,r,t-1} + u_{rt} \tag{6}$$

The results are shown in Panel A of Table 2. As expected, the lagged percent error in GDP has high explanatory power with respect to the current period's error, with a coefficient of .738 and an adjusted R squared of .629. We construct the distribution of counterfactual measurement errors using this model beginning with the actual 1998 measurement errors of the EU regions as the initial condition and simulate alternative paths of measurement error by drawing from the distribution of the estimated model residuals $\hat{u}_{rt}$.[30]

Since key elements of the funding formulas depend on GDP per capita, including the key 75% and 90% thresholds, we also require a counterfactual estimate of GDP per capita. It is likely that if a team is overestimating (underestimating) GDP, they are also overestimating (underestimating) GDP per capita. We thus estimate a regression of regional GDP per capita measurement error on regional GDP measurement error and a lag of regional GDP per capita measurement error:

$$\%Error_{GDPpc,rt} = \lambda_0 + \lambda_1 \%Error_{GDP,rt} + \lambda_2 \%Error_{GDPpc,r,t-1} + v_{rt} \tag{7}$$

The results are shown in Panel B of Table 2 – the estimated coefficient of .864 for current regional GDP (λ_1) confirms that similar errors are being made in estimation of GDP per capita as GDP, while the smaller coefficient of .148 of lagged GDP per capita (λ_2) reflects residual autocorrelation in population estimates. The adjusted R squared of .932 confirms the intuition described above – that the error in GDP and GDP per capita are highly interdependent.

5.2.3 Measurement Error and Regional Characteristics

The identification strategy of Borusyak and Hull (2023) relies upon finding an exogenous shock to construct counterfactual treatments differing from the realized treatment only because of exogenous variation. Our identification would be threatened if measurement error were not randomly assigned. Phrased another way, if our exogenous shock were not truly exogenous. This assumption would be violated if, for example, measurement error was sys-

[30] For the regions where 1998 historical GDP data is not available (Bulgaria, Cyprus, the Brandenburg region of Germany, the autonomous cities of Meililla and Ceuta in Spain, Finland, the autonomous region of Trento/Bolzano in Italy, Malta, and Portugal) we use a random draw from the distribution of all measurement errors as the initial condition.

tematically worse in some countries or in more developed compared to less developed regions. To test whether this is the case, we repeat the specifications in Equations (6) and (7) with some additional controls.

$$\%Error_{GDP,tr} = \gamma_0' + \gamma_1' \%Error_{GDP,t-1,r} + \gamma_2' X_{rt} + \zeta_n + u_{rt}' \tag{8}$$

$$\%Error_{GDPpc,tr} = \lambda_0' + \lambda_1' \%Error_{GDP,tr} + \lambda_2' \%Error_{GDPpc,t-1,r} + \lambda_3' X_{rt} + \zeta_n + v_{rt}' \tag{9}$$

where X_{rt} is a vector of other variables used in the funding determination process. Including this vector is critical. Since these variables are used to determine funding intensity, if measurement error is partially dependent on them we will, by definition, be using some endogenous variation when simulating the measurement error process. ζ_n are member state fixed effects to incorporate potentially correlated errors made by national statistical agencies when estimating GDP.

Table 3 shows the results of the regression for both regional GDP and regional GDP per capita. As is clear from the table, there are a number of other funding determinants and country fixed effects which are statistically significant contributors to the measurement error in GDP and GDP per capita. However, many of these coefficients are of little economic significance. For example, the coefficient on revised GDP in Column (1) is statistically significant, but small at -0.0173. The economic interpretation of this coefficient would be that an increase of regional GDP by one million Euros would increase the percentage error in the GDP estimate by roughly two-hundredths of a percentage point. Given that the standard deviation of the outcome variable is 6.23 percentage points as reported in the bottom of the table, this is of little practical significance.

These statistically significant coefficients are also unlikely to meaningfully change the results of the counterfactual measurement error simulation process: the adjusted R-squared barely increases when adding the additional covariates and country fixed effects, moving it from .629 to .646. The coefficients on the lagged GDPs also do not meaningfully change. For example, the coefficient of .738 in Table 2 declines to 0.719 in Table 3.

Our counterfactual measurement error process centers around simulation of residual terms using the empirical distributions of $\hat{u}_{rt}$ and $\hat{v}_{rt}$ in the case without additional covariates and $\hat{u}_{rt}'$ and $\hat{v}_{rt}'$ in the case with additional covariates. Then, we use the counterfactual lagged GDP year after year to simulate a different walk of GDP measurement error that may have occurred under different circumstances. Since the adjusted R-squared does not meaningfully change with the additional covariates, the distribution of residuals is nearly identical. And since the lag coefficients also barely affected, the walk of GDP measurement error also does not meaningfully change. In a robustness check in Section 6.3 and confirm that results are near-identical when accounting for the relationship between measurement error, other funding determinants, and country fixed effects.

5.3 Measurement Error's Effects on Funding Levels

Measurement error translates into meaningful differences in both which eligibility category Italian regions fall into and between expected and actual funding conditional on assignment. Table 4 shows the average funding envelope of each Italian region after 1000 counterfactual simulations of measurement error conditional on the eligibility bin that regions are sorted into.[31] The mean in each column of the table represents the effect of moving across eligibility bins on funding intensity. For example, in the 2007-2013 funding period, when Puglia was classified as a less developed region in the counterfactuals the average simulated funding package amounted to 1207.42 Euros per capita. When Puglia was classified as a transition region, the average funding package was less than half at 543.73 Euros per capita. The standard deviations, shown in parentheses, show the effect of the "intensive" margin of the funding – i.e. the effects of how far the region is from the key 75% or 90% thresholds. In the case of Puglia, the standard deviations are 131.22 and 14.82 Euros per capita, respectively.

A condition for identification in Borusyak and Hull (2023) is that the expected shock is not perfectly collinear with the realized shock. While this assumption will be tested formally in the regressions, Figure 4 shows the comparison between the actual and expected funding at the commuting-zone level. As is shown on the plot, the expected funding has a strong "first stage" – meaning that expected funding is a good predictor for actual funding. At the same time. we also have many points both above and below the 45-degree line shown in red, demonstrating that some areas received positive and others negative measurement error shocks.

6 Results

In this section, we study the impact of cohesion policy on changes in aggregate productivity at the commuting zone level. We begin by showing the effects of total spending over seven-year funding periods on the change in aggregate productivity. We then decompose aggregate productivity changes to investigate the mechanisms driving our findings.

[31] Appendix Table C.2.1 shows the number of times the regions are classified in each eligibility bin.

6.1 Aggregate Analysis

Table 5 reports the results of our analysis for aggregate productivity growth over the 7-year funding periods. The coefficient on the actual payments in Column (1) suggests that a 1% exogenous increase in the actual per capita funding leads to a 0.0688% decrease in aggregate productivity. In contrast, the coefficient for the expected payments in Column (1) is positive, indicating that commuting zones which receive higher expected payments have higher aggregate productivity growth conditional on funding period and initial productivity level. Although the coefficients are not statistically significant, this suggests that the endogenous demand mechanism discussed in Section 5 dominates – the commuting zones that are successful in attracting more funds per capita are also those that would be expected to have higher productivity growth. Recall that conditioning on the expected amount of funding is crucial, since we want to identify the effect of cohesion policy using only the exogenous part of the variation in funding levels.

In Column (2) we add two additional controls for EU payments. First, payments for projects under the ERDF and ESF that affect the entire region rather than a specific commuting zone, and second, payments from non ERDF and ESF programs such as the European Territorial Cooperation. Controlling for these other programs does not meaningfully change the results, with the new coefficient being -0.0710 and statistically indistinguishable from the result in Column (1). In Column (3) we control for these demographic characteristics, and the aggregate result, while still negative at -0.0245, is no longer statistically significant. Controlling for both other EU payments and demographic characteristics in Column (4) gives a very similar result at -0.0290.

Altogether, this suggests a negative effect of cohesion funds on aggregate productivity growth, albeit a noisy one. However, the scope of cohesion policy is very broad, and it is possible that a noisy/weak negative aggregate effect masks competing positive and negative effects in different areas of the economy.

6.2 Decomposing Aggregate Results

To better understand the mechanisms behind the aggregate productivity impacts of the place-based policy, we decompose the change in aggregate productivity in its four components as discussed in Section 4. Table 6 reports the results of separate regressions including all controls from Column (4) of Table 5. Thus, Column (1) displays the same result as Column (4) of Table 5, i.e. the aggregate productivity effect over the funding period with the full set of controls. In Column (2), the average unweighted productivity effect is shown. The coefficient is near-zero and statistically insignificant, implying that there was no symmetric within-firm productivity change as a result of the place-based policy.

Column (3) reveals that the weak negative aggregate productivity effects of cohesion policy stem from decreases in allocative efficiency in the commuting zone. In other words, in local labor markets that received more cohesion funds than expected, unproductive firms grew (in their relative share of employment) relative to highly productive firms. The coefficient implies that a 1% increase in cohesion policy funding leads to a 0.0636% reduction in allocative efficiency. Furthermore, the losses in labor reallocation across firms are not offset by productivity gains at the firm-level as discussed previously with respect to the results in Column (2). On the contrary, firms in these commuting zones have on average not experienced any gains in productivity, despite the efforts of cohesion policy to provide funding for public goods such as critical infrastructure, which in theory could have the potential to boost productivity of many firms in the targeted commuting zone, leading to a rightward shift of the firm-level productivity distribution. We confirm the lack of a symmetric productivity shift following completions of large (one-million-Euros or more) projects in Appendix A.[32]

The final two Columns (4) and (5) show the effects of funding on the entry and exit components of productivity, respectively. Column (4) shows that, despite the overall negative labor reallocation effects, there are positive and significant effects on the entry margin: new entrants are, on-average, more productive than surviving firms as a result of the place-based policy. In the overall effect, the positive entry effect, with a point estimate of 0.0321, is simply dominated by the larger labor negative reallocation effect in the overall estimate in Column (1).

It is also important to know which industries, if any, are driving the results presented in Table 7. Panel A shows that the negative labor reallocation effects seen in Column (3) are being driven by the Manufacturing sector. Panels B and C show that the positive effects on entry seen in Column (4) of Table 6, by contrast, are being driven by the Services and Construction sectors.[33] Overall, these results imply that cohesion policy has negative effects on aggregate productivity – particularly in the manufacturing sector – driven by reallocation of labor to less productive firms, but smaller in magnitude positive effects on entry in the Services and Construction sectors. Combined, the two lead to a weak negative effect on aggregate productivity.

[32]This is consistent with the finding of Albanese et al. (2021) that cohesion policy does not correlate with TFP growth, measured as unweighted average of firm-level data, in Italan local labor markets.

[33]Somewhat surprisingly, we also observe a symmetric productivity effect in the construction sector, a sector characterized by sluggish productivity growth (Goolsbee and Syverson, 2023). We suspect that effects in the construction sector may be driven by increased prices charged by construction firms aware of the availability of cohesion funds – which would appear in the data as higher value added per worker. Unfortunately, we cannot test this hypothesis since we do not observe prices or quantities in our balance sheet data.

6.3 Robustness

Measurement error exogeneity As discussed in Section 5.2.3, our identification could be threatened if measurement error is correlated with other endogenous characteristics of the region. Repeating the counterfactual simulation process incorporating the additional covariates in equation (8), we show that this is not a concern in practice.[34] The results of the analysis using this more complex method to create $E[Funds]_{ct}$ are shown in Table 8. Reassuringly, the results are extremely similar to the specification with the simpler method for simulating the counterfactual measurement errors.

TFP We test for potential differences in our results when measuring productivity as log TFP rather than log worker value added in Table 9. The results are consistent with the main results using labor productivity. All coefficients of the decomposition are the same sign, and both the negative labor reallocation term and the positive entry term remain statistically significant, though both are smaller in magnitude. The main distinction of the TFP results is that the symmetric component is now marginally significant at the 10% level, though it remains of similar size (.0166 compared to .0107 in the baseline results). As discussed previously, we explore the symmetric effect further in Appendix A, and find no evidence for symmetric productivity shifts for firms in commuting zones where large projects are undertaken for either worker value-added or TFP.

Funding Absorption In Table 10 we explore alternative sample compositions to examine how robust our results are in areas which did not manage to absorb cohesion funding well[35], which has been shown in previous research to be an important driver of potential positive growth effects (Becker et al., 2013). We test if the negative observed effects are confined to areas which were able to utilize most of the funds. In Column (1), we investigate whether our results are being driven by richer or poorer regions by including an interaction term with more developed status as more developed regions in Italy typically have higher absorption (Fritz, Incoronato, and van der List, 2025).[36] In practice, this interaction term can be

[34] This necessitates a few additional steps. First, since we cannot back-code the values of the covariates for some EU regions as discussed in Appendix B, for those regions we use the simple lag model with the coefficients and residual distributions in Table 3. Second, some of the data used are only available from Eurostat beginning in the year 2000, so we fill in the values for 2000 for 1999 to complete the first year of the simulation. Finally, for a small number of regions covariate data is missing for a longer period of time, and in those cases we also fill in the data with the first known value.

[35] See Fritz, Incoronato, and van der List (2025) for detailed description of funding absorption in Italy.

[36] Given the small number of regions in the transition category, we opted for a simpler binary classification. In the 2007-2013 funding period only Basilicata was classified as a transition region, and in the 2014-2020 funding period only Basilicata, Molise, and Sardinia were.

thought of a north/south dummy. In Panel A, the reallocation effect remains statistically significant and similar in magnitude to the main results at -0.061. The entry effect shown in Panel B also remains statistically significant and similar in magnitude to the main results at 0.0296. In both cases, the interaction term brings the overall result closer to zero – being positive in Panel A and negative in Panel B – meaning that the effect of cohesion policy is somewhat less pronounced in more developed regions.

In Column (2), we exclude Sicily where funding absorption was particularly poor and in Column (3) we exclude from our analysis any commuting zones located in region-by-funding-period cells that committed funding amounting to less than 90% of their allocated financial envelope.[37] When excluding Sicily, the point estimates are similar to the main results at -0.0652 and 0.0371 for reallocation and entry, respectively. When imposing the strict funding absorption condition in Column (3), the coefficients of interest become smaller in magnitude and statistically insignificant but remain the same sign.

Heterogeneity over time and alternative span In Table 11 we show the results separately for the 2007-2013 funding period and the 2014-2020 funding period in Panels A and B, respectively. The results are similar for both funding periods, although the point estimates for the 2007-2013 period are smaller in magnitude, and the entry effect is only significant in the 2014-2020 funding period.

Under cohesion policy, regions are permitted to complete projects up to three years after the end of the funding period. So, some spending occurs after the funding period officially ends. In Table 12 we show the results for a 10-year difference (2006-2016) for the 2007-2013 funding period.[38] The results are similar to the 2007-2013 period results in Table 11 Panel A, with a significant negative reallocation effect of -0.0303 and a significant positive entry effect of 0.0279.

Further robustness checks In Appendix Table C.2.2 we show the results of additional robustness checks. In Column (1) we include region fixed effects, in Column (2) we include additional controls for the industrial composition of the commuting zone, and in Column (3) we present results where the key controls are total funding (rather than per capita as in

[37]Note that the commitment condition in Column (3) is very stringent as we include only funding commitments for projects that take place only in a single region and are funded through a single regional funding pot. In other words, a reason that roughly 40% of the regions commit less than 90% of their allocated funds is that we cannot account for multi-region or multi-funding-source projects – which are 7.54% of all funding commitments from the ERDF and ESF which are not from a national program (Fritz, Incoronato, and van der List, 2025).

[38]We are only able to include the 2007-2013 funding period since the Orbis data only ran until 2022 at the time of this draft.

all main results) including an additional control for commuting zone population. The reallocation effect survives all specifications, while the positive entry effect is noisier and only significant in Column (2), but always remains positive.

7 Mechanism Analysis

The negative labor reallocation seen in the decomposition results may be due to two distinct mechanisms. The first is that unproductive firms grow more as a result of cohesion policy, resulting in a larger employment share and consequently a slowdown in aggregate productivity. This may actually be a positive result in the short run from the perspective of workers in the commuting zone if the new hires are pulled from unemployment or those marginally attached to the labor force. The second possibility is that productive firms have less employment than they otherwise would as a result of cohesion policy. This may happen if, for example, productive firms use money from cohesion policy to automate production. Since some of their relative employment would then be replaced by capital, their employment share could decline and lead to the negative labor reallocation effects seen in the decomposition.

7.1 Aggregate Labor Market Effects

The first test we employ to distinguish between these two potential mechanisms is to repeat our aggregate analysis at the commuting-zone level, but with growth in unemployment rates as the dependent variable rather than aggregate productivity. The results are shown in Panel A of Table 13. Unemployment rates grow more in regions with more payments from cohesion policy conditional on the expected payments. This speaks against the hypothesis that unproductive firms used the funds to hire unemployed workers. However, increased unemployment rates could also come from those out of the labor force being induced to enter the labor force in areas more heavily invested by cohesion policy if the labor market improves due to the less productive firms hiring more. Panel B of Table 13 shows that this is unlikely to be the case. Although none of the coefficients are statistically significant, the sign is consistently negative. This suggests that in more heavily invested areas, if anything, labor force participation rates grew less. Overall, both results suggest that the second mechanism – more productive firms growing less – is at play.

7.2 Firm-Level Effects

The second test we employ to distinguish these two potential mechanisms is a firm-level analysis leveraging the VAT tax identifiers available in our funding data to link the implementers to the balance sheet data in Orbis.[39] While, by design, we focus exclusively on a subset of all projects co-financed by the EU – namely, those in which a private firm is listed as a beneficiary[40] – the following analyses allow us to understand what kind of firms select into cohesion policy, and how starting a project co-financed by the ERDF or the ESF affects firm outcomes. Specifically, we first compare descriptively the characteristics of beneficiary firms with those that do not directly benefit from cohesion policy and second conduct firm-level event studies.

Our balance tests show that it is not systematically small unproductive firms that directly benefit from cohesion funding. Specifically, in Table 14 we compare the characteristics of firms that take-up ERDF and ESF funds (in Panels A and B, respectively) with those that do not. Column (3) of both panels shows unconditional differences, while Column (4) controls for region and 2-digit-NACE fixed effects. On average, beneficiary firms are older, have more employees, higher sales and are more productive, conditional on region and industry. Notably, many of the differences are *larger* after adding controls in Column (4) – particularly value-added per worker. This speaks against the hypothesis that bad selection into cohesion policy is the driver for the diminished allocative efficiency.

Next, we run event studies around the start of a EU co-financed project in which a private firm is listed as beneficiary. Some firms benefit from cohesion policy funding multiple times. In such cases, we define treatment as starting in the year of the first project. We run the event studies separately for firms benefiting from the ERDF and the ESF.[41] Analogous to the aggregate analysis, we define the year 2006 as the pre-period and consider any firm treated between 2007 and 2019. We examine two main outcomes Y_{jt}: (log) firm value-added per worker, and (log) firm employment. The event studies are of the form

$$Y_{jt} = \alpha_j + \delta_t + \sum_{\substack{i=-2 \\ i\neq -1}}^{k} \gamma_i \cdot 1(t = \text{Year Treatment}_j + i) + \kappa_m + \zeta_r + \epsilon_{jt} \tag{10}$$

[39]Recall that the OpenCoesione database distinguishes between the project planner (*programmatore*) and three implementation roles: *attuatore*, *beneficiario*, and *realizzatore*. We classify a firm as treated if it is listed as a project beneficiary (*beneficiario*) or implementor (*attuatore*), allowing it to hold multiple roles within the same project (e.g., both beneficiary and executor (*realizzatore*)).

[40]In other words, projects where the beneficiary is an individual, NGO, or public authority are (by design) excluded from the analysis.

[41]If a firm first received ESF funding in 2013 and ERDF funding in 2017, it is considered not-yet-treated in the ERDF event study until 2016.

where j indexes firm, t indexes year, κ_m are industry fixed effects (2-digit NACE) and ζ_r are region fixed effects. We use the estimator of Callaway and Sant'Anna (2021) due to the staggered treatment timing. We run the same analysis twice – using the never-treated and the not-yet treated as control group, respectively. From Table 14 we already know that firms receiving cohesion funding are different from never-treated firms along a number of dimensions and are therefore less likely to represent a suitable control group for treated firms. However, comparing growth trajectories of treated and untreated firms sheds further light on (self-)selection into the program. In a second step, we use yet-to-be-treated firms as the control group under the assumption that the later-treated firms are similar to the earlier-treated firms with the exception of treatment timing. Although it is not a requirement that firms be balanced on pre-treatment characteristics, it is reassuring if firms have similar pre-treatment characteristics as time-specific shocks may have differential impacts based on these characteristics and potentially bias results. As a test of the hypothesis, we show in Table 15 the results of a regression of year-of-treatment on the characteristics of firms (in $t = -1$). Although there are statistically significant differences in treatment timing based on firm characteristics, these differences are not of economic significance. For example, the interpretation of the coefficient of 0.320 on Log real value added per worker is that a 1% increase in log real value added per worker just prior to treatment predicts a later start of a cohesion funded project by .003 years (one day). For both outcomes, there is no evidence of pre-trends when using the yet-to-be-treated a the control group.

The event studies using never-treated firms as a control group serve to visualize pre-treatment growth trajectories and offer informative evidence on how cohesion policy influences firm-level outcomes – providing additional insight into the underlying mechanisms. The results are shown in Figure 5. Panels A and B show the results for productivity for the ERDF and ESF, respectively, and Panels C and D show the results for employment for the ERDF and ESF, respectively. The figure demonstrates that firms that receive cohesion policy funding are not only more productive compared to the never-treated firms, they are also on a stronger growth trajectory in terms of both productivity and employment. Interestingly, employment growth levels off within one or two years after beginning the cohesion project in the case of the ERDF, while in the case of the ESF it begins declining after several years.

To isolate the causal effect of funding, we use not-yet-treated firms as the control group in Figure 6. The results of the regression with the firm's worker value added as the dependent variable are shown in Panels A and B for the ERDF and ESF, respectively. We find noisy results with respect to productivity following the treatment, with little evidence of any effect. The results of the regression with firm employment as the dependent variable are shown in Panels C and D for the ERDF and ESF, respectively. In contrast to the productivity event study, the negative effect on employment is large and persistent. The effect appears in the third year following the start of the project, which is consistent with the average project duration of approximately 1.5 years. In the case of the ESF, the initial increase in employment is consistent with the fact that the vast majority of ESF projects involve either training programs, or subsidies for hiring. From the third year onward, the point estimates grow in magnitude over time and remain statistically significant – with the point estimate five years after the start of the project suggesting funding-induced effects of 10% lower employment compared to the later treated firms for the ERDF and 25% for the ESF, with no effect on firm-level productivity. Combined with the never treated event study, the results demonstrate that firms that take-up cohesion funding level off (and later decline, in the case of the ESF) their employment growth trajectory after beginning the project.

Why is this the case? A potential explanation is that the projects lead to liquidity constraints within the firm, stifling their future growth. First, co-financing under cohesion policy is imperfect – meaning that firms are not funded for the entire cost of their project. In the case that the funds induce firms to invest in lower-value projects that they would not have undertaken in the absence of the funds, the firms may be left liquidity constrained afterwards when the projects do not lead to productivity gains. Second, cohesion policy typically is a reimbursement-based program, meaning firms start the project and are later reimbursed for costs.

We are able to test this hypothesis by running a set of event studies with the non-current liabilities to sales ratio as the dependent variable, a proxy for the financial situation of the firm (Lichter et al., 2025). The results are shown in Panels E and F of Figure 6 for the ERDF and ESF, respectively. The increase in liabilities compared to sales emerges immediately following the start of the project in the case of the ERDF, and after the second year for the ESF, and continues to rise over time.

Altogether, these results suggest that cohesion policy has negative impacts on workers in the short-term. More-productive firms take up the funds, and subsequently their growth trajectory (in employment) slows dramatically, with no effect on productivity. This seems to be due to imperfect co-financing for the projects leading to liquidity constraints on the part of the firms. In the aggregate, this translates into 1) lower aggregate productivity due to the relative negative labor reallocation to less productive firms and 2) higher unemployment rates faced by workers in the more heavily invested areas. Although we find negative medium-term impacts in this study, it is possible that there are long-term positive effects of these investments by the most productive firms which could lead to a "bounceback" for the workers over longer time periods. While such analysis is outside the scope of this paper, it remains an interesting question for future research.

8 Conclusion

How do place-based policies influence productivity in targeted areas? The answer to this question has important policy implications for national governments and supranational organizations such as the EU. Despite high levels of spending on such policies, evidence of their ability to create sustainable growth has been scant. Understanding productivity is key to this question, since productivity growth drives long-term GDP growth. In this paper, we investigate whether place-based policies can increase a location's productivity in the medium term using a novel identification strategy. Specifically, we use cutting-edge methods designed specifically to identify the effects of partially exogenous treatments by constructing counterfactual treatments which might have taken place under different realizations of an exogenous shock. To do so, we exploit quasi-experimental variation induced by measurement error in regional GDP estimates as well as the known funding formulas used in cohesion policy.

We find that cohesion policy, if anything, negatively affected aggregate productivity of Italian commuting zones which received a larger than expected level of funding due to a positive measurement error shock. Decomposing the productivity effects, we find that the results are driven by labor reallocation to less productive firms, particularly in the manufacturing sector. In our mechanism analysis using two complementary approaches, we find evidence that these negative reallocation effects are driven by more productive Italian firms receiving the funds and slowing their employment growth. This suggests that in the medium term, cohesion policy has negative impacts for workers and regions. However, the long-term impacts remain an open research question. Given that the firm-level employment effects we find seem to be driven, at least to some extent, by liquidity constraints faced by the firms it's possible that over longer time horizons the negative effects could be mitigated or even reverse themselves.

Our results speak to the importance of careful policy design and targeting of place-based policies, which have been identified as key to achieving positive effects with such policies (Austin et al., 2018). In fact, studies of place-based policies that are specifically designed to increase employment (Criscuolo et al., 2019; Ku et al., 2020; Grunau et al., 2025; Siegloch et al., 2025) typically find that the policies do what they are designed for. Cohesion policy, by contrast, is characterized by a *lack* of clear and specific policy objectives, with the design emphasizing lifting budget constraints of the less developed regions of the

EU. Where programs are broad in both financial scale and investment scope, many potential effects may offset one another. In our setting, the negative effects emerge dominant.

References

Gustavo Adler, Romain Duval, Davide Furceri, Sinem Kiliç Çelik, Ksenia Koloskova, and Marcos Poplawski-Ribeiro. *Gone with the headwinds: Global productivity.* International Monetary Fund, 2017.

Philippe Aghion, Jing Cai, Mathias Dewatripont, Luosha Du, Ann Harrison, and Patrick Legros. Industrial policy and competition. *American Economic Journal: Macroeconomics*, 7(4):1–32, 2015.

Giuseppe Albanese, Guido de Blasio, and Andrea Locatelli. Does EU regional policy promote local TFP growth? Evidence from the Italian Mezzogiorno. *Papers in Regional Science*, 100(2): 327–348, 2021.

Benjamin A Austin, Edward L Glaeser, and Lawrence H Summers. Jobs for the heartland: Place-based policies in 21st century america. Technical report, National Bureau of Economic Research, 2018.

David Autor, Anne Beck, David Dorn, and Gordon H Hanson. Help for the Heartland? The Employment and Electoral Effects of the Trump Tariffs in the United States. Technical report, National Bureau of Economic Research, 2024.

Guglielmo Barone, Francesco David, and Guido De Blasio. Boulevard of broken dreams. the end of EU funding (1997: Abruzzi, italy). *Regional Science and Urban Economics*, 60:31–38, 2016.

Timothy J Bartik. Using place-based jobs policies to help distressed communities. *Journal of Economic Perspectives*, 34(3):99–127, 2020.

Sascha O Becker, Peter H Egger, and Maximilian von Ehrlich. Going NUTS: The effect of EU Structural Funds on regional performance. *Journal of Public Economics*, 94(9-10):578–590, 2010.

Sascha O Becker, Peter H Egger, and Maximilian von Ehrlich. Absorptive capacity and the growth and investment effects of regional transfers: A regression discontinuity design with heterogeneous treatment effects. *American Economic Journal: Economic Policy*, 5(4):29–77, 2013.

Sascha O Becker, Peter H Egger, and Maximilian von Ehrlich. Effects of EU regional policy: 1989-2013. *Regional Science and Urban Economics*, 69:143–152, 2018.

Filippo Biondi, Sergio Inferrera, Matthias Mertens, and Javier Miranda. Declining Job Reallocation in Europe: The Role of Shocks, Market Power, and Technology. Technical report, IWH Discussion Papers, 2025.

Yashar Blouri and Maximilian von Ehrlich. On the optimal design of place-based policies: A structural evaluation of EU regional transfers. *Journal of international economics*, 125:103319, 2020.

Kirill Borusyak and Peter Hull. Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6): 2155–2185, 2023.

Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230, 2021.

Fabio Canova and Evi Pappa. The macroeconomic effects of EU regional Structural Funds. *Journal of the European Economic Association*, 23(1):327–360, 2025.

Augusto Cerqua and Guido Pellegrini. Do subsidies to private capital boost firms' growth? A multiple regression discontinuity design approach. *Journal of Public Economics*, 109:114–126, 2014.

Augusto Cerqua and Guido Pellegrini. How much does state aid mitigate employment losses? Local policy effects at a time of economic crisis. *Regional Studies*, 56(10):1698–1712, 2022.

Council of the European Union. Council Regulation (EC) No 1083/2006 of 11 July 2006 laying down general provisions on the European Regional Development Fund, the European Social Fund and the Cohesion Fund and repealing Regulation (EC) No 1260/1999, 2006. Accessed: 2025-04-28.

Council of the European Union. Regulation (EU) No 1301/2013 of the European Parliament and of the Council of 17 December 2013 on the European Regional Development Fund and on specific provisions concerning the Investment for growth and jobs goal and repealing Regulation (EC) no 1080/2006, 2013a. Accessed: 2025-04-28.

Council of the European Union. Regulation (EU) No 1303/2013 of the European Parliament and of the Council of 17 December 2013 laying down common provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund, the European Agricultural Fund for Rural Development and the European Maritime and Fisheries Fund and laying down general provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund and the European Maritime and Fisheries Fund and repealing Council Regulation (EC) No 1083/2006, 2013b. Accessed: 2025-04-28.

Chiara Criscuolo, Ralf Martin, Henry G Overman, and John Van Reenen. Some causal effects of an industrial policy. *American Economic Review*, 109(1):48–85, 2019.

Maarten De Ridder. Market power and innovation in the intangible economy. *American Economic Review*, 114(1):199–251, 2024.

Ryan A Decker, John Haltiwanger, Ron S Jarmin, and Javier Miranda. Declining dynamism, allocative efficiency, and the productivity slowdown. *American Economic Review*, 107(5):322–326, 2017.

Marco Di Cataldo. The impact of EU Objective 1 funds on regional development: Evidence from the UK and the prospect of Brexit. *Journal of Regional Science*, 57(5):814–839, 2017.

John Fernald, Robert Inklaar, and Dimitrije Ruzic. The productivity slowdown in advanced economies: common shocks or common trends? *Review of Income and Wealth*, 71(1):e12690, 2025.

Sarah Fritz, Lorenzo Incoronato, and Catherine van der List. Declining free lunch: State capacity and foregone public spending. Technical report, 2025.

Andrew Garin and Jonathan Rothbaum. The long-run impacts of public industrial investment on local development and economic mobility: Evidence from World War II. *The Quarterly Journal of Economics*, 140(1):459–520, 2025.

Mara Giua. Spatial discontinuity for the impact assessment of the EU regional policy: The case of Italian objective 1 regions. *Journal of Regional Science*, 57(1):109–131, 2017.

Edward L Glaeser and Joshua D Gottlieb. The economics of place-making policies. *Brookings Papers on Economic Activity*, pages 155–239, 2008.

Austan Goolsbee and Chad Syverson. The strange and awful path of productivity in the US construction sector. Technical report, National Bureau of Economic Research, 2023.

Philipp Grunau, Florian Hoffmann, Thomas Lemieux, and Mirko Titze. Who benefits from place-based policies? Evidence from matched employer-employee data. Technical report, National Bureau of Economic Research, 2025.

Stephan Heblich, Marlon Seror, Hao Xu, and Yanos Zylberberg. Industrial Clusters in the Long Run: Evidence from Million-Rouble Plants in China. Technical report, National Bureau of Economic Research, 2022.

Vincent Heddesheimer, Hanno Hilbig, and Andreas Wiedemann. Place-Based Policies, Local Responses, and Electoral Behavior. 2024.

Lorenzo Incoronato and Salvatore Lattanzio. Place-Based Industrial Policies and Local Agglomeration in the Long Run. Technical report, 2024.

Charles I Jones. The facts of economic growth. In *Handbook of macroeconomics*, volume 2, pages 3–69. Elsevier, 2016.

Réka Juhász, Nathan Lane, and Dani Rodrik. The new economics of industrial policy. *Annual Review of Economics*, 16, 2023.

Matthias Kehrig and Nicolas Vincent. The micro-level anatomy of the labor share decline. *The Quarterly Journal of Economics*, 136(2):1031–1087, 2021.

Patrick Kline and Enrico Moretti. Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. *The Quarterly Journal of Economics*, 129(1):275–331, 11 2013.

Hyejin Ku, Uta Schönberg, and Ragnhild C Schreiner. Do place-based tax incentives create jobs? *Journal of Public Economics*, 191:104105, 2020.

Andreas Lichter, Max Löffler, Ingo E Isphording, Thu-Van Nguyen, Felix Poege, and Sebastian Siegloch. Profit taxation, R&D spending, and innovation. *American Economic Journal: Economic Policy*, 17(1):432–463, 2025.

Marc J Melitz and Sašo Polanec. Dynamic Olley-Pakes productivity decomposition with entry and exit. *The Rand journal of economics*, 46(2):362–375, 2015.

Matti Mitrunen. War reparations, structural change, and intergenerational mobility. *The Quarterly Journal of Economics*, 140(1):521–584, 2025.

Philipp Mohl and Tobias Hagen. Do EU structural funds promote regional growth? New evidence from various panel data approaches. *Regional Science and Urban Economics*, 40(5):353–365, 2010.

David Neumark and Helen Simpson. Place-based policies. In *Handbook of Regional and Urban Economics*, volume 5, pages 1197–1287. Elsevier, 2015.

Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications industry. *Econometrica*, 64(6):1263–1298, 1996.

Guido Pellegrini, Flavia Terribile, Ornella Tarola, Teo Muccigrosso, and Federica Busillo. Measuring the effects of European Regional Policy on economic growth: A regression discontinuity approach. *Papers in Regional Science*, 92(1):217–233, 2013.

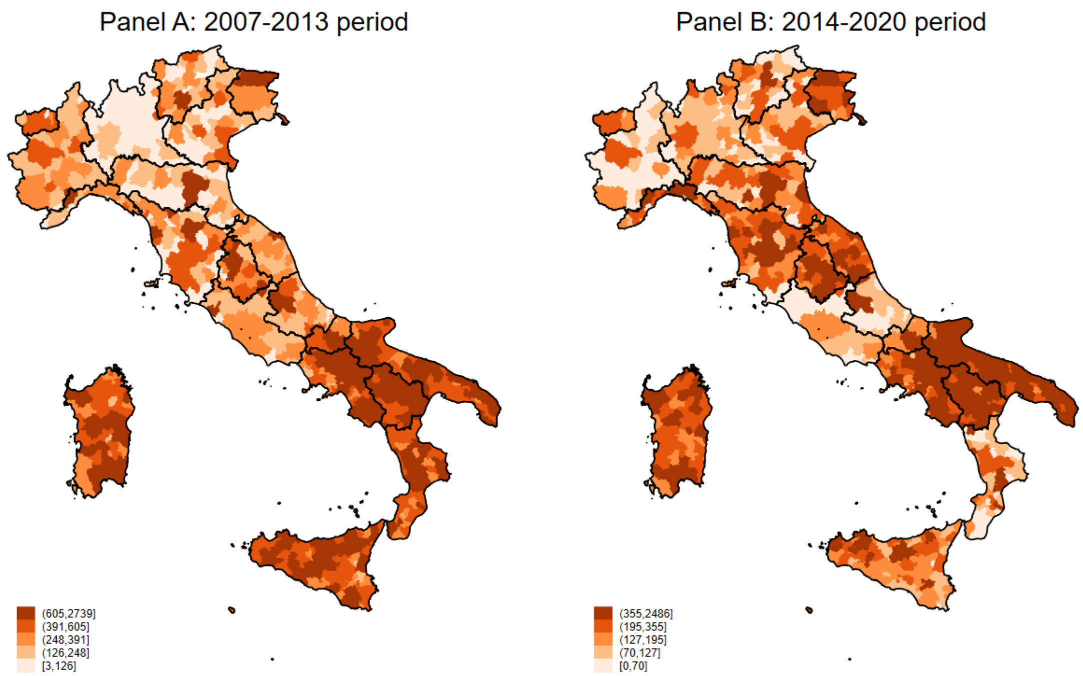
Sebastian Siegloch, Nils Wehrhöfer, and Tobias Etzel. Spillover, Efficiency, and Equity Effects of Regional Firm Subsidies. *American Economic Journal: Economic Policy*, 17(1):144–180, 2025.

Maximilian von Ehrlich and Henry G Overman. Place-based policies and spatial disparities across European cities. *Journal of economic perspectives*, 34(3):128–149, 2020.

Jeffrey M Wooldridge. On estimating firm-level production functions using proxy variables to control for unobservables. *Economics letters*, 104(3):112–114, 2009.

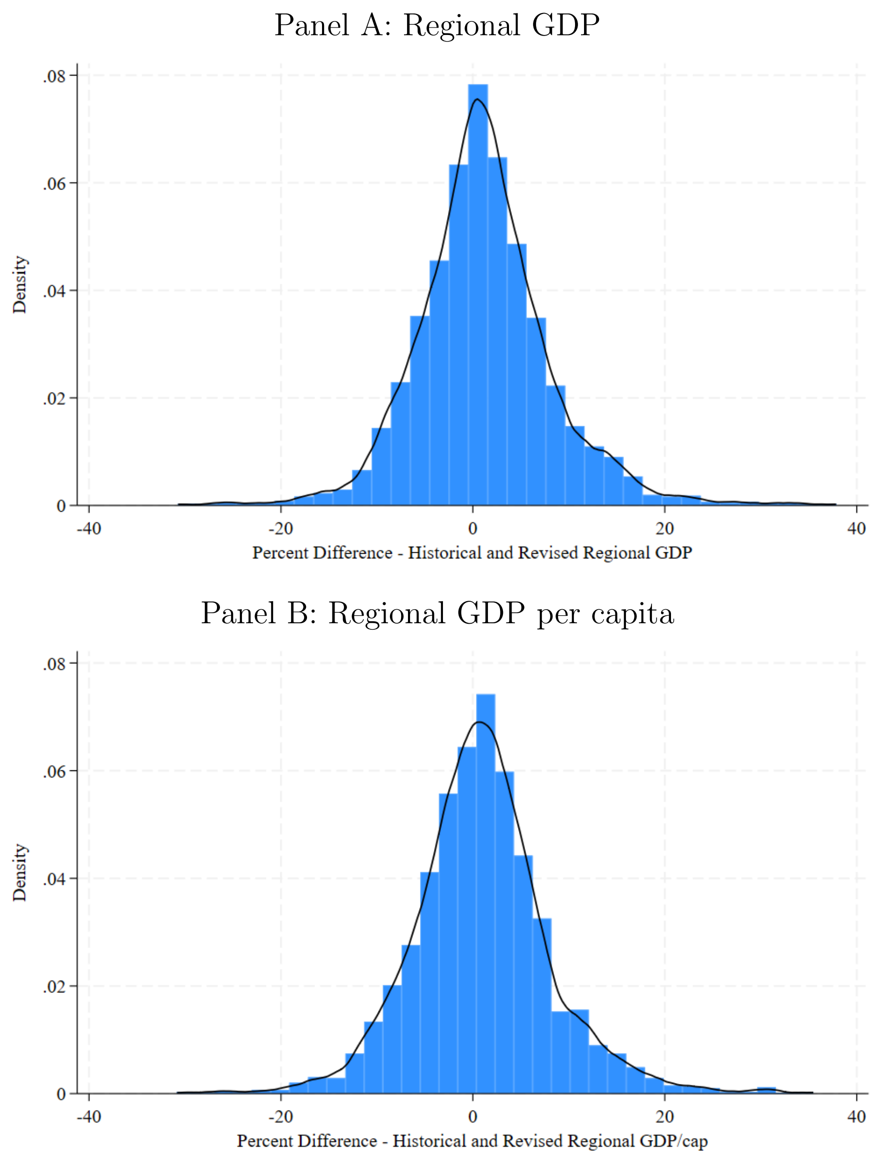
Figures

Figure 1: Payments per capita across Italian commuting zones



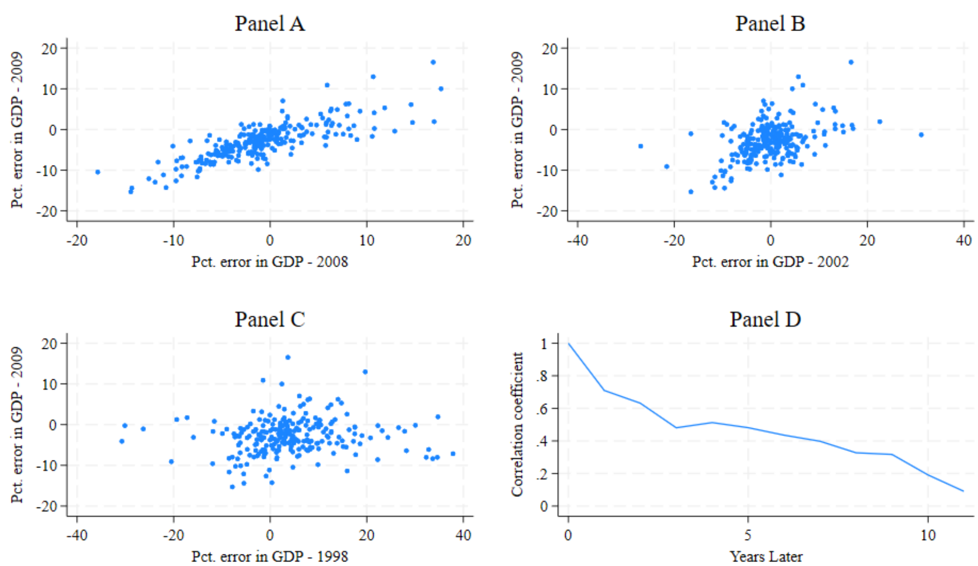
Notes: The sample includes projects funded by the ERDF and ESF excluding multi-funding-source projects as described in Appendix B.1. For projects taking place in multiple commuting zones, equal spending across both commuting zones is assumed. Per capita values are calculated using the population in the commuting zone during the first year of the funding period (2007 and 2014, respectively).

Figure 2: Empirical Probability Density Functions of Measurement Error



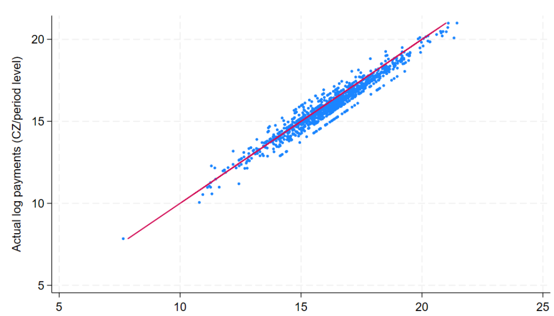
Notes: Authors' calculations using historical NUTS2 regional GDP and GDP per capita estimates for 1998 (or the earliest year available) to 2009 for all EU member states compared to revised figures. The percent error in GDP (per capita) is the percent difference between the historical GDP (per capita) estimates (the first estimate) and the revised estimates.

Figure 3: Serial Autocorrelation in GDP Measurement Error



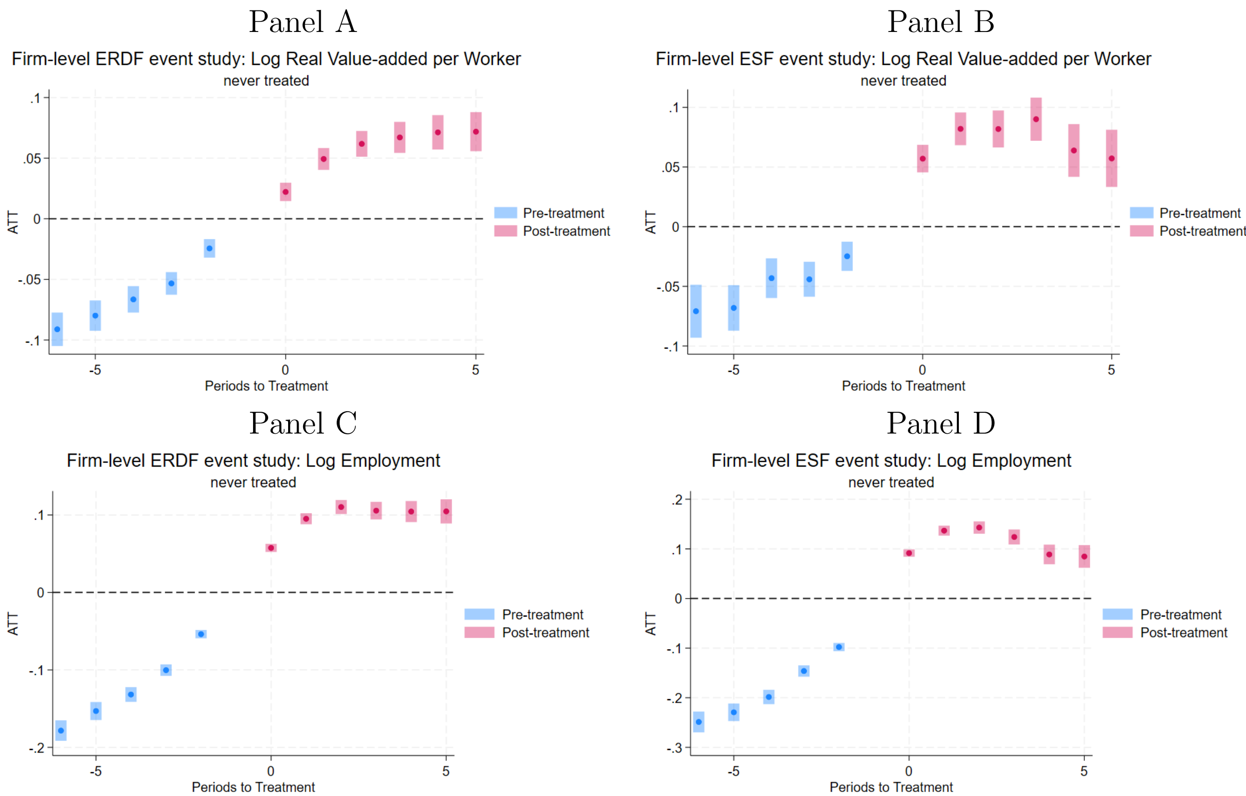
Notes: Panels A, B, and C show the relationship between the percent errors in the first NUTS2 regional GDP estimates compared to the revised ones for 2009 and 2008, 2002, and 1998, respectively, using the data available for all EU member states from the earliest year available. Panel D shows the correlation coefficient between the percent error in regional GDP 1998 and each subsequent year until 2009.

Figure 4: Actual vs. Counterfactual Log Funding



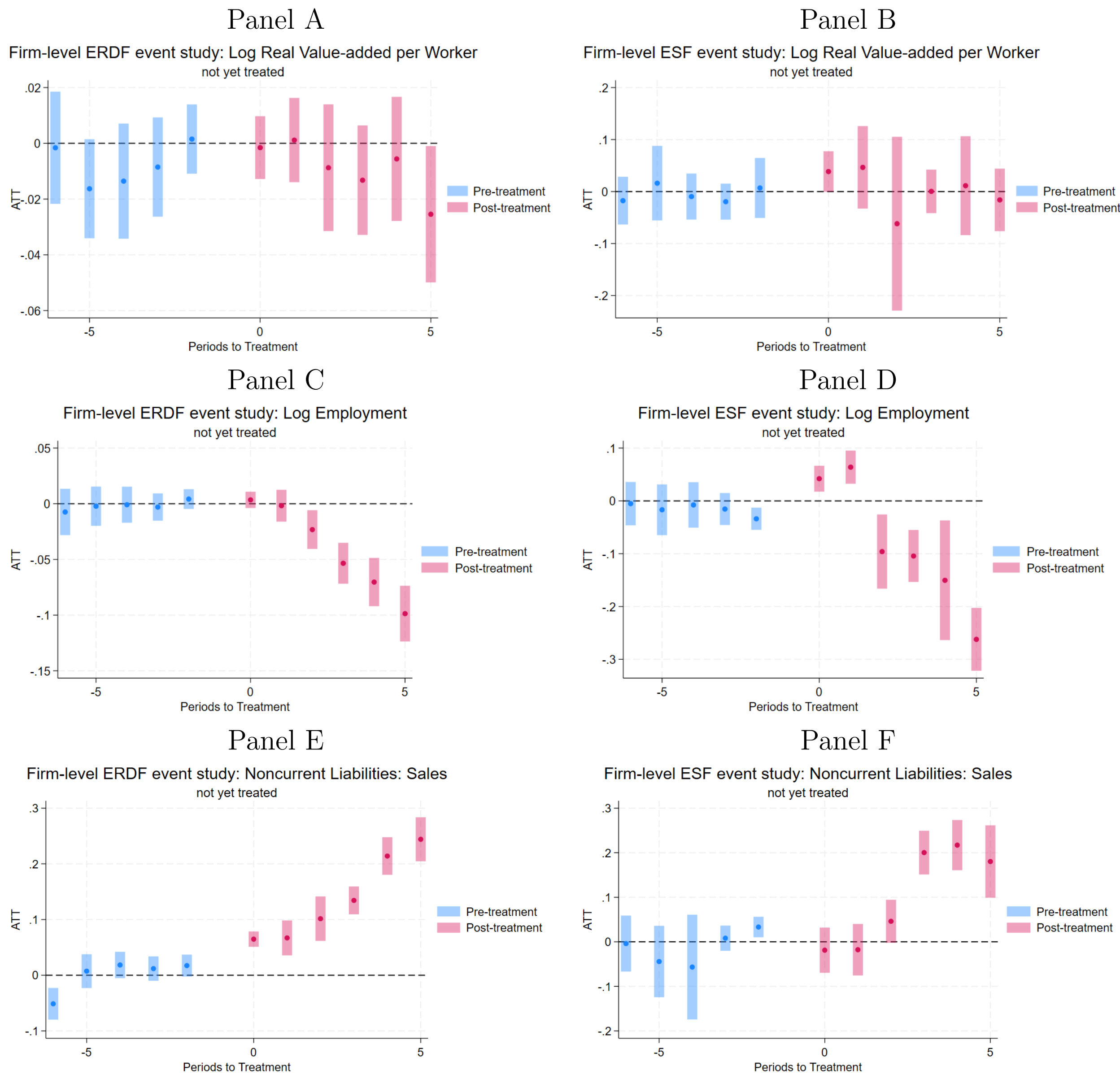
Notes: The sample for actual payments includes projects funded by the ERDF and ESF excluding multi-funding-source projects as described in Appendix B.1. For projects taking place in multiple commuting zones, equal spending across both commuting zones is assumed. Expected payments are calculated using the OpenCoesione data and the simulation procedure described in Section 5. Each dot represents an Italian commuting zone for a given funding period. The red line is a 45 degree line.

Figure 5: Dynamic Selection of Firms in Funding Take-Up



Notes: The dependent variable is log real value-added per worker (Panels A and B) and log number of employees (Panels C and D). All dependent variables are observed at the firm-level. Panels A and C (B and D) refer to projects co-financed from ERDF (ESF). To account for staggered treatment adoption, we use the estimator of Callaway and Sant'Anna (2021). The control group is never-treated firms. The event study time $t = 0$ is the year a project was begun with the private firm in an implementation role ("attuatore" or "beneficiario" in the OpenCoesione data). Control variables are NUTS2-region and 2-digit industry dummies.

Figure 6: Firm-level Event Studies - Not Yet Treated Control



Notes: The dependent variable is log real value-added per worker (Panels A and B), log number of employees (Panels C and D) and the ratio of noncurrent liabilities over sales, capturing liquidity constraints (Panels E and F). All dependent variables are observed at the firm-level. Panels A, C and E (B, D and F) refer to projects co-financed from ERDF (ESF). To account for staggered treatment adoption, we use the estimator of Callaway and Sant'Anna (2021). The control group is not-yet-treated firms. The event study time $t = 0$ is the year a project was begun with the private firm in an implementation role ("attuatore" or "beneficiario" in the OpenCoesione data). Control variables are NUTS2-region and 2-digit industry dummies.

Tables

Table 1: Summary Statistics – Cohesion Policy Projects

Panel A: Project Characteristics

	Funding Period					
	2007-2013			2014-2020		
	Classification of Eligibility			Classification of Eligibility		
	Less Developed	More Developed	Transition	Less Developed	More Developed	Transition
N. projects/1000 persons	4	8	13	3	6	9
Mean Funding commitments/project	195815	37897	101042	193395	37686	41086
Total Funding commitments/cap	851	298	1308	657	244	353
Mean payments/project	154267	35394	89227	115876	33027	28538
Total payments/cap	671	279	1155	394	214	245
Mean Duration (Days)	393	151	412	332	276	440

Panel B: Project Purpose

	Funding Period					
	2007-2013			2014-2020		
	Classification of Eligibility			Classification of Eligibility		
	Less Developed	More Developed	Transition	Less Developed	More Developed	Transition
EU Theme (% of projects)						
Research & Innovation	1.66	3.30	0.91	1.53	3.55	2.77
Digital Networks & Services	2.53	1.32	3.74	1.06	1.02	1.33
Business Competitiveness	12.15	3.71	4.26	47.37	14.04	13.01
Energy	1.25	1.54	2.48	1.33	0.44	1.44
Environment	4.55	0.51	2.44	3.09	0.16	0.48
Culture & Tourism	3.61	0.64	3.68	1.89	0.21	0.66
Transportation	0.84	0.19	0.30	0.55	0.10	0.05
Employment & Work	22.52	71.71	40.09	23.47	64.94	60.97
Social Inclusion & Health	6.12	3.64	6.64	5.04	5.78	9.95
Education & Training	44.25	12.59	31.64	14.20	9.03	9.03
Administrative Capacity	0.52	0.85	3.83	0.46	0.73	0.30
EU Theme (% of payments)						
Research & Innovation	3.57	16.12	2.95	7.06	11.78	12.33
Digital Networks & Services	2.62	2.15	10.73	1.98	2.28	3.68
Business Competitiveness	10.35	7.07	11.37	16.82	16.46	12.32
Energy	2.14	5.78	5.89	3.83	3.18	12.61
Environment	12.21	4.11	8.27	20.76	1.93	8.22
Culture & Tourism	7.66	5.37	7.64	4.52	1.51	7.20
Transportation	16.91	5.00	10.34	13.94	1.47	2.38
Employment & Work	10.81	32.23	13.27	6.59	22.42	20.52
Social Inclusion & Health	13.40	5.11	11.38	8.80	19.47	15.14
Education & Training	18.80	13.67	11.59	11.70	15.63	4.10
Administrative Capacity	1.52	3.37	6.59	4.00	3.86	1.50

Panel C: Project Implementers

	Funding Period					
	2007-2013			2014-2020		
	Classification of Eligibility			Classification of Eligibility		
	Less Developed	More Developed	Transition	Less Developed	More Developed	Transition
Share of Projects						
Government	0.23	0.20	0.52	0.44	0.41	0.43
Firm	0.15	0.52	0.24	0.41	0.56	0.29
Individuals	0.16	0.11	0.14	0.22	0.07	0.29
Other	0.56	0.25	0.18	0.16	0.22	0.05
Share of Payments						
Government	0.45	0.37	0.59	0.58	0.40	0.47
Firm	0.29	0.44	0.36	0.51	0.46	0.41
Individuals	0.02	0.08	0.03	0.18	0.03	0.05
Other	0.31	0.22	0.13	0.26	0.23	0.13

Notes: The sample consists of projects which are funded by a single region's operational program as described in Appendix B.1 and take place within specifiable commuting zones. Observations are aggregated to project initiatives as described in Appendix B.1. Per capita values are calculated using population figures from the first year of the funding period (2007 and 2014, respectively). Project implementers included in Panel C are those with the role "attuatore" or "beneficiario" in the OpenCoesione data, and were classified into categories by the ChatGPT API using the string description of their legal form.

Table 2: Serial Correlation in Regional GDP Estimates

Panel A: Regional GDP

	(1)
	% Error, GDP
Lag % Error, GDP	0.738***
	(0.0108)
Observations	2764
Adjusted R^2	0.629

Panel B: Regional GDP per Capita

	(1)
	% Error, GDP/cap
% Error, GDP	0.864***
	(0.00780)
Lagged % Error, GDP/cap	0.148***
	(0.00719)
Observations	2764
Adjusted R^2	0.932

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

Notes: The sample consists of European NUTS2 regions from 1998 (or the earliest year data is available) to 2009 for all EU member states. The percent error in GDP is the percent difference between the historical GDP estimates (the first estimate) and the revised estimates.

Table 3: Measurement Error and Other Funding Determinants

	(1) % Error, GDP	(2) % Error, GDP/cap
Lag % Error, GDP	0.719***	
	(0.0147)	
Revised GDP (millions)	-0.0173**	
	(0.00782)	
% Error, GDP		0.925***
		(0.00739)
Lagged % Error, GDP/cap		0.109***
		(0.00699)
Revised GDP/cap (thous.)		0.815
		(2.664)
N. Early Leavers 18-24 (thous.)	-0.0239**	0.0112***
	(0.00942)	(0.00321)
N. Tertiary Ed. 30-34 (thous.)	0.000739	0.00469**
	(0.00619)	(0.00211)
N. Employed 15-74 (thous.)	-0.00112	-0.000132
	(0.00143)	(0.000486)
N. Unemp. 15-74 (thous.)	0.00229	-0.00283***
	(0.00275)	(0.000938)
Population (thous.)	0.000584	0.000117
	(0.000547)	(0.000186)
Low Educ. Emps. (thous.)	0.00442***	-0.00234***
	(0.00165)	(0.000561)
Constant	-0.443	-0.330**
	(0.465)	(0.158)
Observations	2001	2001
Adjusted R^2	0.646	0.963
Country FE	Yes	Yes
SD outcome	6.230	6.528

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

Notes: The sample consists of European NUTS2 regions from 1998 (or the earliest year data is available) to 2009 for all EU member states. The dependent variables, the percent error in the GDP (Column 1) and GDP per capita (Column 2), is the percent difference between the historical estimates (the first estimate) and the revised estimates. Control variables are revised variables from Eurostat data releases as described in Appendix B.4.

Table 4: Funding in Counterfactual Simulations

Panel A: 2007-2013 Funding Period

	Classification of Eligibility			
	More Developed	Transition	Less Developed	Total
NUTS2 Name				
Abruzzo	313.06	311.47		313.03
	(30.40)	(10.27)		(30.11)
Basilicata		1114.85	1101.17	1113.14
		(124.92)	(85.93)	(120.79)
Calabria		915.26	1563.61	1561.01
		(8.73)	(148.69)	(153.94)
Campania		736.61	1272.94	1095.41
		(26.77)	(119.63)	(271.23)
Emilia-Romagna	216.41			216.41
	(36.97)			(36.97)
Friuli-Venezia Giulia	269.50			269.50
	(59.09)			(59.09)
Lazio	214.37			214.37
	(34.61)			(34.61)
Liguria	286.77			286.77
	(54.32)			(54.32)
Lombardia	211.74			211.74
	(30.53)			(30.53)
Marche	310.91			310.91
	(33.85)			(33.85)
Molise	314.09	552.40	1052.84	402.98
	(32.68)	(14.48)	(67.78)	(159.14)
Piemonte	259.24			259.24
	(59.39)			(59.39)
Provincia Autonoma Bolzano/Bozen	211.54			211.54
	(30.24)			(30.24)
Provincia Autonoma Trento	212.55			212.55
	(31.81)			(31.81)
Puglia		543.73	1207.42	1098.57
		(14.82)	(131.22)	(273.64)
Sardegna		880.29	1161.81	954.33
		(67.93)	(79.67)	(142.97)
Sicilia		705.32	1303.09	1223.58
		(23.86)	(125.35)	(234.39)
Toscana	280.22			280.22
	(56.83)			(56.83)
Umbria	311.22			311.22
	(33.02)			(33.02)
Valle d'Aosta/Vallée d'Aoste	211.54			211.54
	(30.24)			(30.24)
Veneto	257.03			257.03
	(57.79)			(57.79)
Total	257.43	870.48	1326.71	524.53
	(59.03)	(227.78)	(197.89)	(442.88)

(Continued)

Panel B: 2014-2020 Funding Period

	Classification of Eligibility			
	More Developed	Transition	Less Developed	Total
NUTS2 Name				
Abruzzo	172.54	199.41	698.60	183.94
	(3.11)	(20.84)	(2.11)	(29.58)
Basilicata	173.88	264.93	799.41	366.73
	(2.02)	(28.78)	(63.18)	(214.54)
Calabria		296.33	925.96	900.14
		(15.54)	(105.48)	(162.12)
Campania		322.78	890.62	710.05
		(21.30)	(91.43)	(275.40)
Emilia-Romagna	172.65			172.65
	(3.13)			(3.13)
Friuli-Venezia Giulia	171.88	176.00		171.88
	(3.14)	(.)		(3.14)
Lazio	172.86			172.86
	(3.13)			(3.13)
Liguria	171.96			171.96
	(3.13)			(3.13)
Lombardia	173.83			173.83
	(3.13)			(3.13)
Marche	171.99	189.73		172.29
	(3.13)	(18.56)		(4.52)
Molise	171.90	213.05	772.99	213.11
	(3.17)	(24.68)	(68.68)	(79.13)
Piemonte	172.77			172.77
	(3.13)			(3.13)
Provincia Autonoma Bolzano/Bozen	171.64			171.64
	(3.13)			(3.13)
Provincia Autonoma Trento	171.66			171.66
	(3.13)			(3.13)
Puglia		329.36	957.83	912.58
		(21.18)	(105.69)	(191.87)
Sardegna	177.60	326.04	864.13	393.95
	(2.71)	(30.78)	(65.91)	(188.35)
Sicilia		352.75	943.66	816.02
		(20.70)	(94.57)	(257.47)
Toscana	172.55			172.55
	(3.13)			(3.13)
Umbria	171.80	192.50		172.73
	(3.14)	(17.77)		(6.46)
Valle d'Aosta/Vallée d'Aoste	171.56			171.56
	(3.14)			(3.14)
Veneto	172.86			172.86
	(3.13)			(3.13)
Total	172.32	272.41	921.58	320.85
	(3.20)	(59.68)	(105.55)	(284.56)

Notes: Each cell reports the mean predicted funding pot per capita over the 7-year funding period for each Italian NUTS2 region conditional on being in a particular eligibility category using the measurement error simulation procedure described in Section 5 and Appendix B.4. Standard deviations are in parentheses. Empty cells indicate the region was never in that particular eligibility category in any of the 1,000 simulations.

Table 5: Effects of Cohesion Policy on Aggregate Productivity Growth

	(1) $\Delta\Phi$	(2) $\Delta\Phi$	(3) $\Delta\Phi$	(4) $\Delta\Phi$
Log(Payments)	-0.0688**	-0.0710**	-0.0245	-0.0290
	(0.0279)	(0.0308)	(0.0184)	(0.0210)
Log(E(Payments))	0.0339	0.0500	0.0315	0.0414
	(0.0351)	(0.0433)	(0.0256)	(0.0297)
Observations	1040	1040	1040	1040
Demo. Controls	No	No	Yes	Yes
Other EU Payments Controls	No	Yes	No	Yes

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

Notes: The dependent variable is the change in aggregate productivity at the commuting-zone-by-funding-period level. At the firm-level, productivity is measured as value added per worker. Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5. All specifications include funding period fixed effects and controls for a commuting zone's initial productivity level. Demographic controls include education (share illiterate, literate, elementary school, middle school, and college/university – high school excluded), age (under 14 and over 65, 15-64 excluded), share male, and share foreign. Other EU payments include ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data. Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.

Table 6: Decomposition of Aggregate Productivity Growth Effects

	(1) $\Delta\Phi$	(2) $\Delta\overline{\phi}$	(3) Δcov	(4) Entry	(5) Exit
Log(Payments)	-0.0290	0.0107	-0.0636***	0.0321**	-0.00830
	(0.0210)	(0.0152)	(0.0174)	(0.0116)	(0.00815)
Log(E(Payments))	0.0414	0.000150	0.0694***	-0.0341**	0.00590
	(0.0297)	(0.0177)	(0.0215)	(0.0136)	(0.00883)
Observations	1040	1040	1040	1040	1040

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

Notes: The dependent variables are the change in decomposed components of aggregate productivity using the method of Melitz and Polanec (2015) (Column (1) is the aggregate effect, Column (2) is the symmetric, Column (3) is the reallocation, Column (4) is the entry, and Column (5) is the exit component) at the commuting-zone-by-funding-period level. At the firm-level, productivity is measured as value added per worker. Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5. All specifications include funding period fixed effects and controls for a commuting zone's initial productivity level, demographic characteristics (share illiterate, literate, elementary school, middle school, and college/university – high school excluded – age under 14 and over 65 – 15-64 excluded, share male, and share foreign), and other EU payments (ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data). Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.

Table 7: Regional Productivity Changes - By Sector

Panel A: Manufacturing

	(1) $\Delta\Phi$	(2) $\Delta\overline{\phi}$	(3) Δcov	(4) Entry	(5) Exit
Log(Payments)	-0.0654**	0.00665	-0.0651**	-0.00447	-0.00246
	(0.0290)	(0.0116)	(0.0279)	(0.0101)	(0.00949)
Log(E(Payments))	0.0643*	-0.0108	0.0694**	0.000386	0.00537
	(0.0326)	(0.0111)	(0.0283)	(0.0105)	(0.0126)
Observations	831	831	831	831	831

Panel B: Construction

	(1) $\Delta\Phi$	(2) $\Delta\overline{\phi}$	(3) Δcov	(4) Entry	(5) Exit
Log(Payments)	0.0755***	0.0709**	-0.0130	0.0280**	-0.0103*
	(0.0214)	(0.0319)	(0.0243)	(0.00999)	(0.00502)
Log(E(Payments))	-0.0809***	-0.0473	-0.0121	-0.0332**	0.0116**
	(0.0266)	(0.0377)	(0.0351)	(0.0128)	(0.00521)
Observations	816	816	816	816	816

Panel C: Services

	(1) $\Delta\Phi$	(2) $\Delta\overline{\phi}$	(3) Δcov	(4) Entry	(5) Exit
Log(Payments)	0.0135	0.0119	-0.0212	0.0395**	-0.0167
	(0.0309)	(0.0157)	(0.0171)	(0.0177)	(0.0109)
Log(E(Payments))	0.0211	0.00381	0.0417	-0.0332*	0.00871
	(0.0407)	(0.0194)	(0.0273)	(0.0172)	(0.0134)
Observations	888	888	888	888	888

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

Notes: The manufacturing sector is NACE Rev. 2 sectors B-E, the construction sector is NACE Rev. 2 sector F, and the services sector is NACE Rev. 2 sectors H-S (excluding O). The dependent variables are the change in decomposed components of aggregate productivity using the method of Melitz and Polanec (2015) (Column (1) is the aggregate effect, Column (2) is the symmetric, Column (3) is the reallocation, Column (4) is the entry, and Column (5) is the exit component) at the commuting-zone-by-funding-period level. At the firm-level, productivity is measured as value added per worker. Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5. All specifications include funding period fixed effects and controls for a commuting zone's initial productivity level, demographic characteristics (share illiterate, literate, elementary school, middle school, and college/university – high school excluded – age under 14 and over 65 – 15-64 excluded, share male, and share foreign), and other EU payments (ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data). Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.

Table 8: Alternative Counterfactual Funding Construction

	(1) $\Delta\Phi$	(2) $\Delta\overline{\phi}$	(3) Δcov	(4) Entry	(5) Exit
Log(Payments)	-0.0278	0.0108	-0.0627***	0.0303**	-0.00629
	(0.0197)	(0.0140)	(0.0165)	(0.0109)	(0.00825)
Log(E(Payments))	0.0401	0.0000361	0.0684***	-0.0320**	0.00354
	(0.0282)	(0.0167)	(0.0209)	(0.0125)	(0.00871)
Observations	1040	1040	1040	1040	1040

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

Notes: The dependent variables are the change in decomposed components of aggregate productivity using the method of Melitz and Polanec (2015) (Column (1) is the aggregate effect, Column (2) is the symmetric, Column (3) is the reallocation, Column (4) is the entry, and Column (5) is the exit component) at the commuting-zone-by-funding-period level. At the firm-level, productivity is measured as value added per worker. Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5 taking into account the additional economic characteristics of the region as described in Section 5.2.3. All specifications include funding period fixed effects and controls for a commuting zone's initial productivity level, demographic characteristics (share illiterate, literate, elementary school, middle school, and college/university – high school excluded – age under 14 and over 65 – 15-64 excluded, share male, and share foreign), and other EU payments (ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data). Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.

Table 9: Results - Total Factor Productivity

	(1) $\Delta\Phi$	(2) $\Delta\overline{\phi}$	(3) Δcov	(4) Entry	(5) Exit
Log(Payments)	0.00938	0.0166*	-0.0119**	0.00425*	-0.00123
	(0.00726)	(0.00846)	(0.00440)	(0.00244)	(0.00220)
Log(E(Payments))	-0.00526	-0.0103	0.00774	-0.00389	0.00212
	(0.00899)	(0.00952)	(0.00555)	(0.00250)	(0.00257)
Observations	1030	1030	1030	1030	1030

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

Notes: The dependent variables are the change in decomposed components of aggregate productivity using the method of Melitz and Polanec (2015) (Column (1) is the aggregate effect, Column (2) is the symmetric, Column (3) is the reallocation, Column (4) is the entry, and Column (5) is the exit component) at the commuting-zone-by-funding-period level. At the firm-level, productivity is measured as Total Factor Productivity calculated using the estimator of Wooldridge (2009). Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5. All specifications include funding period fixed effects and controls for a commuting zone's initial productivity level, demographic characteristics (share illiterate, literate, elementary school, middle school, and college/university – high school excluded – age under 14 and over 65 – 15-64 excluded, share male, and share foreign), and other EU payments (ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data). Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.

Table 10: Heterogeneity: Funding Absorption

Panel A: Reallocation Effect

	(1)	(2)	(3)
	Δcov	Δcov	Δcov
Log(Payments)	-0.0610***	-0.0652***	-0.0330
	(0.0163)	(0.0186)	(0.0306)
Payments x MoreDev	0.0122**		
	(0.00566)		
Log(E(Payments))	0.0586***	0.0702***	0.0341
	(0.0205)	(0.0227)	(0.0328)
Observations	1040	911	627
Excluding Sicily	No	Yes	No
Commitment Condition	No	No	Yes

Panel B: Entry Effect

	(1)	(2)	(3)
	Entry	Entry	Entry
Log(Payments)	0.0296**	0.0371***	0.00362
	(0.0113)	(0.0124)	(0.0204)
Payments x MoreDev	-0.00786**		
	(0.00347)		
Log(E(Payments))	-0.0253*	-0.0383**	-0.00125
	(0.0139)	(0.0147)	(0.0252)
Observations	1040	911	627
Excluding Sicily	No	Yes	No
Commitment Condition	No	No	Yes

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

Notes: The dependent variables are the change in the reallocation component (Panel A) and the entry component (Panel B) of aggregate productivity using the method of Melitz and Polanec (2015) at the commuting-zone-by-funding-period level, respectively. At the firm-level, productivity is measured as value added per worker. Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. MoreDev denotes more developed status (regional GDP per capita > 90% of EU average). The commitment condition in Column (3) excludes regions which committed less than 90% of their total funding allocation (pooled ERDF and ESF) during a funding period. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5. All specifications include funding period fixed effects and controls for a commuting zone's initial productivity level, demographic characteristics (share illiterate, literate, elementary school, middle school, and college/university – high school excluded – age under 14 and over 65 – 15-64 excluded, share male, and share foreign), and other EU payments (ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data). Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.

Table 11: Splitting by Funding Period

Panel A: 2007-2013 Period

	(1)	(2)
	Δcov	Entry
Log(Payments)	-0.0457**	0.0238
	(0.0171)	(0.0163)
Log(E(Payments))	0.0553**	-0.0150
	(0.0254)	(0.0211)
Observations	506	506

Panel B: 2014-2020 Period

	(1)	(2)
	Δcov	Entry
Log(Payments)	-0.0935***	0.0395***
	(0.0242)	(0.0102)
Log(E(Payments))	0.0959***	-0.0435***
	(0.0272)	(0.00980)
Observations	534	534

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

Notes: The dependent variables in Columns (1) and (2) are the change in the reallocation component and the entry component of aggregate productivity using the method of Melitz and Polanec (2015) at the commuting-zone-by-funding-period level, respectively. At the firm-level, productivity is measured as value added per worker. Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5. All specifications include controls for a commuting zone's initial productivity level, demographic characteristics (share illiterate, literate, elementary school, middle school, and college/university – high school excluded – age under 14 and over 65 – 15-64 excluded, share male, and share foreign), and other EU payments (ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data). Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.

Table 12: Alternative Time Span

	(1) $\Delta\Phi$	(2) $\Delta\overline{\phi}$	(3) Δcov	(4) Entry	(5) Exit
Log(Payments)	-0.0288	-0.00140	-0.0303*	0.0279**	-0.0250*
	(0.0284)	(0.0159)	(0.0146)	(0.0101)	(0.0140)
Log(E(Payments))	0.0341	0.0196	-0.00510	-0.0102	0.0298*
	(0.0333)	(0.0196)	(0.0233)	(0.0139)	(0.0161)
Observations	506	506	506	506	506

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

Notes: The dependent variables are the change in decomposed components of aggregate productivity using the method of Melitz and Polanec (2015) (Column (1) is the aggregate effect, Column (2) is the symmetric, Column (3) is the reallocation, Column (4) is the entry, and Column (5) is the exit component) between 2006 and 2016 at the commuting zone level. At the firm-level, productivity is measured as value added per worker. Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5. All specifications include controls for initial productivity level, demographic characteristics (share illiterate, literate, elementary school, middle school, and college/university – high school excluded – age under 14 and over 65 – 15-64 excluded, share male, and share foreign), and other EU payments (ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data). Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.

Table 13: Regional Labor Market Responses

Panel A: Unemployment Rates

	(1)	(2)	(3)	(4)	(5)
	$\Delta URate$	$\Delta URate$	$\Delta URate$	$\Delta URate$	$\Delta URate$
Log(Payments)	1.006**	1.317**	0.841	0.912*	0.880
	(0.478)	(0.470)	(0.534)	(0.471)	(0.537)
Log(E(Payments))	-0.666	-1.222*	-0.723	-0.950	-0.734
	(0.563)	(0.653)	(0.654)	(0.601)	(0.711)
Payments x MoreDev					-0.328***
					(0.113)
Observations	1040	1040	1040	1040	1040
Demo. Controls	No	No	Yes	Yes	Yes
Other EU Payments Controls	No	Yes	No	Yes	Yes

Panel B: Labor Force Participation Rates

	(1)	(2)	(3)	(4)	(5)
	$\Delta LFPR$	$\Delta LFPR$	$\Delta LFPR$	$\Delta LFPR$	$\Delta LFPR$
Log(Payments)	-0.0493	-0.199	-0.385	-0.375	-0.417
	(0.224)	(0.206)	(0.258)	(0.268)	(0.267)
Log(E(Payments))	-0.109	0.0840	0.275	0.233	0.320
	(0.258)	(0.264)	(0.353)	(0.360)	(0.366)
Payments x MoreDev					-0.0178
					(0.0954)
Observations	1040	1040	1040	1040	1040
Demo. Controls	No	No	Yes	Yes	Yes
Other EU Payments Controls	No	Yes	No	Yes	Yes

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

Notes: The dependent variables are the change in unemployment rates ($URate$) and labor force participation rates ($LFPR$) at the commuting-zone-by-funding-period level, respectively. Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5. All specifications include funding period fixed effects and controls for initial productivity level, demographic characteristics (share illiterate, literate, elementary school, middle school, and college/university – high school excluded – age under 14 and over 65 – 15-64 excluded, share male, and share foreign), and other EU payments (ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data). Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.

Table 14: Selection of Firms into Funding

Panel A: ERDF

	Never-Treated Mean	Treated Mean	Mean Difference	Mean Difference (Controls)	P-value (Controls)
Firm age	12.574	15.348	2.775	2.131	0.000
Log Employment	1.594	2.415	0.821	0.713	0.000
Log Sales	12.853	13.911	1.059	1.035	0.000
Log Revenue per Employee	11.661	11.775	0.114	0.223	0.000
Log Value-Added per worker	10.470	10.645	0.175	0.211	0.000
Public limited companies (0/1)	0.031	0.107	0.076	0.071	0.000
North Italy (0/1)	0.489	0.422	-0.067	.	.
Manufacturing (0/1)	0.198	0.422	0.224	.	.
Construction (0/1)	0.155	0.091	-0.063	.	.
Retail (0/1)	0.244	0.133	-0.111	.	.

Panel B: ESF

	Never-Treated Mean	Treated Mean	Mean Difference	Mean Difference (Controls)	P-value (Controls)
Firm age	12.771	12.597	-0.174	0.673	0.000
Log Employment	1.620	2.435	0.815	0.843	0.000
Log Sales	12.888	13.753	0.864	1.071	0.000
Log Revenue per Employee	11.674	11.547	-0.127	0.081	0.000
Log Value-Added per worker	10.483	10.451	-0.032	0.092	0.000
Public limited companies (0/1)	0.034	0.086	0.052	0.058	0.000
North Italy (0/1)	0.493	0.241	-0.253	.	.
Manufacturing (0/1)	0.209	0.268	0.058	.	.
Construction (0/1)	0.152	0.112	-0.040	.	.
Retail (0/1)	0.239	0.190	-0.050	.	.

Notes: The sample of firms is those in the Orbis analysis sample after making the restrictions described in Section 3. Column (4) includes region, 2-digit NACE, and year-of-treatment fixed effects. Column (5) shows the p-value for the coefficient printed in Column (4).

Table 15: Treatment Timing and Initial Firm-Level Characteristics

	(1) ERDF	(2) ESF
Firm age	0.0132***	0.0156***
	(0.00154)	(0.00213)
Log Employment	-0.375***	-0.357***
	(0.0632)	(0.0927)
Log Sales	-0.171***	0.102
	(0.0594)	(0.0895)
Log Real Operating Revenue per Employee	-0.611***	-0.584***
	(0.0684)	(0.0963)
Log Real Value-added per Worker	0.320***	0.324***
	(0.0376)	(0.0431)
Public limited companies (0/1)	-0.467***	-0.514***
	(0.0714)	(0.0949)
Observations	33884	18375

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

Notes: The dependent variable is the year of receipt of ERDF (Column 1) or ESF (Column 2) funding. The sample of firms is those in the Orbis analysis sample after making the restrictions described in Section 3 who received funding. Both specifications include region and 2-digit NACE fixed effects. All independent variables are measured in the year prior to beginning the ERDF or ESF projects, $t = -1$.

A Large-Scale Projects and Local Productivity

The preceding analysis suggests that cohesion policy does not generate significant growth in firm-level productivity and, on aggregate, may even reduce productivity growth through lower allocative efficiency. These findings raise concerns about the overall effectiveness of the policy in promoting structural transformation. However, average effects may conceal more targeted successes. In this section, we explore whether a specific subset of investments – so-called "million-euro projects" – can generate localized productivity gains consistent with the "big push" theory of regional development (Kline, 2010; Greenstone et al., 2010; Kline and Moretti, 2013; Cerrato, 2024). The underlying idea is that sufficiently large public investments may overcome local market failures, trigger (further) productivity growth through agglomeration spillovers, and generate welfare gains.

To test this hypothesis, we exploit staggered treatment adoption (Callaway and Sant'Anna, 2021) and compare average firm-level productivity in commuting zones that experience the completion of a large-scale project ($\geq$ €1,000,000) earlier in the sample period to later in the sample period. During our period of observation, about 2500 million euro projects have been carried out ranging from transport or tourism infrastructure projects to R&D investments.

There are many reasons to expect that areas completing a large-scale project during the sample period differ systematically from those that do not. Most notably, not every commuting zone may have the administrative or political capacity to attract and execute such large investments – indeed, in approximately 40% of commuting zones, no million-euro project is ever completed. However, there is reason to believe that the *timing* of treatment – that is, the year in which a commuting zone completes its first million-euro project – is plausibly exogenous.[42] First, large-scale projects, particularly those involving physical infrastructure and construction, are often subject to delays driven by exogenous factors such as procurement issues, weather, or legal disputes. In our data, we are able to observe if the planned completion date of the project aligns with the actual completion date of the project. In approximately 53% of cases, the two dates do not coincide, suggesting that the finalization of projects is frequently influenced by idiosyncratic shocks. Second, we are able to explicitly test whether commuting zones with particular characteristics are systematically treated earlier or later by regressing the year of treatment on the characteristics of the commuting zone from the 2001 census. The results are shown in Table A.0.1. None of the coefficients are statistically significant with the exception of population, suggesting that areas with particular demographic characteristics are not systematically treated earlier or later.

We estimate the commuting-zone-level impact of million-euro project completions on average firm productivity using the following empirical specification:

$$Y_{ct} = \alpha_c + \delta_t + \sum_{\substack{i=-k \\ i\neq -1}}^{k} \gamma_i \times 1(t = \text{Year Treatment}_c + i) + x_0 + \epsilon_{ct} \tag{11}$$

[42] In some years, several million-euro projects are completed within the same commuting zone, especially in larger areas. We treat those as one event. When multiple such projects are completed in different years, we define the treatment year as the year the first million-euro project was completed.

where α_c are commuting zone fixed effects, δ_t are year fixed effects, and $YearTreatment_c$ is the year the area experienced its first large-scale project completion. We define Y_{ct} as the unweighted average of the log of firm-level productivity within commuting zone c and year t, where productivity is measured as either log value-added per worker or log TFP – analogous to the first component of the Dynamic Olley-Pakes Decomposition (shown in Column (2) of Table 6). We control for the initial characteristics in the commuting zone, x_0. Our coefficients of interest are γ_i, corresponding to the effect of project completions i years after a commuting zone completed its first large-scale project.

The results for both measures of productivity are shown Panels A and B of Figure A.0.1, for value added per worker and TFP, respectively. The event studies corroborate our earlier finding that cohesion policy does not induce substantive increases in value-added per worker or TFP – even when focusing on a subset of large-scale projects that are arguably the most likely to generate firm-level productivity gains.

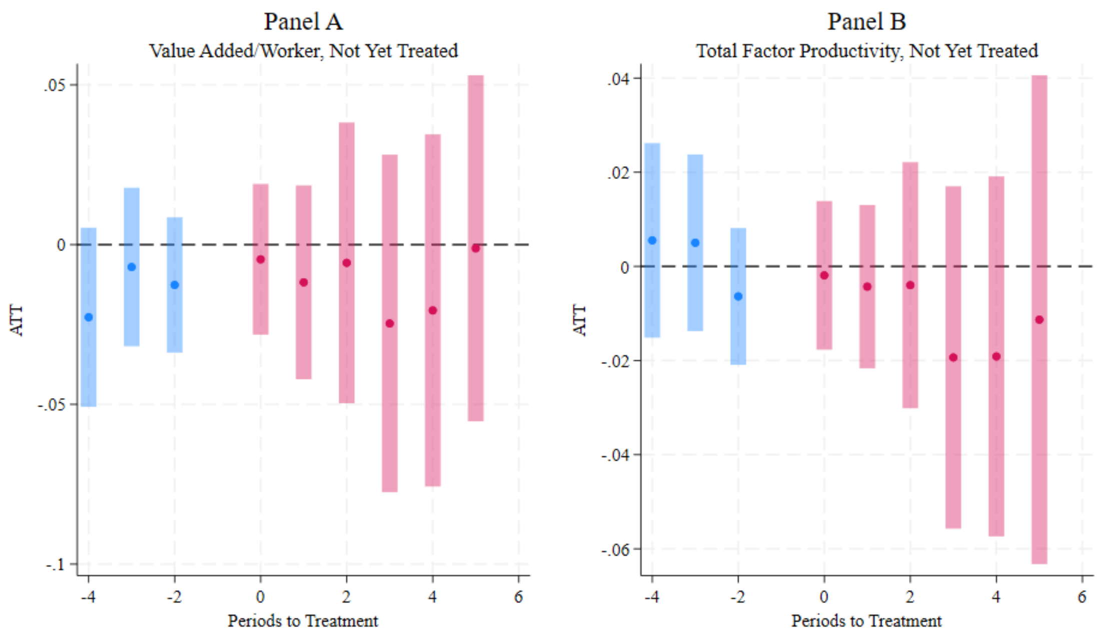
Table A.0.1: Treatment Timing and Initial Commuting-zone-level Characteristics

	(1)
	Year of initial treatment
Population	-1.219***
	(0.250)
Sh. male	-24.76
	(38.16)
Sh. pop. less 14	16.96
	(20.62)
Sh. pop. over 65	7.080
	(15.31)
Sh. foreign born	18.79
	(11.99)
Unemployment Rate	3.611
	(14.06)
Sh. Illiterate	-17.10
	(21.32)
Sh. Elem. School	1.702
	(10.46)
Sh. Middle School	-1.596
	(10.65)
Sh. College/Uni	-12.78
	(24.32)
Observations	318

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

Notes: The dependent variable is the first year of completion of an ERDF or ESF initiative with one million Euros or more in payments. Within an initiative, the first completion date of a sub-project is used as the completion date. Control variables are from the 2001 census for commuting zones treated in the 2007-2013 funding period, and from the 2011 census for commuting zones treated in the 2014-2020 funding period.

Figure A.0.1: Local Productivity Effects of Large-Scale Projects



Notes: The dependent variable is the average of firm-level productivity within commuting zone and year. Productivity is measured as log value-added per worker (in Panel A), or log TFP (in Panel B). To account for staggered treatment adoption, we use the estimator of Callaway and Sant'Anna (2021). The control group is not-yet-treated commuting zones. The event study time $t = -1$ is the year the first million-euro project was completed in a commuting zone. Controls included are NUTS2 region fixed effects, population, unemployment rates, demographic characteristics of residents (share male, share foreign, education levels, and age structures), and payments to the commuting zone from all other EU sources.

B Data Appendix

B.1 Project Data

Data on EU co-financed projects in Italy come from the OpenCoesione database. OpenCoesione is a national open data initiative managed by the Department for Cohesion Policy. Data on *all* projects undertaken under cohesion policy are covered and published on the portal. All datasets are at the project level. The main dataset contains information on projects, including their basic financial characteristics, milestone dates, location, information about the major actors, and more. We supplement this information with additional datasets containing further details concerning project actors, payments, and project dates.

Financial Information The main project file contains detailed breakdowns of funding commitments from each source (i.e. the EU Structural Funds, matched funding from the Italian government, private funding, etc). The variables specifying the commitment value from EU Structural Funds are *oc_finanz_ue_fesr_netto* and *oc_finanz_ue_fse_netto* for the ERDF and the ESF, respectively. National co-financing commitments are given by the variables *oc_finanz_stato_fondo_rot_netto* and *oc_finanz_stato_altri_prov_netto*.

However, funding commitments do not represent actual expenditures. Payments in the project file and supplemental payments file (which provides the exact dates of project payments and some information concerning funding source) do not specify the percentage from the EU funds and the national matches. In order to construct a measure of actual payments from the EU and the national government, we assume that the percentage of payments from each source is the same as the percentage of funding commitments from that source. The percentage is calculated using the variable *oc_finanz_tot_pub_netto* in combination with the above specified commitment variables, as there are other sources of funding aside from the EU and national funds which are accounted for in the data.

Funding Source Restrictions We restrict our sample to projects funded by a single region's Operational Program (POR). This excludes national funds such as the ESF inclusion fund, as well as programs funded by multiple regional PORs. We exclude the latter because we cannot determine how much funding came from each source due to data limitations. These projects represent 27.3% of projects in the ERDF and ESF databases, but only 4.73% of payments.

Location The project data also provide information on project location at the municipal, provincial, commuting zone, and regional level.[43] There are several types of projects, those that take place in a single location (specified as a municipality, province, or region) and those which take place in multiple locations (i.e. two municipalities). In the case of multi-location projects, we do not have information about the breakdown of spending between locations as the finance variables relate to the entire project. In such cases, we assume that the funds are split equally between each project location. In our POR-restricted sample, these projects

[43]Even if an actor such as a firm is in a different location, location provided in the project data concerns the location of the project.

are between 1.02% and 3.36% of the total number of projects, and between 5.84% and 7.99% of payments depending on the fund and funding period.

Initiatives For our summary statistics and million-euro-project event studies, we collapse individual projects into project "initiatives" – or a set of projects that are part of one larger initiative. For example, a scholarship fund that awards multiple scholarships, where each scholarship is one observation in the project data. Or a municipal transportation project where one line in the data is buying new buses, and another line is electrifying existing buses. To identify such initiatives, we flag projects in the data that take place in the same municipality, have the same starting date, have the same string flags for project "type", and the same VAT tax identifiers for planners and executors. As a second step to ensure the quality of our algorithm, we check that those projects the algorithm flags as being part of the same "initiative" have similar titles.

Actors The main project dataset only contains information on up to six actors, so in order to obtain full information on the actors needed for the firm-level event studies we supplemented the main data with the "subjects" data file available on OpenCoesione. The subjects file also contains details on actors involved in the project such as their Codice Fiscale, address, and ATE3 industry. We also have information about each actors' role in the project (planner, implementer, beneficiary, creator). We do not observe the "level" of involvement of each actor beyond their role. For example, if we observe two construction companies as beneficiaries, we cannot say which did more work for the project.

Date In the million-euro-project event study analyses, we use the effective and projected execution dates available in the project data to calculate project delays. Specifically, the variables used are *data_fine_eff_esecuzione* and *data_fine_prev_esecuzione*. For the firm-level event studies, we use the year the project was launched (*oc_data_inizio_progetto*). During the process of cleaning the data, we found that a number of the date variables were mislabelled in the project-level dataset metadata, which we verified using the sequence of dates from the supplemental dates dataset. The correct dates are the ones indicated by logic of the variable name rather than the label in the metadata file.

B.2 Orbis

B.2.1 Sample composition

The financial and balance sheet information provided in the Orbis database comes from national business registers. While in certain countries, legal and administrative filing requirements are dependent on firm size, filing requirements in Italy depend on the legal form (see Kalemli-Özcan et al. (2024) for more details).

Legal form In Italy, Orbis data mainly covers companies such as private limited companies, public limited companies, foreign companies and partnerships. However, Orbis also includes information public authorities, non profit organizations as well as sole traders/ proprietorship. For the purposes of our analysis, we concentrate on firms (in the strict sense) and therefore drop the latter three categories. We also exclude branches since Orbis does not contain their balance sheet information.

Location Orbis contains information on firm locations which we map to a commuting zone. We therefore exclude firms for which information on location at the municipal level is missing and can not be deduced from the other available variables (recall that Italy consists of 20 regions (NUTS2), 107 provinces (NUTS3), 610 commuting zones and approximately 8000 municipalities).[44] Specifically, we use five variables to determine a firm's location: municipality name, zip code, name of province, region and macro-region. In cases where only the municipality name is missing but the zip code is consistent with the regional information, we deduce the municipality name from the zip code for the largest cities in every Italian region. When we only have the municipality name, we assume that the given information is correct. In some cases, we are able to confirm that the location is consistent with the firm's tax identifier (more below).

B.2.2 Financials

Double counting To avoid double counting, we drop any consolidated accounts of company-headquarters whenever the company also presents an unconsolidated account.

Duplicate reports Certain firms file more than one report or various types of reports in one year. We follow Kalemli-Özcan et al. (2024) to tackle duplicate reports. First, we keep only reports that refer to an entire year and drop quarterly reports. Second, if there are several reports for the same calendar year, we keep the one with the closing date that is conceptually closest to the end of the calendar year.[45] Occasionally, a company is reported twice in the same year under different consolidation codes, but the balance sheet data would coincide. In such cases, we keep the report type that has the longer time-series and keep the unconsolidated account in case of a draw. When the two files are very similar but not identical, we keep the report that contains more information. Finally, a few remaining duplicates are due to the fact that a firm might file both an annual report and a local registry file. We again keep the longer time-series and prioritize local registry filing over annual reports.

Data cleaning Certain reports contain obvious mistakes. We again follow the procedure proposed by Kalemli-Özcan et al. (2024) (steps 3 to 10 in section A.5.3 of their Online Appendix), adapting their recommendations to the Italian context. For instance, we drop firms that report number of employees greater than Italy's largest employer, Poste Italiane ($<$ 200,000 employees).

[44] There are only three cases where a commuting zone exactly corresponds to a province.

[45] Following the procedure suggested in Kalemli-Özcan et al. (2024) (Section A.5.1 of their online Appendix), we assign the balance sheet to the previous calendar year if the date is any day before June 1. Specifically, a report filed on May 31 would be the one closest to the end of the calendar year.

Accounting equalities or approximations Most reports contain at least some missing values. Whenever possible, we fill missing information using accounting equalities or approximations based on the variables available in the data. For instance, we compute Earnings before interest, taxes, depreciation, and amortization (EBITDA) following its definition: *EBITDA = operating revenue - cost of goods sold - other operating expenses + depreciation.*

Longitudinal imputation The balance sheet data exhibits irregular reporting patterns, where information is not consistently available for all years. This is due to two main factors: first, firms' balance sheet information is often available only years after their establishment, and second, there are instances of non-consecutive reporting, where data for certain years is missing. We handle missing data for years within the range of observed reporting through imputation using linear interpolation, but we do not extrapolate information beyond the first or last year of available balance sheet data for a firm.

B.2.3 Firm entry and exit

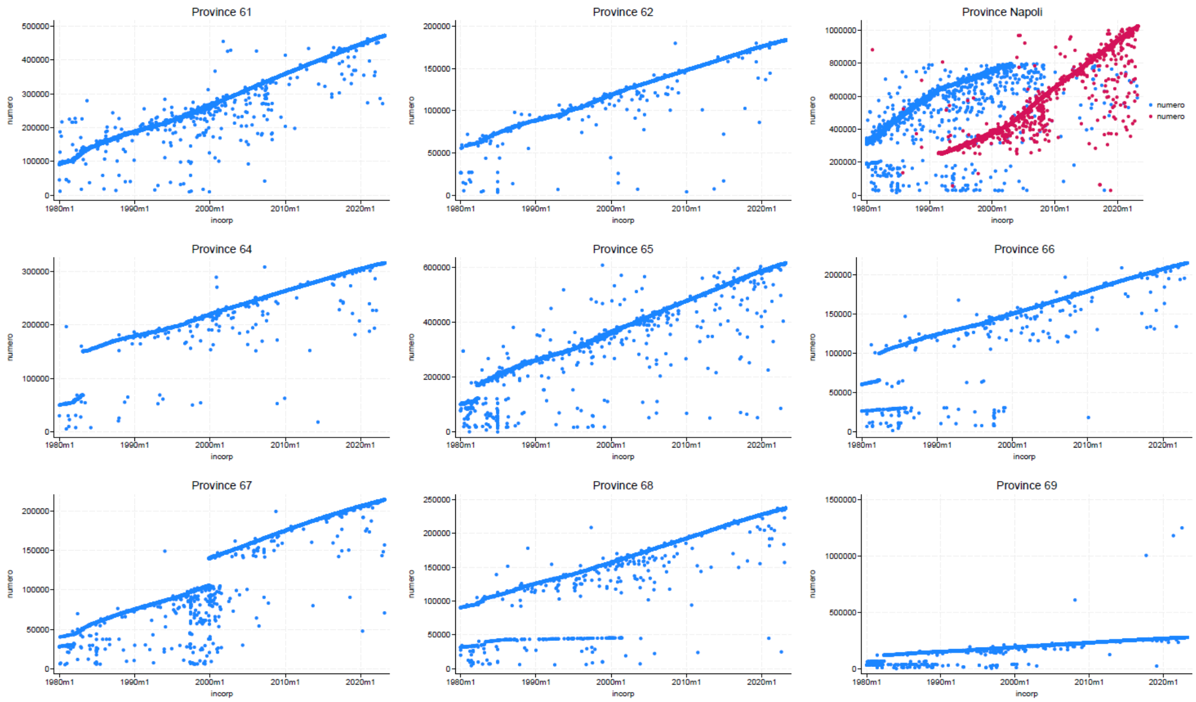
Orbis data contains information on the date of incorporation of the firm as well as information on its status indicating whether a firm is still active. In the case that date of incorporation is missing, we are able to very precisely impute it based on the fiscal code (partita IVA). By systematically integrating status updates and financial data, we can determine or approximate a firm's active period and exit date.

Firm entry For about half a million firms for which Orbis reports legal information, the date of incorporation is unknown. In most cases, we are able to impute the date of incorporation based on the fiscal code (partita IVA) of a firm making use of the fact that the Italian business register assigns the codes in a sequential manner. Specifically, a firm's fiscal code is an 11-digit code composed of a 7-digit number, followed by 3 digits that identify the province where the firm is located, and finally a control number. Within a province, the 7-digit number is linearly increasing over time. In other words, the younger a firm, the higher its 7-digit number. Figure B.2.1 shows the linear trend relationship between the 7-digit number and the date of incorporation.

Concretely, we regress the date of incorporation (month and year) on the 7-digit number by province and use the two province-specific coefficients from the linear regression to impute the day of incorporation for firms where the information is missing. We take into account that the provinces of Milano, Firenze, Brescia, Genova, Roma, Bologna and Napoli each have a discontinued province code and a new province code. The correlation between predicted date of incorporation and true date of incorporation (whenever available) is 98.6 %. We correctly predict the year of incorporation for 64% of firms and standard error of the prediction error is 1.7 years. In the final firm-level sample, we use the imputed entry date for about 500 firms only (0.04% of the sample).

Firm exit Orbis provides information on the status of firms, indicating whether they are active, filing for bankruptcy, in liquidation, inactive, or dissolved. Typically, Orbis vignettes

Figure B.2.1: Date of Incorporation and Fiscal Code



Notes: Scatter plots of firms' 7-digit number that is part of the fiscal code and date of incorporation ($\geq$ 1980) reported in Orbis for provinces 61-69. The province of Naples has two province codes (63 shown in blue and 121 shown in red).

include only the most recent status update and the corresponding date. To enhance this information, we supplemented it with the full history of status updates from BvD. We define a firm's exit year as follows:

1. **Firms with Dissolved or Inactive status**: We assume firms exited in the year their status was first set to "dissolved" or "inactive".

2. **Firms with Liquidation or Bankruptcy status**: For firms with a (latest) status of "liquidation" or "bankruptcy", we assume the firm exited in the year of its last available balance sheet, provided two conditions are met:

   (a) the gap between the last status update and the balance sheet date is no more than two years, and

   (b) employment is zero and either turnover or operating revenue is exactly zero.

   If multiple years satisfy these criteria, we assign the earliest year as the exit year.

If a firm fulfills both criteria, we assume a firm exited in the year of its last available balance sheet.

Active firms Using Orbis data, we identify the years during which a firm was actively operating. A firm is classified as still active based on the most recent status update, exit year, or evidence of activity from financial indicators. Specifically, the active years are determined as follows:

1. **Firms Marked as "Active"**: If the latest status is "Active", the firm is considered active up to the most recent status update or the last year for which data (e.g., financial statements) are available.

2. **Firms with an Exit Year**: For firms with a defined exit year, their last active year is recorded as the year prior to exit.

3. **Firms with Liquidation or Bankruptcy Status**: For firms marked as "in liquidation" or "bankruptcy", we check whether their latest financial information suggests continued operations. Specifically, the firm is assumed active until the last observed year if financial indicators (operating revenue, turnover, and employment) are all positive and the last update is within two years of the last financial record. If multiple years satisfy these conditions, the most recent year is selected as the last active year.

B.2.4 Firm-level productivity

Value-added per worker For our main productivity measure, value-added per worker, we use the value-added variable reported in Orbis. Where missing, we compute or approximate value-added as follows:

1. *value-added = profit for period + depreciation + taxation + interests paid + cost of employees*

2. *value-added* ≈ *cost of employees + EBITDA*

3. *value-added* ≈ *turnover - material costs* (when weakly positive)

4. *value-added* ≈ *sales - material costs* (when weakly positive)

Total Factor Productivity As robustness check, we calculate TFP as alternative productivity measure. We apply the estimator of Wooldridge (2009) and run the regressions separately for 2-digit industries.[46] We use nominal log value-added as the outcome variable. To measure capital, we sum tangible and intangible fixed assets.

Trimming measures of productivity To exclude outliers and since productivity is hard to measure, we trim the bottom and top 5 % of real added-value and log TFP, pooling the years 2000-2022 and all industries. After trimming, the distribution of firm-level labor productivity shown in Appendix Figure C.1.2 exhibits the characteristic shape of a Pareto distribution, with a heavy right tail indicative of a small number of highly productive firms coexisting with a larger number of less productive ones.

[46] We regroup firms in small industries. For instance, we consider the 2-digit industries 10, 11 and 12 as one industry.

B.3 Comparison of Orbis to Alternative Data Sources

In this section, we compare the coverage of Orbis to alternative aggregate sources of data about firms and establishments available for Italy in order to validate the breadth and quality of Orbis in the Italian context. We will first compare the Orbis database to Registro Aziende, a commercial website listing all active firms in Italy. We also assess Orbis coverage using alternative aggregate data sources available from the Italian Government. Specifically, we will use Istat's Registro Statistico delle Imprese Attive (ASIA) as well as data from MovImprese (`https://www.infocamere.it/movimprese`).

Recall that the financial and balance sheet information provided in the Orbis database comes from national business registers. While in certain countries, legal and administrative filing requirements are dependent on firm size, filing requirements in Italy depend on the legal form (see Kalemli-Özcan et al. (2024) for more details). Thus, the expected level of coverage for Italy for smaller firms is relatively high – if firms are required to register with the business register, we expect them to appear in the Orbis data.

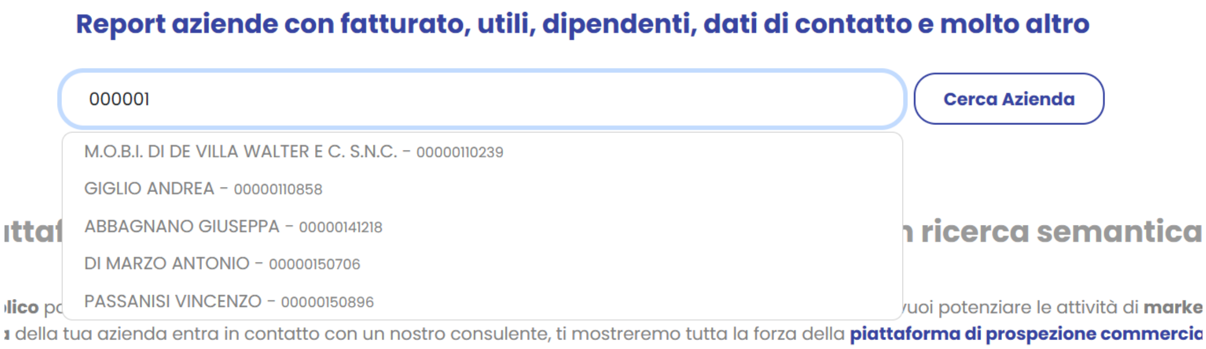
B.3.1 Registro Aziende Comparison

Registro Aziende (`https://registroaziende.it/`) is a commercial website with an associated propriety platform designed to assist commercial enterprises in marketing activities, searching for suppliers, etc. The public website provides basic information about companies from the Italian Business Register. In order to find information about a company on the website, you input their VAT identifier – their partita IVA (PIVA) code – into the search box, and the platform returns basic information about the company such as its name.

In order to use this as a basis of comparison for the Orbis data, we queried the REST API associated with the search box using Python. The REST API returns up to five PIVAs at a time, along with the name(s) of the companies associated with those tax identifiers. We implemented the following basic algorithm, with further details given below:

1. Beginning from the lower bound of possible PIVAs, query the REST API.
2. If the API returns five results:
   (a) Record the PIVAs, name of the company, and slugs pointing to the detailed company information pages on the Registro Aziende website.
   (b) Begin again from step (1) using the next possible valid PIVA code after the fifth code returned
3. If the API returns four or fewer results:
   (a) Record the PIVAs, name of the company, and slugs pointing to the detailed company information pages on the Registro Aziende website.
   (b) Begin again from step (1) using the next possible valid PIVA code after the final code returned

Figure B.3.1: Visual Example of REST API Queries



Note: Screenshot of the searchbox from the website `https://registroaziende.it/`.

We carried out this process between October 31, 2023 and November 22, 2023 on four computers.[47] As the Registro Aziende platform is designed to assist with marketing, they only include companies on the platform which are to the best of their knowledge currently active, so the data can be understood as a snapshot of active firms in Italy as of November 2023 in the range of PIVAs we were able to query.

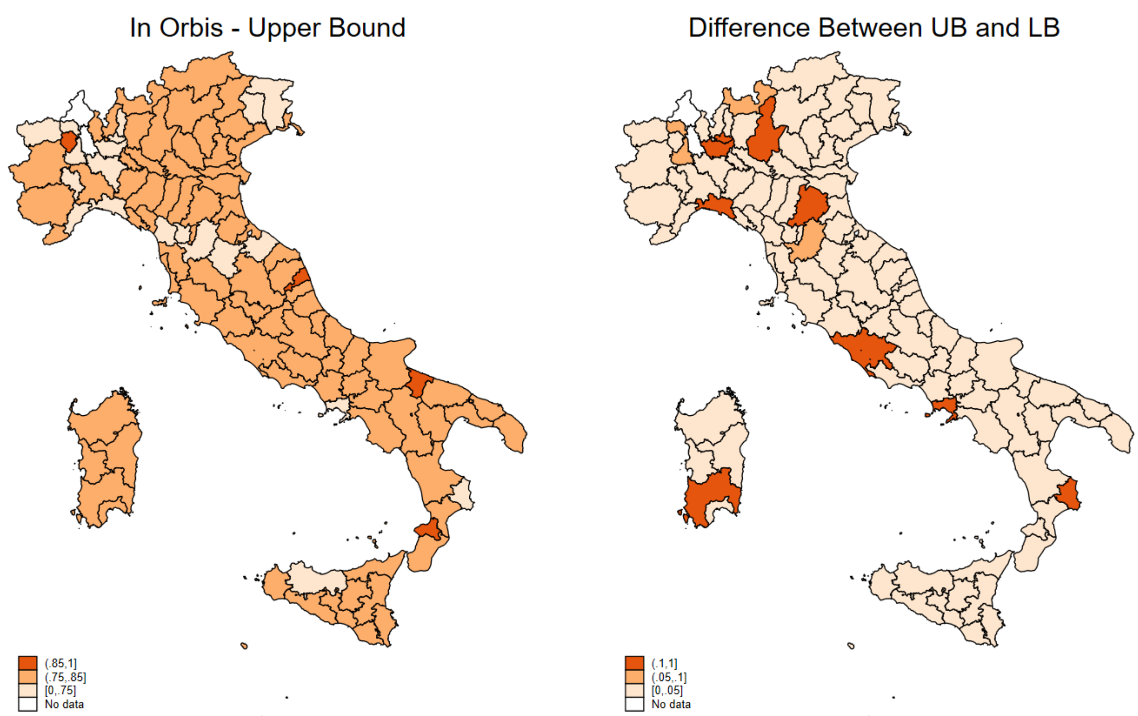
Structure of PIVA codes A firm's PIVA code is an 11-digit code composed of a 7-digit number, followed by 3 digits that identify the province where the firm is located, and finally a checksum. So, essentially, a ten-digit number. Querying the universe of such numbers one-by-one using the REST API to find which are in use is theoretically possible but impractical, as it represents ten billion potential values. The scale of the problem can be reduced with the knowledge that the first digits 8 and 9 represent public enterprises (not included in our sample, so not necessary to query), and the fact that the province codes only take on the values 1-100, 120, 121, 888, and 999. Even with these restrictions, it still represents a large problem.

Python Algorithm The Registro Aziende platform's REST API returning up to five results at a time helps with simplifying this problem. For example, Figure B.3.1 shows (visually) the results of querying the search box with the number 000001 – which returns five results. Since five is the maximum number, we need to provide the REST API with a more precise query to ensure we are obtaining information on all PIVA codes in this range. However, we also know that we can begin with the next valid PIVA after 00000150896, since that is the fifth result returned by the API and results are sorted sequentially (the lowest five PIVAs are shown). If instead querying the search box with the number 000001 had shown four results, our algorithm would have next queried the number 0000002.

[47] We stopped before completion because we were blocked by the platform.

Comparison to Orbis Figure B.3.2 shows the percentage of PIVA codes from Registro Aziende which are also in Orbis.[48] The left panel shows the upper bound of Orbis coverage, which assumes that every PIVA code in Orbis but not in the Registro dataset (947,051) is still active. In almost all provinces, the upper bound (UB) of Orbis coverage is between 75 and 85%. The lower bound (LB) is be the opposite assumption, that every one of those firms had exited before November 2023 but we have not flagged an exit using our rules described in Appendix Section B.2.3. We tested the validity of assuming the upper bound vs. the lower bound by checking the last date Orbis confirmed the firm was still active. For 201,296 of the 947,051 PIVA codes only in Orbis, Orbis confirmed activity in 2023 or later. Hence, some of the firms not in the Registro dataset are confirmed to be active (i.e. the Registro Aziende coverage is also imperfect). The right hand panel shows the difference between the upper and lower bounds, which is less than 5 percentage points in all but (mostly) the largest cities. Overall, these figures suggest that Orbis contains at least 75% of active firms in Italy as of 2023, with coverage being relatively even across provinces.

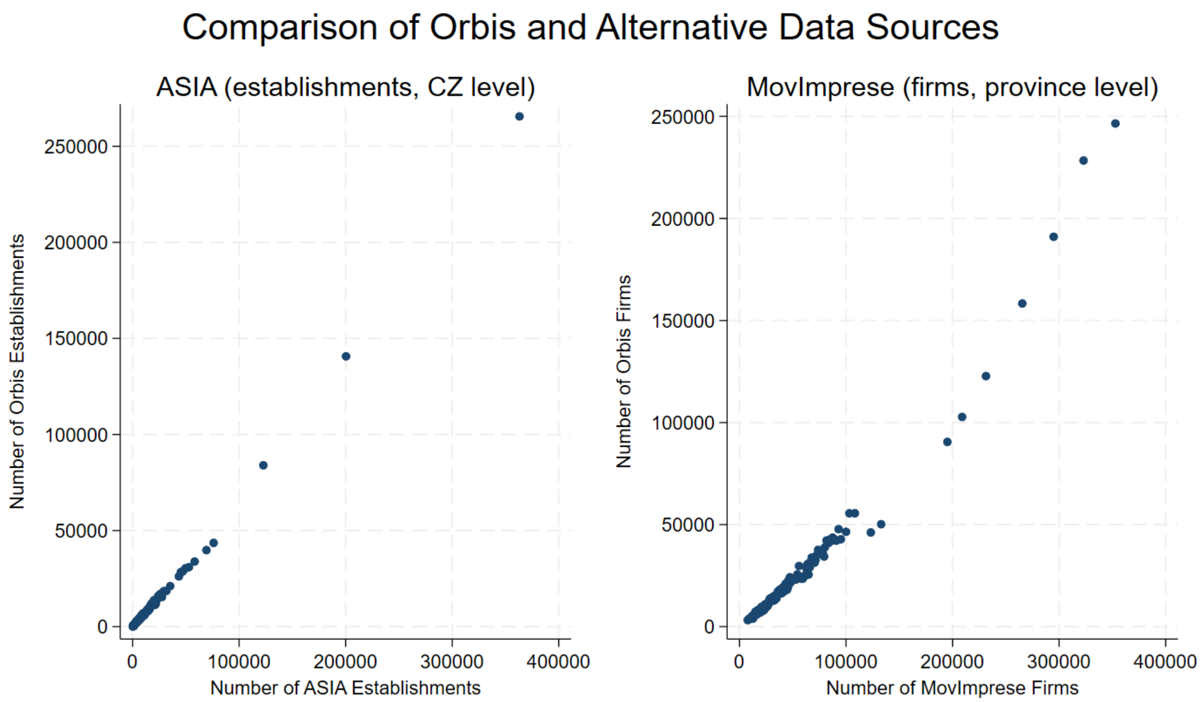
Figure B.3.2: Comparison of Orbis and Registro Aziende



Notes: The Orbis comparison sample consists of firms with a PIVA code identifier that had not exited the market prior to 2023 using the procedure described in Appendix Section B.2.3. The sample is also restricted to the range of PIVA codes queried using the procedure described in Appendix Section B.3.1. The left panel assumes that all firms contained in Orbis but not the Registro Aziende dataset are still active, while the right panel shows the percent difference in coverage when making the opposite assumption.

[48] Since the Registro Aziende database consists of firms which they believed to be active as of the end of 2023, we only compare the Registro sample to firms in Orbis which had not exited as of 2023. We also restrict to PIVA codes in the ranges we were able to query before being blocked by Registro Aziende.

Figure B.3.3: Comparison of Orbis to ASIA and MovImprese



Notes: The figure shows binned regressions using the binsreg command of Cattaneo et al. (2024). The left panel shows the relationship between the number of branches and firm headquarters in Orbis (dependent variable) on the number of establishments from the ASIA database (left panel), using data between 2007 and 2017 at the commuting-zone level. The right panel shows the relationship between the number of firms in Orbis (dependent variable) on the number of firms from the MovImprese database (left panel), using data between 2007 and 2022 at the provincial (NUTS3) level.

B.3.2 Comparison to ASIA (establishments) and MovImprese (firms)

We also compare Orbis to publicly available alternative aggregate data sources. Istat provides the ASIA-Imprese database on the number of active establishments between 2007 and 2017 at the 3-digit-ATECO-by-municipality level.[49] The Italian Chamber of Commerce maintains the MovImprese database, which contains information about the number of active firms at the provincial level by incorporation form and NACE sector. We first assess the overall coverage of the Orbis database. We then compare it to the above-mentioned datasets under the same sample restrictions applied in our analysis.

Overall Coverage Figure B.3.3 shows the comparison of the number of branches and firm headquarters in Orbis to the ASIA data, and the comparison of the number of firms in Orbis to the MovImprese data. Compared to both data sources, Orbis contains about 50% of active firms and establishments across all years, with coverage rates being similar between commuting zones and provinces.

We also assess overall Orbis coverage across space and time. Figure B.3.4 shows the percentage of firms in Orbis for the beginning (2007) and end (2022) of our analysis period.

[49]The ASIA-Imprese database does not include the NACE sectors A, O, T and U.

Panel A shows the percentage of firms in Orbis for all forms of incorporation, the same sample previously shown in Figure B.3.3. The Orbis coverage is relatively even across space in both 2007 and 2022, but improving over time, with most provinces covering 35-45% of all active firms in 2007, and 40-55% in 2022. Panel B shows the share of all firms in Orbis which are included in our analysis dataset – applying our sample restrictions described in Section B.2 to the Orbis data only. In both 2007 and 2022, between 5 and 15% of all firms are in our analysis dataset.

Analysis Sample Coverage The numbers in Figure B.3.4 represent all forms of incorporation, but our analysis dataset is almost entirely limited firms and partnerships since we exclude sole proprietorships due to their small size and lack of information about their financials available in Orbis. Thus, the more relevant comparison group is the population of incorporated firms and partnerships covered in the MovImprese dataset. In Panel A of Figure B.3.5 we show the share of limited firms/partnerships available in Orbis (Panel A). Coverage rates for limited firms and partnerships are much higher, more than 60% in almost all provinces in 2007, and more than 75% in almost all provinces in 2022 (consistent with the scraped data presented above). In some provinces, we have over-coverage, likely reflecting the imprecise nature of our exit measure – particularly for firms for which we lack financial information which are not in our analysis dataset. Panel B shows the share of limited firms/partnerships which are in our analysis dataset and supports this hypothesis as coverage across space is more even. For our analysis dataset, coverage is also more even over time, with between 25 and 40% of limited firms/partnerships included in both 2007 and 2022.

B.3.3 Employment coverage

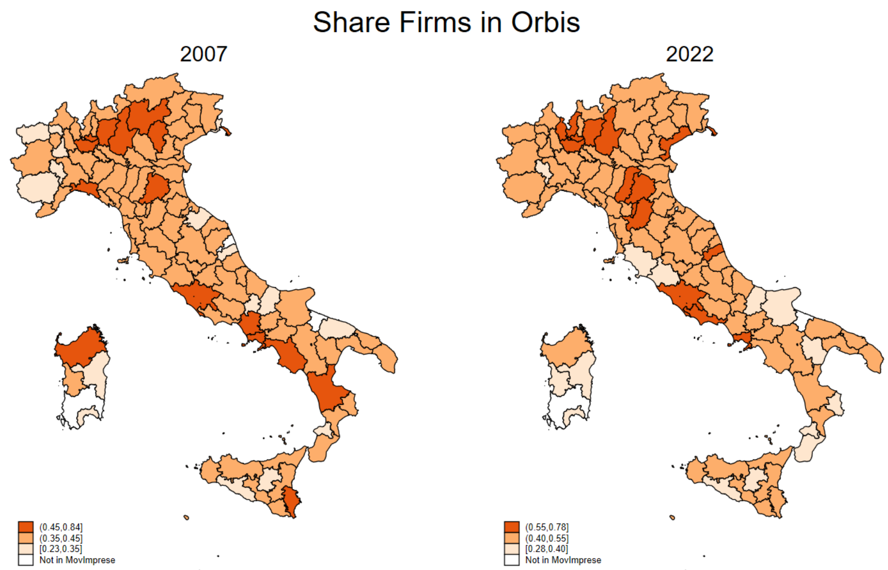
Having shown the overall good coverage of Orbis with respect to the share of firms included in the database, we now assess the quality of the Orbis coverage in terms of employment using ASIA. Figure B.3.6 shows the comparison of employment numbers between the two sources for all firms for which we have employment information in Orbis. Overall, approximately 60% of official employment is covered in Orbis between 2007 and 2017, with coverage being higher in the largest commuting zones. Figure B.3.7 maps the coverage rates across commuting zones for the years 2007 and 2017. As is clear from the figures, the overall employment coverage of Orbis increases over time.

Figure shows the overall percentages of employment which are included in our analysis dataset[50], while shows the spatial/time dimension of the overall coverage. Overall, about 20% of all employment is included in our analysis dataset. When making our analysis restrictions, our coverage becomes more even, with the majority of commuting zones having between 15 and 35% of their overall employment included in our analysis dataset in both years.

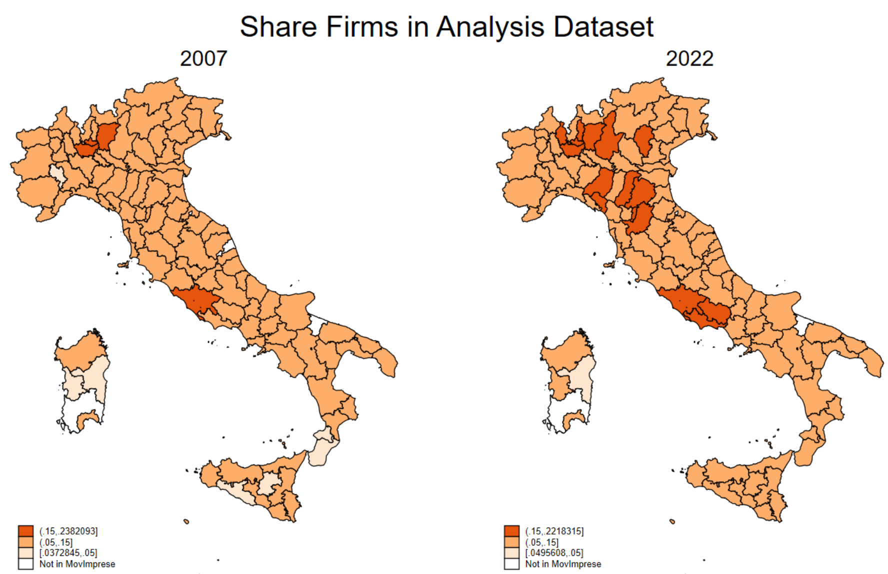
[50] We restrict the ASIA employment to sectors which we include in our analysis, we do not make restrictions upon incorporation form as it is not available in the municipal-level dataset from which we derive the commuting-zone level employment.

Figure B.3.4: Comparison of Orbis and MovImprese

Panel A



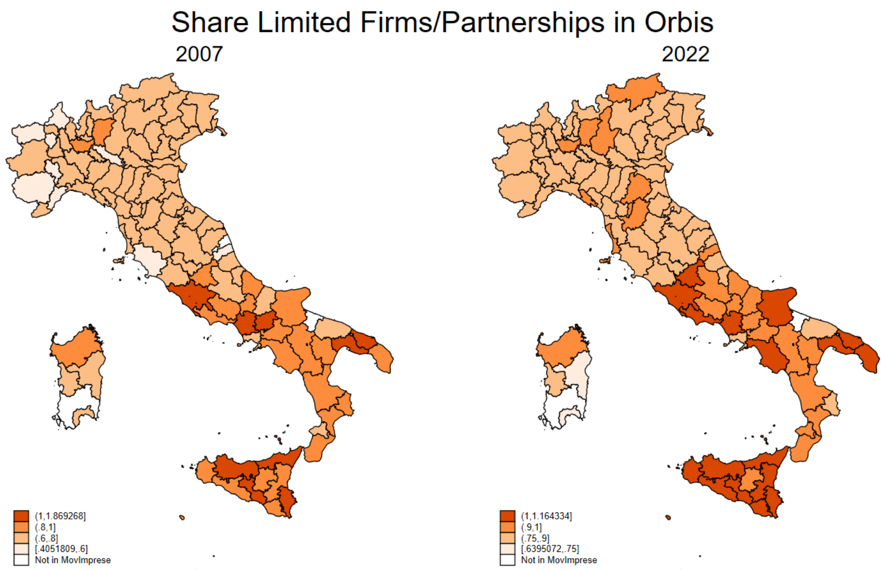
Panel B



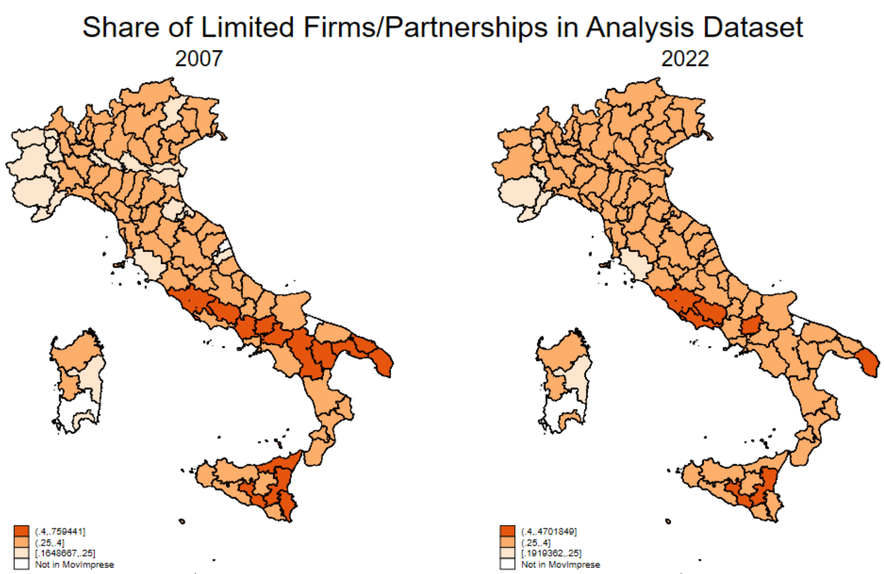
Notes: The figures show the share of firms contained in Orbis, calculated at the provincial (NUTS3) level using the aggregate numbers available in the MovImprese database. Panel A shows the share of the total contained in MovImprese, while Panel B shows the share of firms after making the sample restrictions described in Section 3.

Figure B.3.5: Comparison – Limited Firms/Partnerships

Panel A

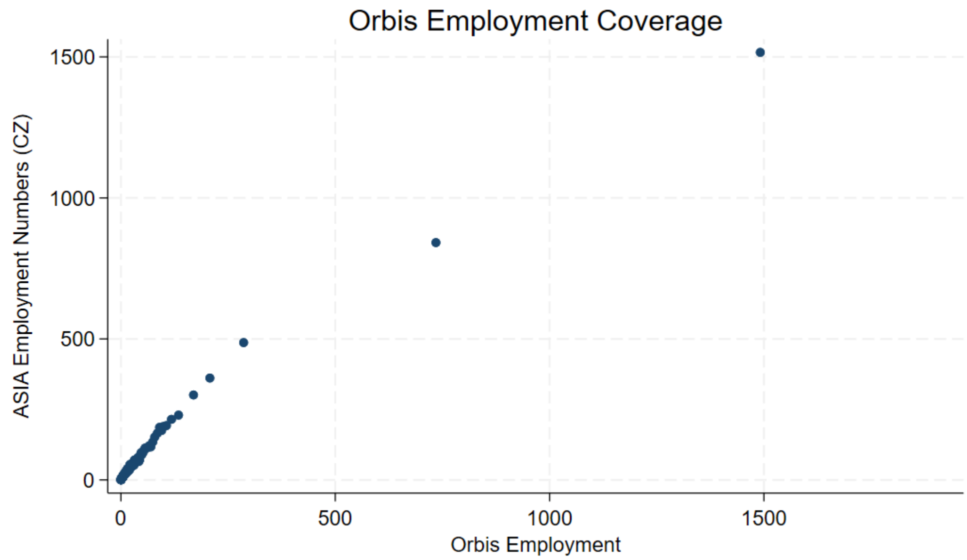


Panel B



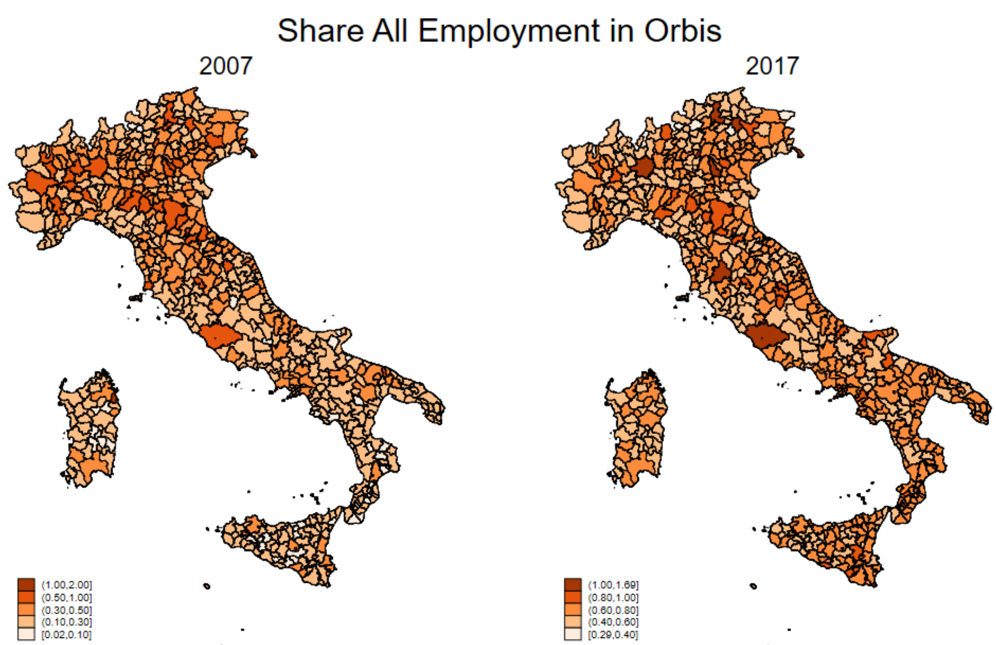
Notes: The figures show the share of firms contained in Orbis, calculated at the provincial (NUTS3) level available in the MovImprese database using only limited firms/partnerships in both sets of data. Panel A shows the share of the total contained in MovImprese, while Panel B shows the share of firms after making the sample restrictions described in Section 3.

Figure B.3.6: Aggregate Employment Comparison



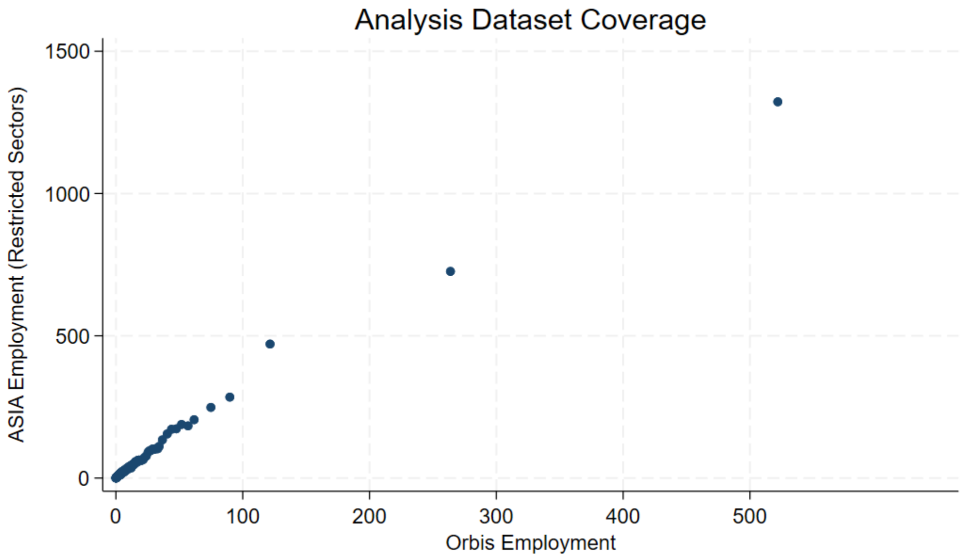
Notes: The figure shows a binned regression using the binsreg command of Cattaneo et al. (2024). The figure shows the relationship between the total commuting-zone level employment calculated from the ASIA dataset (dependent variable), compared to the number of employees covered by the Orbis sample (independent variable), using data between 2007 and 2017 at the commuting-zone level.

Figure B.3.7: Aggregate Employment Comparison



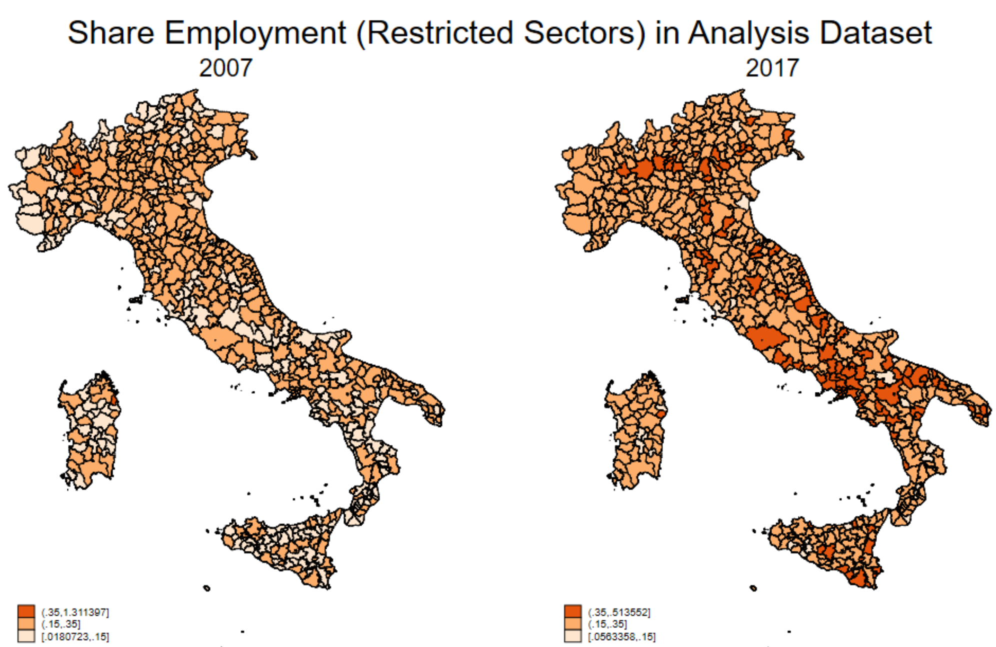
Note: The figures show the share of employment covered in Orbis, calculated at the commuting zone level using data available from the ASIA database.

Figure B.3.8: Analysis Dataset Coverage



Notes: The figure shows a binned regression using the binsreg command of Cattaneo et al. (2024). The figure shows the relationship between the commuting-zone level employment from the ASIA dataset (dependent variable), compared to the number of employees covered by the Orbis sample (independent variable), using data between 2007 and 2017 at the commuting-zone level after making the restrictions with respect to economic sector described in Section 3.

Figure B.3.9: Analysis Dataset Coverage



Note: The figures show the share of employment covered in Orbis, calculated at the commuting zone level using data available from the ASIA database after making the restrictions with respect to economic sector described in Section 3.

B.4 Counterfactual Funding Simulations

In this section, we outline the relevant institutional details we use to calculate the counterfactual levels of funding in our empirical strategy described in Section 5. The regulation for the 2007-2013 funding period is EU Regulation 1083/2006, while for 2014-2020 funding period it is Regulation 1303/2013.

First, we calculate the empirical cumulative density function of the measurement error in both regional GDP and population using the actual and historical GDP and GDP per capita estimated. We then simulate measurement error 1000 times accounting for serial autocorrelation in our main analysis (see Section 5.2.2), as well as for correlations of measurement error and other economic factors in our robustness checks (see Section 5.2.3). In each simulation we then calculate the eligibility position of the region under the counterfactual GDP per capita using the actual (revised) GDP and the simulated measurement error. Then, we apply the relevant funding formula to obtain the counterfactual financial envelope of the region in a particular simulation. The first relevant year of data used in the funding formulas is 2002. In the case that data is only available beginning in a year after 2002 (a very small number of cases), we use the earliest year of data available. Finally, we calculate the commuting zone level funding using the empirical shares as shown in Equation (5).

We are unable to precisely replicate the use of the three-year-averages of GDP and GDP per capita in the eligibility determination process discussed in Section 2 due to limitations in the historical GDP (per capita) data releases. Recall that for the 2014-2020 funding period, funding envelopes are determined based on regional GDP per capita estimates for the years 2007-2009, released in 2012. This data released by Eurostat in 2012, for example, reports a three year average for regional *GDP per capita* between 2007 and 2009 but only the single year estimate for regional *GDP* (not per capita) for the year 2009. To construct a 3-year moving average for regional GDP we cannot simply combine this single-year estimate and documents from previous years (i.e. use 2011 Eurostat releases to get the 2008 regional GDP per capita figures) because initial revisions for the 2008 figures will have already taken place between 2011 and 2012. Therefore, it is impossible for us to back out what the 2012 estimate was for the 3-year moving average of regional GDP. Although we cannot incorporate the 3-year moving average for regional GDP, we can use the 3-year moving average of the GDP per capita to compare the distribution as that information is provided directly in the historical GDP releases. Appendix Figure C.1.1 shows that the incorporation of the moving average does not substantially change the distribution of the measurement error. This is precisely because the errors made in the calculations are serially correlated over short time periods, as discussed in the main text in Section 5.2.2.

B.4.1 2007-2013 Funding Period

Regulation 1083/2006 Annex II paragraph 9 specifies that the NUTS2 regional GDP per capita figures used by the Commission to calculate funding eligibility are those published as of April 2005 (Eurostat press release STAT/05/47, published April 7, 2005). GDP per capita was measured in PPS. There are four eligibility bins that regions may be sorted into (see Articles 5(1), 6 and 8 of Regulation 1083/2006 for details).

1. **Less developed regions** are those with a GDP per capita less than 75% of the *EU-25* average GDP per capita.

2. **Phasing out regions** are those that would have been classified as less developed if the EU Enlargement of 2004 had not occurred.[51] More precisely, phasing-out regions are those below the 75th percentile of the *EU-15* average GDP per capita.

3. **Phasing in regions** are those that were less developed regions during the final year of the 2000-2006 funding period, but their relative GDP per capita was now above 75% EU-15 average GDP per capita.

4. **More developed regions** are all other regions.

Conditional on their respective eligibility bin, regions are subject to different funding formulas. Below, we discuss any relevant interpretations of each regulation section individually, the data used in each calculation, as well as any additional assumptions made during the process.

Less developed degions The funding formula for less developed regions in the 2007-2013 funding period is given in Annex II paragraph 1 of Regulation 1083/2006. Concerning the national Gross National Income per capita referred to in paragraph (b), we use data from the World Bank (as Eurostat only has the data available beginning in 2019). The World Bank data is revised, and we do not attempt to simulate measurement error in GNI per capita for two reasons. First, historical numbers are not easily available. Second, we expect measurement error in GNI to be negligible as national numbers are typically measured more accurately.

Concerning the aid intensity premium for the number of unemployed persons in paragraph (c), we additionally assume that the percent error in the population in a specific age group is the same as the overall percent error in the population estimate. We make this assumption for all other calculations involving population numbers. For the data on the revised number of unemployed persons, we use the data series *lfst_r_lfu3pers* from Eurostat.

Phasing Out and Phasing In Regions The funding formula for Phasing in and Phasing out regions is given in Annex II paragraph 6 of Regulation 1083/2006. These regions were subject to a linear reduction in their per capita aid intensity from 2006 – the final year of the 2000-2006 funding period – alongside a premium for unemployed persons similar to that used in the less developed regions. Per capita aid intensity for the year 2006 was obtained from the European Union's "Historical EU Payments by MS & NUTS2 Region" tool[52] for regions which were classified as less developed in 2006.

[51] In 2004, ten countries joined the EU: Cyprus, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia. Because those countries were poorer than the old EU-15 member states, the EU average dropped. This lead to the "statistical effect" – regions crossing the 75% *EU-25* threshold because of *relative (not absolute)* improvements in their regional development.

[52] See `https://cohesiondata.ec.europa.eu/EU-Level/Historic-EU-payments-by-MS-NUTS-2-region-filter-by/2qa4-zm5t`.

More Developed Regions The funding formula for more developed regions is given in Annex II paragraph 4 in Regulation 1083/2006. We first discuss the interpretation of the formula outlined in Annex II paragraph 4. we then specify the data used and further assumptions needed to apply the formula.

The sum of the shares referred to in the formula requires taking the member states' most developed regions' share of each variable, and then summing those shares together with the given weights. For instance, the first variable entering the funding formula is total population with a weight of 0.5. Concretely, assume that Italy's more developed regions represent 4% of the total population of all more developed regions across the EU. Then the value for the first condition would be .04*.5 for Italy's more developed regions. Then if Italy's share of the second condition – the "number of unemployed people in NUTS level 3 regions with an unemployment rate above the group average (weighting 0,2)" among the member states' more developed regions – was 5%, the second value would be .05*.2. If there were only two conditions, we would obtain a summed share of .04*.5 + .05*.2 = .03 for Italy's more developed regions. As a last step, we would obtain the financial aid available to all Italian more developed regions multiplying this summed share by the total envelope available to most developed regions during the funding period specified in Article 18 of Regulation 1083/2006.

To compute the above mentioned shares, we had to make further assumptions. For the second share condition (the "number of unemployed people in NUTS level 3 regions with an unemployment rate above the group average") we were unable to find data on the number of unemployed by NUTS3 regions for all European regions, so we use the same method as was used in the 2014-2020 funding period in Regulation 1303/2013, which is discussed below. The final share condition of "low population density" is also interpreted the same way as specified in the 2014-2020 funding period in Regulation 1303/2013 discussed below. For the third share condition "number of jobs needed to reach an employment rate of 70%" we use the Eurostat series *demo_r_d2jan* to calculate the counterfactual estimated population ages 20-64. Finally, for the fourth condition, we use the Eurostat series *lfst_r_lfe2eedu* and interpret low education as isced11=ED0-2.

B.4.2 2014-2020 Funding Period

Regulation 1303/2013 Annex VII paragraph 12 specifies that the NUTS2 regional GDP figures used by the Commission to calculate funding eligibility are those published as of May 2012 (Eurostat press release 38/2012, published March 13, 2012). There are three eligibility bins that regions may be sorted into (see Article 90 paragraph 2 of Regulation 1303/2013 for details).

1. **Less developed regions** are those with a GDP per capita less than 75% of the EU average.
2. **Transition regions** are those with a GDP per capita between 75% and 90% of the EU average.
3. **More developed regions** are all other regions (more than 90% of the EU average).

Less Developed Regions The funding formula used for less developed regions in the 2014-2020 funding period is nearly identical to that used in the 2007-2013 funding period, and we use the same assumptions discussed in the paragraph concerning less developed regions in Section B.4.1.

Transition Regions Transition regions in the 2014-2020 funding period receive financial aid calculated using a linear interpolation between a maximum and a minimum value. The maximum aid intensity is 40% of what the transition region would receive if it was instead a less developed region with a GDP per capita exactly at 75% of the EU average. The minimum intensity is the aid intensity of that country's most developed regions. We calculate the linear interpolation assuming that some epsilon above the 75% threshold a region receives the maximum value, and some epsilon below the 90% threshold the region receives the minimum value.

More Developed Regions The funding formula for more developed region is, as in the 2007-2013 funding period, a sum of shares multiplying an aid intensity. They key difference is that in the 2014-2020 funding period, the multiplied value is a per capita annual aid intensity for the population of the country's more developed regions rather than a total Euro value. For step (b) –"number of unemployed people in NUTS level 2 regions with an unemployment rate above the average of all more developed regions" – we use the same method as discussed in the paragraph concerning less developed regions in Section B.4.1 for calculation of the number of unemployed persons and the unemployment rates. Step (b) is the method used for the second share condition discussed in the paragraph concerning more developed regions in Section B.4.1.

Furthermore, as discussed in the paragraph concerning less developed regions in Section B.4.1, we assume the overall measurement error in the population to be the same as the measurement error for the particular population subgroups used in steps (c) to (e). For these steps, we use the Eurostat datasets: *estat_lfst_r_lfe2emp* for employment, *estat_edat_lfse_04* for tertiary education, and *estat_edat_lfse_16* for early leavers. For step (g) (also the final share condition in the paragraph concerning more developed regions in Section B.4.1) we use historical NUTS3 population data from the ARDECO database. Historical NUTS3 geographic areas are calculated using GIS from projections available from the European Commission.[53]

B.5 Other Data

B.5.1 Municipal Geocode Harmonizations

The OpenCoesione data are coded to the latest municipal codes – the 2022 codes as of our download. However some of our data (such as ASIA and MovImprese) are coded using previous vintages of the regional codes. The Italian Statistical Institute provides detailed

[53]`https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts`.

information on municipal changes.[54] Using information about these changes, we forward-coded municipal codes using the information about the type of change available via Istat. Changes involving partial splits of municipalities (i.e. cessions of territory to another municipality) were deemed unacceptable, while absorption of one municipality by another, province changes, and simple name changes were deemed acceptable. Fortunately, in the Italian context partial splits of municipalities are very rare, less than 10 cases during our sample period out of roughly 8000 Italian municipalities, and are all extremely small in terms of population. We drop these municipalities from our analysis sample.

With respect to our identification strategy, we do not need to make any adjustments due to regional boundary changes. Italian regional boundaries have remained constant throughout our entire sample period with the exception of seven small municipalities in the region Emilia-Romagna that detached from the Marche region in the year 2009. Thus, using the historical boundaries and expected funding based on historical data poses no problems in our analysis.

Commuting zone boundaries are defined using the mappings of municipal codes to the 2011 revisions of the SLL codes provided by Istat.

B.5.2 Municipal Population

Municipal population is available using the 2019 municipal codes, downloaded from Istat.

References

Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230, 2021.

Matias D Cattaneo, Richard K Crump, Max H Farrell, and Yingjie Feng. On binscatter. *American Economic Review*, 114(5):1488–1514, 2024.

Andrea Cerrato. How big is the big push? the macroeconomic effects of a large-scale regional development program. *Unpublished, University of California, Berkeley*, 2024.

Michael Greenstone, Richard Hornbeck, and Enrico Moretti. Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, 118 (3):536–598, 2010.

Şebnem Kalemli-Özcan, Bent E Sørensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yeşiltaş. How to construct nationally representative firm-level data from the orbis global database: New facts on smes and aggregate implications for industry concentration. *American Economic Journal: Macroeconomics*, 16(2):353–374, 2024.

Patrick Kline. Place based policies, heterogeneity, and agglomeration. *American Economic Review*, 100(2):383–387, 2010.

Patrick Kline and Enrico Moretti. Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. *The Quarterly Journal of Economics*, 129(1):275–331, 11 2013.

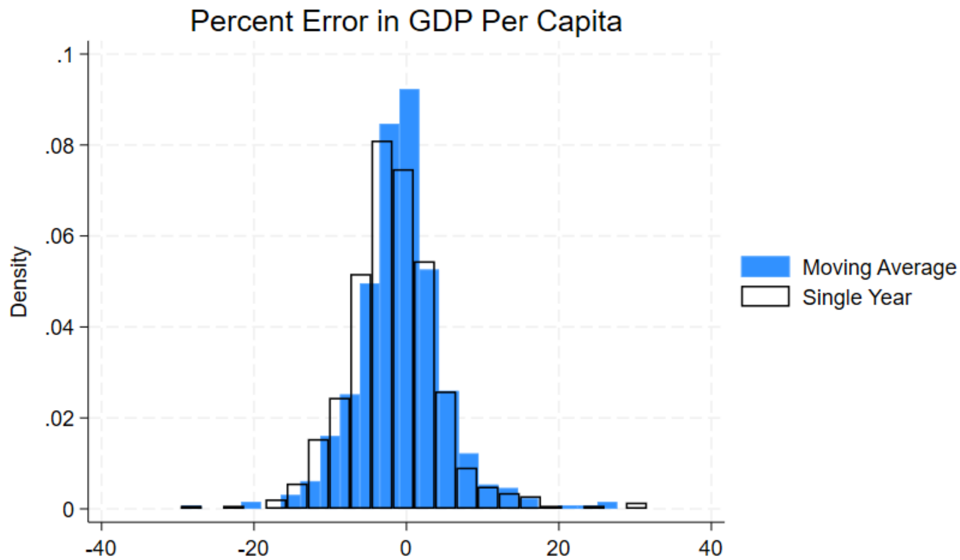
[54] See `https://www.istat.it/it/archivio/6789`.

Jeffrey M Wooldridge. On estimating firm-level production functions using proxy variables to control for unobservables. *Economics letters*, 104(3):112–114, 2009.

C Additional Figures and Tables

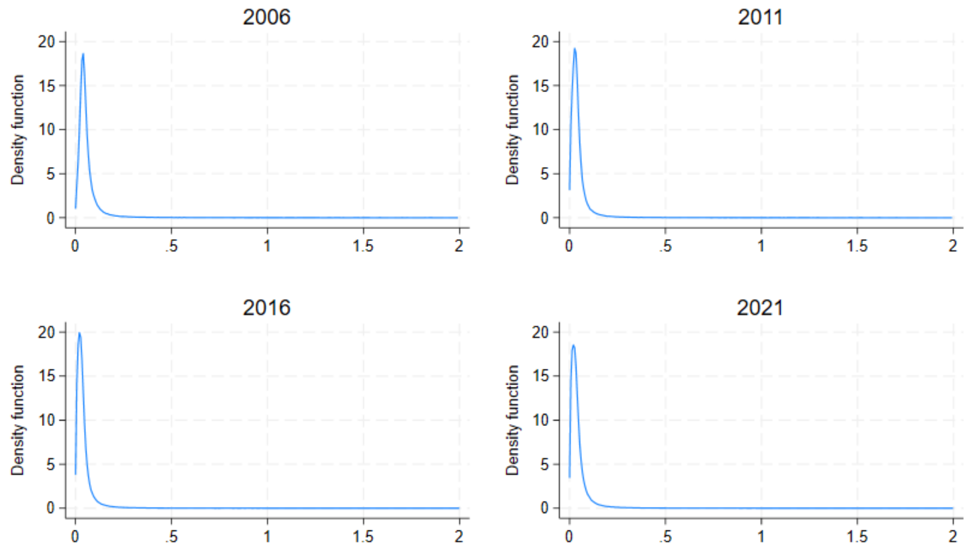
C.1 Additional Figures

Figure C.1.1: Empirical Probability Density Function of Measurement Error with Moving Average



Notes: Authors' calculations using historical regional GDP per capita estimates for 2002 and 2009 for all EU member states, both for the three-year average figures released in that year. They are compared to a three year average of revised figures at the NUTS2 regional level.

Figure C.1.2: Real worker added-value (million Euros)



Note: Density function of real worker added-value for the years 2006, 2011, 2016 and 2021 after trimming the top and bottom 5% of the firm-level real value-added distribution.

C.2 Additional Tables

Table C.2.1: Eligibility in Counterfactual Simulations

Panel A: 2007-2013 Funding Period

	Classification of Eligibility		
	More Developed	Transition	Less Developed
NUTS2 Name			
Abruzzo	978	22	
Basilicata		875	125
Calabria		4	996
Campania		331	669
Emilia-Romagna	1,000		
Friuli-Venezia Giulia	1,000		
Lazio	1,000		
Liguria	1,000		
Lombardia	1,000		
Marche	1,000		
Molise	690	280	30
Piemonte	1,000		
Provincia Autonoma Bolzano/Bozen	1,000		
Provincia Autonoma Trento	1,000		
Puglia		164	836
Sardegna		737	263
Sicilia		133	867
Toscana	1,000		
Umbria	1,000		
Valle d'Aosta/Vallée d'Aoste	1,000		
Veneto	1,000		

(Continued)

Panel B: 2014-2020 Funding Period

	Classification of Eligibility		
	More Developed	Transition	Less Developed
NUTS2 Name			
Abruzzo	613	385	2
Basilicata	9	799	192
Calabria		41	959
Campania		318	682
Emilia-Romagna	1,000		
Friuli-Venezia Giulia	999	1	
Lazio	1,000		
Liguria	1,000		
Lombardia	1,000		
Marche	983	17	
Molise	230	753	17
Piemonte	1,000		
Provincia Autonoma Bolzano/Bozen	1,000		
Provincia Autonoma Trento	1,000		
Puglia		72	928
Sardegna	21	847	132
Sicilia		216	784
Toscana	1,000		
Umbria	955	45	
Valle d'Aosta/Vallée d'Aoste	1,000		
Veneto	1,000		

Note: Each cell is the number of simulations (out of 1000) in which each Italian region is classified into a particular eligibility category using the measurement error simulation procedure described in Section 5.

Table C.2.2: Additional Robustness Checks

Panel A: Reallocation Effect

	(1)	(2)	(3)
	Δcov	Δcov	Δcov
Log(Payments)	-0.0898**	-0.0602***	-0.0531**
	(0.0421)	(0.0204)	(0.0191)
Log(E(Payments))	0.0958**	0.0672**	0.0626***
	(0.0431)	(0.0262)	(0.0192)
Observations	1040	1040	1040
Region FE	Yes	No	No
Industry Composition Controls	No	Yes	No
Independent Variable Total Payments	No	No	Yes

Panel B: Entry Effect

	(1)	(2)	(3)
	Entry	Entry	Entry
Log(Payments)	0.0641	0.0299**	0.0109
	(0.0409)	(0.0111)	(0.00988)
Log(E(Payments))	-0.0681	-0.0335**	-0.0186*
	(0.0416)	(0.0128)	(0.00943)
Observations	1040	1040	1040
Region FE	Yes	No	No
Industry Composition Controls	No	Yes	No
Independent Variable Total Payments	No	No	Yes

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

Notes: The dependent variables are the change in the reallocation component (Panel A) and the entry component (Panel B) of aggregate productivity using the method of Melitz and Polanec (2015) at the commuting-zone-by-funding-period level, respectively. At the firm-level, productivity is measured as value added per worker. Payments are measured per capita using the population of the commuting zone in the first year of the funding period (2007 and 2014, respectively) and include both European and national funds. Expected payments per capita are calculated using the OpenCoesione data and the simulation procedure described in Section 5. All specifications include funding period fixed effects and controls for a commuting zone's initial productivity level, demographic characteristics (share illiterate, literate, elementary school, middle school, and college/university – high school excluded – age under 14 and over 65 – 15-64 excluded, share male, and share foreign), and other EU payments (ERDF and ESF projects taking place across the entire region, and payments from other EU funds specific to the commuting zone or taking place across the entire region from the OpenCoesione data). Standard errors are reported in parentheses and are clustered at the NUTS2 regional level. The sample includes only commuting zones within a single region, and commuting zones with less than ten firms are excluded.