

Bachtrögler, Julia; Hammer, Christoph; Reuter, Wolf Heinrich; Schwendinger, Florian

**Working Paper**

## Spotlight on the beneficiaries of EU regional funds: A new firm-level dataset

Arbeitspapier, No. 02/2017

**Provided in Cooperation with:**

German Council of Economic Experts

*Suggested Citation:* Bachtrögler, Julia; Hammer, Christoph; Reuter, Wolf Heinrich; Schwendinger, Florian (2017) : Spotlight on the beneficiaries of EU regional funds: A new firm-level dataset, Arbeitspapier, No. 02/2017, Sachverständigenrat zur Begutachtung der Gesamtwirtschaftlichen Entwicklung, Wiesbaden

This Version is available at:

<https://hdl.handle.net/10419/162836>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Spotlight on the beneficiaries of EU regional funds:  
A new firm-level dataset

---

Julia Bachtrögler  
(Vienna University of Economics and Business)

Christoph Hammer  
(Vienna University of Economics and Business)

Wolf Heinrich Reuter  
(Staff of the German Council of Economic Experts)

Florian Schwendinger  
(Vienna University of Economics and Business)

Working Paper 02/2017\*)  
July, 2017

# Spotlight on the beneficiaries of EU regional funds: A new firm-level dataset\*

Julia Bachtrögler<sup>1</sup>, Christoph Hammer<sup>1</sup>, Wolf Heinrich Reuter<sup>2</sup>, and Florian Schwendinger<sup>1</sup>

<sup>1</sup>Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria. E-mail: julia.bachtroegler@wu.ac.at, christoph.hammer@wu.ac.at, florian.schwendinger@wu.ac.at

<sup>2</sup>Staff of the German Council of Economic Experts, Gustav-Stresemann-Ring 11, 65189 Wiesbaden, Germany. E-mail: wolf.reuter@svr-wirtschaft.de

## Abstract

This study introduces a new firm-level dataset containing over two million projects co-funded by the European Union's (EU) structural and Cohesion funds in 25 EU member states in the multi-annual financial framework 2007-2013. Information on individual beneficiary firms and institutions published by regional authorities is linked with business data from Bureau van Dijk's ORBIS database. Moreover, we show how modern text mining techniques can be used to categorise EU funded projects into fifteen thematic categories proposed by the European Commission.

A first analysis of the dataset reveals substantial heterogeneity of beneficiaries and projects across and within countries. While in the majority of lagging regions the largest project expenditure is dedicated to transportation and energy infrastructure, in most other regions the major part is assigned to innovation and technological development as well as business (including SME) support. In an econometric analysis we control for project and firm characteristics and find that the highest single project values are associated with older beneficiary firms that are larger in size. Furthermore, the projects with topmost expenditure are carried out in Dutch and British regions.

**Keywords:** Distribution of EU structural funds, Regional policy, Firm-level data, Cohesion, European Union

**JEL Classification:** E61, H77, R11, R58

---

\*Corresponding author: Julia Bachtrögler (julia.bachtroegler@wu.ac.at). Project funding by the Vienna University of Economics and Business (Institute for International Economics and "small-scale project by WU junior faculty") is greatly appreciated. The authors gratefully acknowledge assistance by the following students: Konrad Fätkenheuer, Maximilian Fritz, Maria Kodajova, Christian Lubi, Andrej Mijakovic, Anna Stelzer and Tinja Zerzer. Moreover, we thank participants at the WU Department of Economics seminar 2016, ERSA 2016, the Geoffrey Hewings Regional Economics Workshop 2016, a workshop at the Joint Research Centre in 2016 and a seminar at the German Council of Economic Experts for very helpful discussions. The opinions expressed in this paper are those of the authors only and do not necessarily reflect the official viewpoint of the German Council of Economic Experts.

# 1 Introduction

Through its structural funds and the Cohesion Fund, the European Union (EU) redistributes over EUR 348 billion among its regions to foster regional development and cohesion. The design of regional policy programmes is a prominent, recurring theme in the academic as well as public debate. There is a broad literature that investigates different aspects of the effectiveness of regional funds across European regions. Due to a lack of data, research so far has mainly focused on the effects (e.g., in terms of increasing income growth) at an aggregated level or has been limited to specific funding programmes. However, more detailed information on the distribution of funds within a region allows to investigate the optimal allocation of regional funds further.

The European Regional Development Fund (ERDF), the European Social Fund (ESF) and the Cohesion Fund (CF), in this study referred to as the EU's regional funds, co-finance projects that are part of operational programmes (OP) targeted at strategic priorities like strengthening the labour market, improving social infrastructure or building better traffic networks.<sup>1</sup> The projects are carried out by firms, institutions or other entities that receive the co-funding and are selected by an OP's managing authority (a public or private body nominated by the member state).<sup>2</sup> Thus, the beneficiaries are *de facto* responsible for the single projects' success that contributes to achieving the corresponding OP's target and, therefore, the overall effectiveness of a region's regional policy implementation.<sup>3</sup>

The five main contributions of this study to the literature are the following: First, we introduce a novel dataset that contains over two million projects in 25 EU member states co-funded by EU regional funds in the programming period 2007-2013. Second, the data offers new project-level insights into the distribution of EU regional funds with respect to geographical, industrial and other characteristics of the funds' beneficiaries. A thorough investigation of both the individual project and beneficiary entails interesting policy conclusions regarding effective regional funds distribution that have been hidden so far. Third, we provide an R package (*fastTextR*) which we use for classifying projects into fifteen themes (thematic categories) defined by the European Commission.<sup>4</sup> Fourth, the econometric analysis reveals the determinants of the size of single regional policy projects. Fifth, the in-depth analysis of the dataset structure includes key suggestions for other researchers working with similar data.

The novel database contains 2,055,375 projects in 25 EU member states that are carried

---

<sup>1</sup>Note that we do not consider projects co-funded by the European Agricultural Fund for Rural Development (EAFRD) and the European Maritime and Fisheries Fund (EMFF) which are also EU regional policy instruments.

<sup>2</sup>Refer to European Council (2006b) for detailed information on regional policy implementation.

<sup>3</sup>The European Commission considers various indicators, e.g., the number of jobs created or the number of direct investment aid projects to small and medium-sized enterprises, for its evaluation of regional policy (see [http://ec.europa.eu/regional\\_policy/en/policy/evaluations/ec/2007-2013/#1](http://ec.europa.eu/regional_policy/en/policy/evaluations/ec/2007-2013/#1)).

<sup>4</sup>Refer to Section 3 for a detailed description of the classification task.

out by 1,076,097 beneficiaries. The beneficiaries are matched with the ORBIS business database by Bureau van Dijk in order to add some of their business characteristics. For each project, the database contains three blocks of information related to i) the project and corresponding operational programme, ii) its funding (co-financing) structure and iii) additional information on the beneficiary from ORBIS.

The dataset enables us to draw a detailed picture of the distribution of EU regional funds within regions. We find strong differences in the number of projects carried out in regions and their average project values. Moreover, in low-income regions, the largest expenditure tends to be allocated to transportation or energy infrastructure, environment or human capital and labour market projects. Contrarily, in the majority of the other regions, the largest project expenditure is assigned to innovation and technological development as well as support for small and medium-sized enterprises (SME). The focus on innovation as well as the fact that the sum of total project values is highest for projects carried out by small firms is in line with the Community's strategic guidelines on cohesion (European Council 2006a). Conditional on project- and firm-level characteristics, econometric analyses reveal that the largest single projects in terms of their total value are co-funded by the ERDF, aimed at fostering the transport infrastructure in a region (*Road, Rail and Other Transport*). Finally, projects with highest conditional project values are conducted in regions of the United Kingdom or the Netherlands.

The remainder of this paper is structured as follows: Chapter 2 gives an overview on the EU regional policy design and places the database into the context of existing literature. Chapter 3 describes the content of the database in detail and compares it with existing official data. Chapter 4 shows the distribution of total project volumes across regions, strategic priorities of EU regional policy and firm characteristics. Chapter 5 provides econometric analyses with a focus on the individual project level and, finally, Chapter 6 concludes.

## 2 The distribution of EU regional funds

Next to expenditures of EUR 412,611 million in the “natural resources” programme which includes agricultural subsidies, the EU's regional funds of EUR 348,865 million are the second largest item of its budget in the period (multi-annual financial framework, MFF) 2007-2013.<sup>5</sup> The EU's regional policy objective is to increase economic and social cohesion across European regions by co-financing member states' initiatives targeted at specific priorities. First, a region's principal eligibility for funding is in line with three main objectives, i) *Convergence* (former Objective 1), ii) *Regional Competitiveness and Employment*

<sup>5</sup>See [http://ec.europa.eu/budget/figures/fin\\_fw0713/fw0713\\_en.cfm#cf07\\_13](http://ec.europa.eu/budget/figures/fin_fw0713/fw0713_en.cfm#cf07_13).

(former Objective 2), and iii) *Territorial cooperation* (former Objective 3).<sup>6</sup> Second, the funds are allocated to national or regional operational programmes that (in the case of a national OP always) have a specified thematic target, e.g., improving regional human capital or social infrastructure. Private and public firms carry out projects that are co-financed in order to contribute to the OP's targets. The new dataset is based on lists of these beneficiary firms, institutions, non-governmental organizations or other types of entities (in the following we refer to all types of beneficiaries by the term *firm*) that have to be made public since the MFF 2007-2013 (Article 7 in European Commission 2006).

## 2.1 Cohesion policy design

The EU's cohesion policy works under the principle of shared management (refer to Article 14 and 15 in European Commission 2006). That is why multiple European and national institutions are involved in the allocation process of regional funds.

For each MFF, the European Council prepares a document with strategic guidelines on reducing economic, social and territorial disparities. These guidelines for 2007-2013 encompass three priorities (European Council 2006a): i) improving transport infrastructure, environmental and energy issues, ii) creating more and better jobs, and iii) a focus on knowledge transfer and innovation. Regarding the latter, special emphasis is put on supporting small and medium-sized enterprises (SMEs) which “often represent the highest source of employment at the regional level” (European Council 2006a, p. 19). The strategic guidelines on cohesion serve as a basis for the so-called national strategic reference framework that needs to be provided by each member state. The national strategic reference framework provides an overview of fields for intervention of cohesion policy in the particular country and undergoes a review process by the European Commission. Besides a proposal for the annual allocation of regional funds across the period, the member states need to generate a list of operational programmes (per objective and fund) for the objectives *Convergence* and *Regional Competitiveness and Employment*. Each operational programme is prepared and implemented by a managing authority, a private or public body appointed by the member state (Article 59 of European Council 2006b). In most cases, they refer to particular NUTS-2 or NUTS-1 regions, however, there are also national, NUTS-0, programmes (see Appendix A.1 and Title III in European Council 2006b).<sup>7</sup> The OPs must incorporate reasons for focusing on specific priority axes proposed by the Commission (see Annex IV of European Council 2006b).<sup>8</sup> From these priority axes,

<sup>6</sup>In the MFF 2007-2013, regions with a gross domestic product (GDP) per capita below 75% of the EU-25 average in 2000-2002 are eligible for funds under the *Convergence* objective. The remaining regions are eligible for transfers under the *Regional Competitiveness and Employment* objective and *Territorial cooperation* initiatives (Article 3 in European Council 2006b). Member states can apply for support from the Cohesion Fund if their gross national income (GNI) per capita lies below 90 % of the EU average.

<sup>7</sup>NUTS regions at different levels are defined according to the *Nomenclature des unités territoriales statistiques* 2010 (NUTS 2010) (European Commission 2011).

<sup>8</sup>See also “ERDF/ESF/CF Priority theme overview 2007-2013” at [http://ec.europa.eu/regional\\_policy/en/policy/evaluations/data-for-research/](http://ec.europa.eu/regional_policy/en/policy/evaluations/data-for-research/).

the European Commission derives fifteen so-called themes.

In the next step, the OPs' managing authorities select appropriate projects that are carried out by firms, i.e., the beneficiaries. According to Article 2 of Council Regulation (EC) No 1083/2006, a beneficiary is defined as "an operator, body or firm, whether public or private, responsible for initiating and implementing operations". An operation is referred to as "a project or groups of projects selected by the managing authority of the operational programme [...] allowing achievement of the goals of the priority axis to which it relates" (European Council 2006*b*). The amount of EU co-funding for each project depends on the eligible expenditure and the designated co-financing rate.<sup>9</sup>

The structure of the new database builds on these regulations. The dataset includes the OP to which each observation (project) is assigned to, the corresponding fund and objective. Furthermore, we assign a respective theme (as defined, but not provided by the EC) to each observation. Thus, we are able to check the validity of our data by comparing it with official numbers by DG REGIO in different dimensions (see Section 3.3).

## 2.2 Literature review

While there is a broad literature on the evaluation of cohesion policy, the present paper presents new data on beneficiaries in 25 EU member states. Most studies that investigate effects in a pan-European setting focus on the regional (in many cases NUTS-2) level (see, e.g., Hagen and Mohl 2009) for a survey). The new data enables project- and firm-level analyses which contributes interesting conclusions on the effectiveness of cohesion policy to the academic and political debate.<sup>10</sup>

Generally speaking, there is no consensus in the literature regarding the outcome of regional policy. Most studies find a conditional positive effect of regional funds assignment (e.g., Becker et al. 2013, Cappelen, Castellacci, Fagerberg and Verspagen 2003, Ferrara et al. 2016), while others provide results that even suggest a negative impact (Breidenbach, Mitze and Schmidt 2016). In recent years, the reasons for heterogeneous (conditional) regional policy effects have gained major attention: First, Rodríguez-Pose and Fratesi (2004) move the focus to expenditure categories of the main funding instruments. They find that investments in infrastructure or agriculture do not have sustainable effects on regional growth, though, projects that foster human capital lead to sustainable positive effects on

---

<sup>9</sup>Annex III of European Council (2006*b*) reports the ceilings for co-financing rates, i.e., the maximum percentage of eligible expenditure that is financed by a regional fund. E.g., the maximum co-financing rate for Spain amounts to 80% for the *Convergence* and to 50% for the *Regional Competitiveness and Employment* objective. The detailed regulation on the eligibility of expenditure can be found in Council Regulation (EC) No 1083/2006 and Commission Regulation (EC) No 1828/2006.

<sup>10</sup>The effectiveness is measured, e.g., as a positive effect on gross domestic product (GDP) per capita growth (e.g., Pellegrini, Terribile, Tarola, Muccigrosso and Busillo 2013), investments per capita (Becker, Egger and von Ehrlich 2013) or regional research and development activity (Ferrara, McCann, Pellegrini, Stelder and Terribile 2016).

economic cohesion. This and other studies use data on the distribution of expenditure across NUTS-2 region and themes including *Social Infrastructure, Rail, Road* or *Cultural Heritage and Tourism* (e.g., Percoco 2013, Ferrara et al. 2016). Second, regional heterogeneity as a determinant of policy effectiveness has become a topic of interest and is often modelled by a region's capacity to take advantage of regional funds. Becker et al. (2013) indicate that human capital and institutional quality matter for the effectiveness of Objective 1 funds in terms of their effect on GDP per capita growth and investment. Institutions are confirmed as influencing factor for the success of regional policy by other authors as well (e.g., Cappelen et al. 2003, Bachtler, Mendez and Oraže 2014, Rodríguez-Pose 2013). Third, Becker, Egger and Von Ehrlich (2012) take the amount of regional funds expenditure spent in a region into account and conclude that there is a maximum efficient level of funds and paying more does not increase the effectiveness any more (see also Kyriacou and Roca-Sagalés 2012, Rodríguez-Pose and Garcilazo 2015). Fourth, Becker, Egger and von Ehrlich (2016) and Bachtrögler (2016) analyse the effects of structural funds (on income growth) in lagging regions over time and in the context of the economic and financial crisis starting in 2007. The latter finds that the effectiveness of regional policy in terms of increasing GDP per capita growth appears to decrease in the crisis compared to former periods when controlling for regional structural characteristics. Barone, David and de Blasio (2016) also take the time dimension into account and show that regional policy effects are not persistent over time. Finally, spatial spillovers are considered to play a role for cohesion policy effectiveness (Dall'Erba and Le Gallo 2008, Breidenbach et al. 2016, Maynou, Saez, Kyriacou and Bacaria 2016).

All studies named above are based on data at the regional or local level. Turning to the beneficiaries as observation unit, De Zwaan and Merlevede (2013) evaluate the effects of Objective 1 and Objective 2 payments in the programming period 2000-2006 in 25 EU member states on productivity and employment growth of firms in treated and non-treated regions. However, they do not use actual recipients of regional funds but compare all manufacturing firms (available in the ORBIS database) located in treated regions with the manufacturing firms in non-treated regions. Additional firm-level analyses are available for sub-national geographical units (Bernini and Pellegrini 2011) and certain types of regional funds (Hartsenko and Sauga 2012). However, to the best of the authors' knowledge there is no EU-wide study of regional policy using data on actual beneficiaries.

As we overcome this lack of data with the new database, this paper contributes to the literature by giving the first detailed insights on actual beneficiaries (and projects) of regional policy in 25 EU member states between 2007-2013. That allows to identify regional funds allocation patterns across regions and countries (see Sections 4 and 5). A combined analysis of the projects' theme, firm-level characteristics of corresponding beneficiaries and the size of the projects' expenditure may help to explain heterogeneous effects of regional funds allocation found in the literature, and thereby, lead to important policy implications.



### 3 A novel dataset

The dataset provides a detailed description of over two million projects co-funded by the structural funds and the Cohesion Fund. In the following, we present the content of this cross-sectional data for the MFF 2007-2013 and compare the sum of project values with official data on regional funds by DG REGIO.

#### 3.1 Content of the dataset

The European Commission’s Directorate-General for Regional and Urban Policy (DG REGIO) provides a collection of links to national or regional websites that make this information available.<sup>11</sup> Unfortunately, the degree of detail of these lists’ content as well as their structure vary significantly across countries, regions and even operational programmes. Moreover, the documents are provided in mostly national languages, different data formats and using non-standardised definitions. The collected lists of beneficiaries of each operational programme contain project and financial data, whereby the extent (and degree of detail) of information issued by the particular managing authority varies significantly between the OPs.<sup>12</sup> The information on beneficiaries is augmented with data from the ORBIS business database by Bureau van Dijk.

The set of variables can be grouped into three blocks, namely, i) project information, ii) funding (co-financing) information, and iii) business characteristics of the beneficiary retrieved from ORBIS. Table 1 shows the project data using the “Top five” projects in terms of highest total project values as examples. It includes the country and NUTS region in which the project is carried out according to the OP and list of beneficiaries, respectively.<sup>13</sup> Moreover, it covers the co-financing fund, the objective and corresponding OP to which the project is assigned. As already noted, the dataset includes projects co-funded by the ERDF, the ESF and the CF, under the objectives of *Convergence* and *Regional Competitiveness and Employment*. Next, a project name or description, the start and end date as well as the theme of the project are included. The theme is not reported by all managing authorities, which is why we classify the remaining projects according to available project information using supervised text classification (see Subsection 3.2 for a detailed description). The fifteen themes are: i) *Capacity Building*, ii) *Culture, Heritage and Tourism*, iii) *Energy*, iv) *Environment*, v) *Human Capital*, vi) *Innovation & Research and Technological Development (RTD)*, vii) *IT Services and Infrastructure*, viii) *Labour Market*, ix) *Other SME and Business Support*, x) *Other Transport*, xi) *Rail*, xii) *Road*, xiii) *Social Inclusion*, xiv) *Social Infrastructure*, and xv) *Urban and Territorial Dimension*.

---

<sup>11</sup>See [http://ec.europa.eu/regional\\_policy/en/atlas/beneficiaries/](http://ec.europa.eu/regional_policy/en/atlas/beneficiaries/).

<sup>12</sup>In Appendix A.1, we provide an overview of all OPs together with information on each one covered in our database.

<sup>13</sup>If that information is not evident in the managing authority’s report, we consider the NUTS-2 region of the beneficiary in ORBIS if available.

The second group of variables is presented in Table 2 and describes the funding structure. It contains the committed co-financing amounts by the EU (C\_EU) and the national government (C\_NAT) as defined in the beginning of the MFF 2007-2013.<sup>14</sup> If the amount borne by the firm itself (ineligible cost, Inelig<sup>15</sup>) is reported, it is added to the sum of co-funding commitments in order to arrive at a total value for project  $i$ :

$$\text{Total value}_i = \text{C\_EU}_i + \text{C\_NAT}_i + \text{Inelig}_i \quad (1)$$

In addition to the commitments, values actually paid out by regional (Paid\_EU) or national (Paid\_NAT) public funds are available for a subset of observations. If only the actually paid-out value is declared, the total project value represents the sum of regional or national payments:

$$\text{Total value}_i = \text{Paid\_EU}_i + \text{Paid\_NAT}_i \quad (2)$$

Furthermore, the declaration date refers to the time of reporting of the respective list of beneficiaries. In case it was not available we use the date of download.<sup>16</sup>

The third information block relates to the beneficiary. This data is produced by a matching exercise (using the name of the firm and its home country) with the ORBIS business database. We are aware of several shortcomings of this database (Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych and Yesiltas 2015), however, it represents the most comprehensive and accessible international business database.<sup>17</sup> Table 3 presents the ORBIS data again for the “Top five” projects: First, the firm’s name in ORBIS and its location. Second, its founding year and information on the industry in which it operates (NACE Rev.2 industry and code). Third, a size classification by ORBIS that is based on at least one of the following variables: The firms’ number of employees, total assets, operating revenue and whether it is listed at the stock exchange.<sup>18</sup> Besides the firms’

<sup>14</sup>Appendix A.1 states the degree of detail in which the financial information on projects is provided in the lists of beneficiaries.

<sup>15</sup>Article 56 in European Council (2006b) states the definition of project expenditure that is eligible for co-funding.

<sup>16</sup>Payments corresponding to the MFF 2007-2013 have been transferred until the end of 2015, however, not all have been reported at the time of data collection.

<sup>17</sup>Information on ORBIS can be retrieved from <http://www.bvdinfo.com/en-gb/our-products/company-information/international-products/orbis>. ORBIS, a global database, includes large parts of the European AMADEUS business database. One difference is a longer data coverage in AMADEUS for some countries, however, ORBIS keeps firms for a longer period than AMADEUS after they have not reported data for five years (Kalemli-Ozcan et al. 2015). For the purpose of this study, this characteristic represents an advantage of using ORBIS.

<sup>18</sup>Very large company: Listed, operating revenue greater or equal to EUR 100 million or total assets greater or equal to EUR 200 million or 1,000 or more employees. Large company: Operating revenue greater or equal to EUR 10 million or total assets greater or equal to EUR 20 million or 150 or more employees. Medium-sized company: Operating revenue greater or equal to EUR 1 million or total assets greater or equal to EUR 2 million or 15 or more employees. Small company: Companies that do not fall in any of the other categories. As companies are classified as small ones if no data points are available, we do not consider this variable in regressions in Section 5. However, comparing the distribution of the ORBIS size category with the number of employees (more than 15, 150 or 1000) shows a similar picture.

Table (1) – Content of database – Part I: Project information

Country	NUTS	Type of fund	Objective	OP	Project Name	Project Start	Project End	Theme
PL	PL12	ERDF or CF	Convergence	2007PL161PO002	II linia metra w Warszawie [...]	2007-01-01	2015-06-30	Road
ES	ES	ERDF or CF	Convergence	2007ES161PO009	GP LEVANTE. TRAMOS: [...]	-	-	-
IT	ITG1	ERDF	Convergence	2007IT161PO011	Raddoppio Palermo C.le [...]	-	-	-
PL	PL42	ERDF or CF	Convergence	2007PL161PO002	Budowa Terminalu Regazyfikacyjnego	2008-01-10	2015-11-30	Energy
RO	RO11	ERDF or CF	Convergence	2007RO161PO003	Reabilitarea liniei de cale ferată Brasov	-	-	Rail

Notes: "Top five" projects in terms of total project value. Source: Lists of beneficiaries provided publicly by managing authorities (see Appendix A.1), own classification for theme. OP: Operational programme. Some OPs are co-funded by the ERDF and the CF.

Table (2) – Content of database – Part II: Funding information

Beneficiary	Curr.	Committed EU	Committed national	Committed sum	Committed beneficiary	Total project value	Paid-out EU	Paid-out national	Paid-out sum	Declaration date
1 City of Warsaw	EUR	-	-	1,152	-	1,508	774	-	774	2015-06-30
2 ADIF. Alta Velocidad	EUR	-	-	1,154	-	1,154	-	-	931	2015-10-13
3 RFI SPA	EUR	281	-	281	-	1,077	193	-	257	2015-02-28
4 Polskie LNG S.A.	EUR	-	-	367	-	906	-	-	209	2015-06-30
5 Compania Națională de Căi Ferate CFR SA	EUR	544	96	640	213	853	-	-	-	2015-07-31

Notes: "Top five" projects in terms of total project value. Source: Lists of beneficiaries provided publicly by regional authorities, own calculation (sums), in million Euros. *Committed sum* (sum of committed funding) is the sum of columns *Committed EU* (committed EU co-funding) and *Committed national* (committed national co-funding). The total value adds *Committed beneficiary* (the projected cost for the beneficiary itself, also referred to as ineligible cost) to *Committed sum*. For some operational programmes or projects, we do not know committed values but paid-out amounts by the EU (*Paid-out EU*) or national authorities (*Paid-out national*) or their sum *Paid-out sum*, respectively. If this is the case, the variable *Total value* is equal to *Paid-out Sum*. Curr.: Currency. In Bulgarian, Croatian (partly), Czech, Danish, Polish, Romanian, Swedish (partly) and British lists of beneficiaries, the funding amounts are reported in national currency. Therefore, we use the respective average exchange rates across the years from 2007 to 2013 to calculate the Euro amounts.

Table (3) – Content of database – Part III: Business information (ORBIS)

ORBIS Name	City	Founding year	NACE Rev.2 sector	NACE code	ORBIS Size category	No. of employees last avail. year	Sales last avail. year	No. of entities in corporate group
1 MIASTO STOLECZNE WARSZAWA	WARSZAWA	2002	O	8411	Medium-sized company	35	-	36
2 ADIF. Alta Velocidad	MADRID	-	-	-	Small company	-	-	-
4 POLSKIE LNG S.A.	SWINOUJSCIE	2007	D	3521	Very large company	-	-	1156
5 COMPANIA NAȚIONALĂ DE CĂI FERATE "CFR" S.A.	CLUJ-NAPOCA	2003	H	5221	Small company	-	-	-

Notes: "Top five" projects in terms of total project value. Source: ORBIS business database. NACE Rev. 2 sector and NACE code (subcategory): Statistical classification of economic activities in the European Community (Eurostat 2008). Sectors: A: Agriculture, forestry, fishing; B: Mining, quarrying; C: Manufacturing; D: Energy; E: Water supply, sewerage, waste management; F: Construction; G: Wholesale; H: Transportation; I: Accommodation and food services; J: Information and communication; K: Financial services; L: Real estate services; M: Professional, scientific and technical activities; N: Administrative and support service activities; O: Public administration, defence, social security; P: Education; Q: Human health and social work; R: Arts, entertainment, recreation; S: Other services; T: Household services. The last available year refers to the last period for which ORBIS contains information on each firm.

number of employees and sales volume, the database contains information on whether a firm belongs to a corporate group and, if so, on the number of entities in this group.

Table 4 shows all variables and their coverage in the database, i.e., the share of all observations for which it is available. The OP, its location, the concerned fund, the objective as well as the name of the project and the beneficiary is available for each project. Moreover, the dataset provides at least the total project value for each observation. 39% of the observations could be matched with ORBIS.

**Table (4)** – List of variables and coverage in dataset

<b>Variable</b>	<b>Coverage</b>
<b>Project information</b>	
Country (NUTS-0)	100 %
Location (NUTS-2, NUTS-1)	97 %
Type of fund (ERDF, ESF, CF)	99 %
Objective (unique)	75 %
Operational programme	99 %
Name or description of project	97 %
Detailed project description	39 %
Project start date	19 %
Project end date	14 %
Theme	83 %
Name of the beneficiary	100 %
<b>Funding information</b>	
Currency	100 %
Amount of EU support - committed	15 %
Amount of national support (co-financing) - committed	5 %
Non-eligible cost paid by beneficiary	2 %
Total project value	100 %
Amount of EU support - paid out	15 %
Amount of national support (co-financing) - paid out	3 %
Declaration date	100 %
<b>Business information (ORBIS)</b>	
Name of the firm	39 %
Detailed address information	36 %
Founding year	32 %
NACE Rev.2 industry (main category)	33 %
NACE Rev.2 code (4 digits)	33 %
ORBIS size category (four categories)	38 %
Number of employees (last available year)	18 %
Volume of sales (last available year)	22 %
Number of companies in corporate group	17 %

Notes: Total number of observations: 2,055,375. The coverage denotes the share of observations for which each variable is available. The location refers to the corresponding OP or, if the OP cannot be assigned to a region, we use the matched location information from ORBIS.

### 3.2 Missing Themes

As described in Section 2, the projects can be categorised into fifteen themes as defined by DG REGIO. Since not all managing authorities publish these themes in their lists of beneficiaries but most of them provide a project description and a project name, we employ supervised text classification to predict the missing project themes.

In order to train the classification algorithm (classifier) we use theme labels reported by some managing authorities, augmented with manually assigned theme labels. Since some of the project descriptions are given in a language other than English, we first use `Google Cloud Translation API` to translate the project descriptions and project names into English. Although we cannot quantify how many errors have been introduced during the translation process, we report the overall accuracy of the classification, where some part of error is attributed to translation errors. The records which cannot be translated are left unchanged. Then we remove those observations where the project name and description together have fewer than 30 characters. Thus, we overall use 1,698,191 projects (82.62% of all observations) and 588,713 labeled projects (28.64% of all observations) to train and evaluate the classifier.

Choosing an appropriate classifier is a non-trivial task. We find the recently published `fastText`<sup>19</sup> library (Joulin, Grave, Bojanowski and Mikolov 2016) to perform well on our dataset in terms of the performance metrics precision, recall and accuracy.<sup>20</sup>

In text classification, it is often desirable not to use the words of a text directly for estimation but to first map the text into a vector space with a much lower dimension.

The `fastText` library which uses a single hidden layer neural network can be applied for text classification and to learn the vector representations of words. We outline the basic idea of the model in two steps: First, the data is mapped into a low dimensional vector space (i.e., each sentence is mapped into a numeric vector) in such a way that similar texts have similar vector representations. Second, multinomial logistic regression is used to predict the labels.

For the evaluation of the classification we use the *ten*-fold cross validation method (Stone 1974). In *k*-fold cross validation the data is randomly split into *k* parts, where *k* - 1 parts are used for training the model and the remaining part is used for model evaluation. To test the model on all available data, this process is typically repeated *k* times. In Figure 1 we report the confusion matrix (a special type of contingency table) for one of the *ten*-cross folds, specifically, we choose to report the results for the cross fold with the lowest accuracy. The rows indicate the true known themes and the columns the predicted themes. Let *X* be the confusion matrix, then precision, recall and accuracy are defined as follows:

$$\text{precision}_j = \frac{X_{jj}}{\sum_i X_{ij}}, \quad \text{recall}_i = \frac{X_{ii}}{\sum_j X_{ij}}, \quad \text{accuracy} = \frac{\sum_i X_{ii}}{\sum_{ij} X_{ij}}. \quad (3)$$

Therefore, precision, recall and accuracy are equal to one if all the predicted themes are correct (the confusion matrix has only entries in the main diagonal) and zero if all the

<sup>19</sup><https://github.com/facebookresearch/fastText>

<sup>20</sup>Due to the size of our data some of the alternative methods are overly time-consuming or run out of memory and are therefore excluded from further considerations.

predicted themes are wrong (the confusion matrix has no entries in the main diagonal). The rectangles in Figure 1 visualise the row and column percentages of the confusion matrix. The width of the rectangles represents the row percentages and the height of the rectangles the column percentages. Therefore, the width of the rectangles of the diagonal corresponds to the recall, the height of the rectangles of the diagonal to the precision and the volumes of the rectangles of the diagonal to the squared G-measure; ( $G\text{-measure}_i = \sqrt{\text{precision}_i \times \text{recall}_i}$ ).

Known Labels	Predicted Labels														
	Capacity Building	Other transport	Rail	Road	Social Inclusion	Social infrastructure	Urban and territorial dimension	Culture heritage and tourism	Energy	Environment	Human capital	Innovation & RTD	IT services and infrastructure	Labour Market	Other SME and Business support
Capacity Building	973			1	5	4	1	2	1	5	29	21	3	4	18
Other transport	1	105		7	1	1	1	3		4		3			4
Rail		4	70	2						2					1
Road	1	7	1	436		1	2	9		3		1			4
Social Inclusion	6			1	3922	11	2	3		35	86	29	8	96	8
Social infrastructure	2			10	9	1940	8	46	1	12	18	15	3	5	39
Urban and territorial dimension	1	1		3		2	641	20	1	13		3	1	1	4
Culture heritage and tourism	2	1		1	8	29	11	1028	2	53	17	14	3	2	33
Energy	2			2	4	42	4	2	1409	21	1	6	1	1	14
Environment	5	1		4	18	16	8	20	13	4281	15	64	7	50	19
Human capital	9	1		1	69	14		26	1	10	12494	41	7	93	38
Innovation & RTD	9			1	31	15	1	8	2	99	32	4575	17	59	382
IT services and infrastructure	3				30	10	3	8	1	11	18	36	1198	16	91
Labour Market	2			1	139	5		7		48	112	56	9	7082	16
Other SME and Business support	5	1	1	6	9	58	4	30	10	18	38	258	31	1	15252

**Figure (1)** – Confusion matrix of the theme prediction.

Notes: The numbers in the figure show the results of the classification on the held out data set. The width of the rectangles of the diagonal give the recall and the height of the rectangles of the diagonal give the precision

In the confusion matrix, we see that, given the true theme *Other Transport*, in 105 cases the model is able to predict the true label and in seven cases the model predicts the theme *Road*. That the themes *Other Transport* and *Road* are hard to distinguish can also be seen in the word cloud (Figure 2), which shows that in both themes the words “Road” and “Construction” occur very often. The word cloud shows the 50 most frequent words

in the project names and descriptions for each theme, the size of the words increases if a word occurs more often. Therefore, the word cloud can give insights about which words are most commonly used in the project description for a given theme. E.g., in projects assigned to the theme *Energy*, there are many projects related to heating (boiler), biomass and water systems.



**Figure (2) – Word cloud of the themes.**  
Notes: The figure shows the most frequent words for each theme. The size of the words is dependent on their frequency.

Overall, we obtain an average classification accuracy of 0.94. However, for completeness we note that there are duplicates in the trainings and test set. Accounting for this, the average accuracy without duplicates is 0.90.

In order to make the classification results easily reproducible, we assemble the **R** (R Core Team 2016) package `fastTextR` that contains an interface to the `fastText` library and is available at CRAN<sup>21</sup>.

### 3.3 Comparison of the dataset with official data

We assess the validity of the assembled data by checking for outliers and plausibility (see Section 4) and comparing its dimension with official data on regional funds assignment published by DG REGIO. Table 5 shows the sum of total project values in the database

<sup>21</sup><https://CRAN.R-project.org/package=fastTextR>

**Table (5)** – Comparison with official DG REGIO data (in million Euros)

Country	Obj.	Sum of total project values in database	Maximum EU co-financing rate	Maximum EU co-financing	DG REGIO Sum of allocated funds
AT	1	232	75%	174	102
	2	2,063	50%	1,032	473
BE	1	1,849	75%	1,387	451
	2	4,226	50 %	2,113	546
BG	1	2,673	85%	2,272	6,313
	2	0	85%	0	0
CZ	1	22,717	85%	19,309	17,555
	2	408	85%	347	206
DE	1	18,445	75%	13,834	10,599
	2	16,798	50%	8,399	4,530
DK	1	0	75%	0	0
	2	491	50%	246	250
EE	1	1,848	85%	1,571	2,836
	2	0	85%	0	0
ES	1	28,415	80%	22,732	12,415
	2	10,354	50%	5,177	3,792
FI	1	0	75%	0	0
	2	6,754	50%	3,377	998
FR	1	7,300	75%	5,475	1,876
	2	35,042	50%	17,521	5,289
HR	1	1,592	85%	1,353	377
	2	0	85%	0	0
IE	1	0	75%	0	0
	2	6,193	50%	3,097	414
IT	1	29,443	75%	22,082	17,955
	2	16,533	50%	8,267	3,171
LT	1	9,186	85%	7,808	5,556
	2	0	85%	0	0
LU	1	0	75%	0	0
	2	102	50%	51	24
LV	1	7,171	85%	6,095	3,894
	2	0	85%	0	0
MT	1	948	85%	806	708
	2	0	85%	0	0
NL	1	0	75%	0	0
	2	6,263	50%	3,132	866
PL	1	127,045	85%	107,988	57,138
	2	0	85%	0	0
PT	1	23,880	85%	20,298	10,538
	2	1,912	85%	1,625	642
RO	1	36,888	85%	31,355	16,010
	2	0	85%	0	0
SE	1	0	75%	0	0
	2	2,571	50%	1,286	925
SI	1	5,180	85%	4,403	2,641
	2	0	85%	0	0
SK	1	13,640	85%	11,594	8,144
	2	132	85%	112	79
UK	1	5,004	75%	3,753	1,762
	2	9,588	50%	4,794	3,527

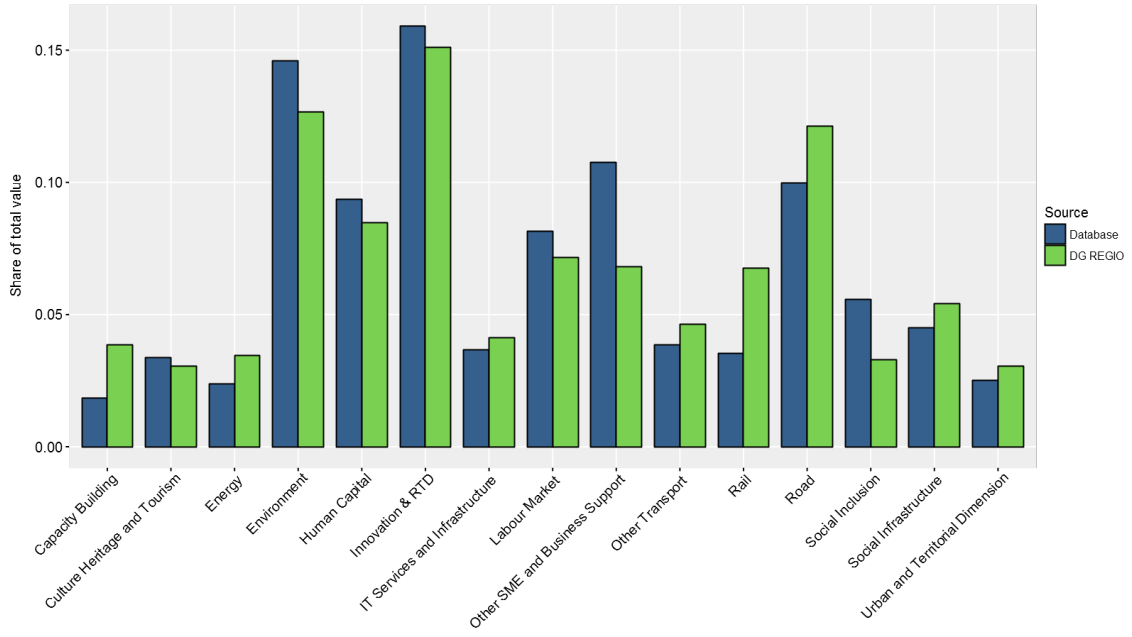
Notes: Source: DG REGIO, Open Data Portal for European Structural and Investment Funds. Obj. (Objective): 1 stands for *Convergence*, 2 for *Regional Competitiveness and Employment*. From our database, we do not consider funds where a unique assignment of an objective is not possible. From DG REGIO data, we do not consider commitments assigned to multiple objectives (Multi Objective).



per country and objective, excluding projects which cannot be assigned to a specific objective. As noted above and in Appendix A.1, the total values in the database in general do not only consist of committed values by the EU. Multiplying the total project values with the maximum co-financing rate per country and objective (see Section 2.1) results in the highest amount the EU should provide. If the official committed value, given in the last column of Table 5, is lower or equal to the maximum EU's co-funding, we expect the sum of total project values in our database to be plausible. The latter is true for the large majority of member states. For Bulgaria, lists of beneficiaries are available for only less than a third of their operational programmes. The Estonian source is an online database which might not yet contain all projects. For Denmark, the gap may arise due to the total project value summing up paid-out and not committed amounts in the database.

Next, we compare the distribution of regional funds among co-financing instruments and themes as reported in the database and by DG REGIO. First, 55% of the sum of total project values correspond to projects co-funded by the ERDF (unequivocal classification) while according to data by DG REGIO around the same amount (58%) of structural funds and the Cohesion Fund is transferred via the ERDF. For ESF, the share of the total project values amounts to 22% which is exactly the same one as in official data. As several operational programmes are co-funded by the ERDF and the CF and there is no more detailed information reported in the lists of beneficiaries, we are able to attribute only about 4% of total project values in the database to the Cohesion Fund. For about 20%, we cannot clearly say which one of the funds is the supporting one. Following DG REGIO data, about 20% of structural and Cohesion funds commitments are settled via the Cohesion Fund, i.e., it is likely that the major part of the not uniquely assigned total values in the lists of beneficiaries can be attributed to the Cohesion Fund.

Regarding the distribution of funds and total project values across the fifteen themes (project categories), our database proves to be consistent with official data (Figure 3). The highest project expenditure is dedicated to *Innovation & RTD* and *Environment* according to the new dataset and as also reported by DG REGIO.

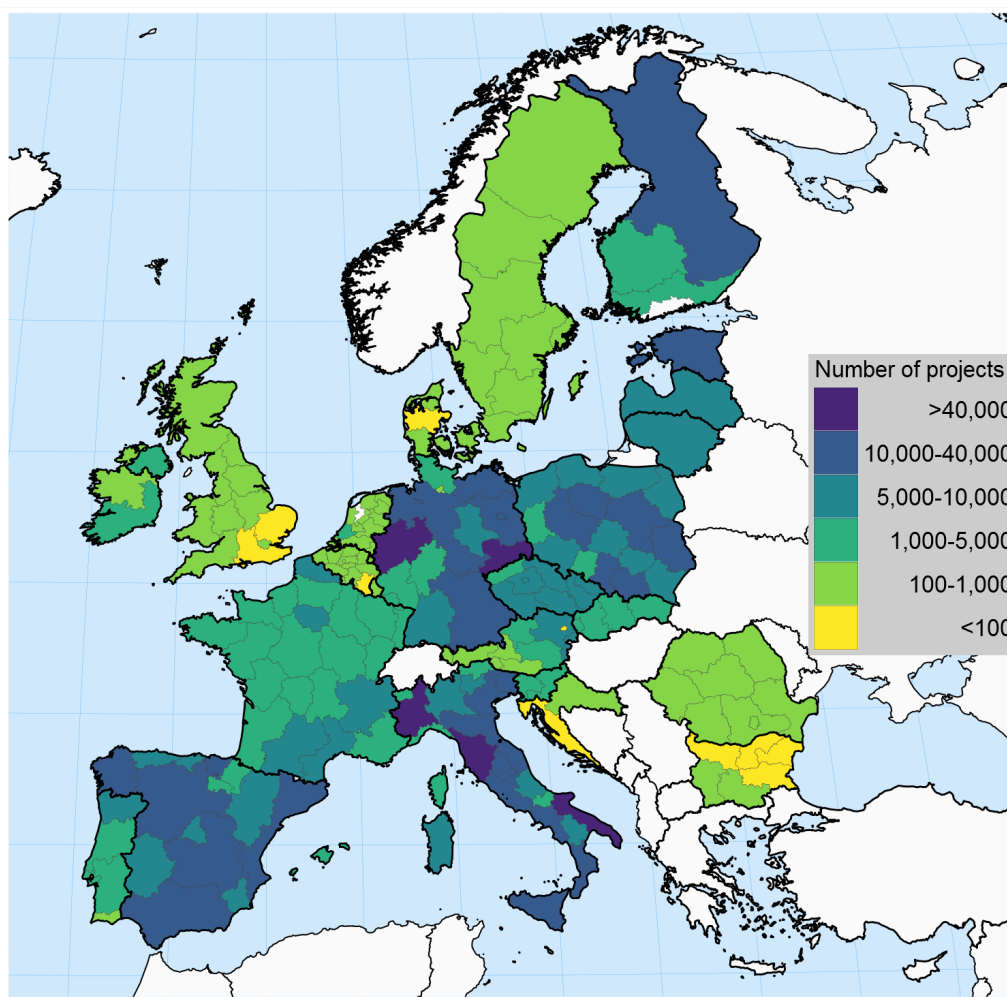


**Figure (3)** – Comparison of distribution of committed regional funds with DG REGIO data: Shares (%) of sum of total project values.  
 Notes: Source DG REGIO: “ERDF/ESF/CF Priority theme overview 2007-2013” downloadable at [http://ec.europa.eu/regional\\_policy/en/policy/evaluations/data-for-research/](http://ec.europa.eu/regional_policy/en/policy/evaluations/data-for-research/).

## 4 Descriptive statistics

This section presents a descriptive summary of the database by providing insights into the total volumes of projects co-funded by the regional funds. We study their distribution across regions followed by an analysis across characteristics of the projects and the beneficiaries.<sup>22</sup>

### 4.1 Regional distribution



**Figure (4)** – Number of projects (observations) per region.

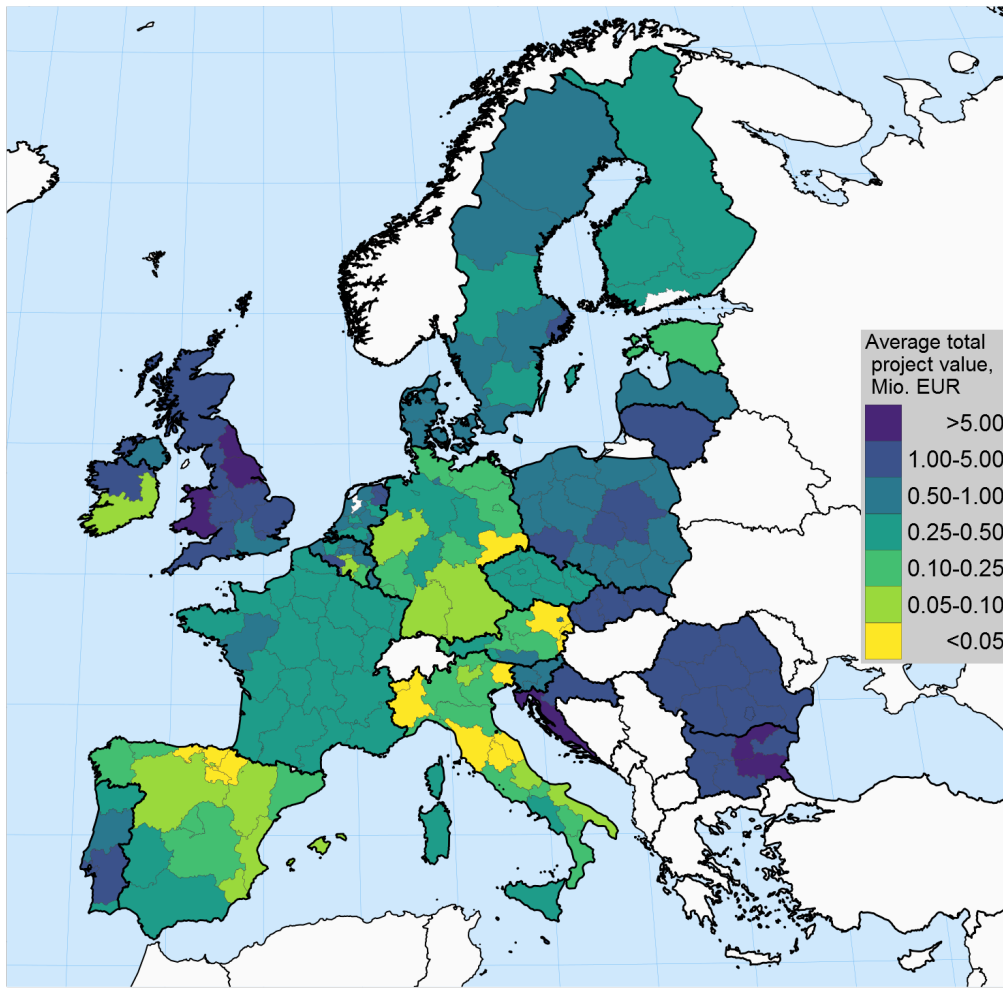
Notes: Min.: 32, 1st Qu.: 569, Median: 3,220, Mean: 6,526, 3rd Qu.: 8,326, Max: 85,420. While the lists of beneficiaries published by the managing authorities do not permit to assign projects to regions Helsinki, Finland, as well as Dutch regions, the figure shows observations for Dutch regions that are matched with ORBIS which contains information on the location of firms.

The regional number of projects realised across Europe varies widely (Figure 4), between 32 in South-East England, UK, and over 85,000 in Puglia, Italy. There are regions with many projects like Puglia, Italy, or North Rhine-Westphalia, Germany, which are typically characterised by a high number of projects related to the themes *Labour Market* or *Human Capital*. Both themes are usually associated with smaller project amounts (see

<sup>22</sup>Note that all subsequent figures, except the first, show total project values as defined in Section 3.

Figure 8). Other regions, like South-East England, UK, or Vienna, Austria, have few but, on average, large projects. Some of the remaining variation in the number of projects can be explained by poor data availability. In Croatia, not all projects could be assigned to a NUTS-2 region due to the reporting of the managing authority. For Bulgaria we only have information on three out of nine operational programmes. Finally, some regions appear to report only intermediate beneficiaries like institutions on a municipal level. While those institutions apply for the funds, they can redistribute them further or use them to carry out projects for smaller entities. In those cases, the ultimate beneficiaries are not known publicly.

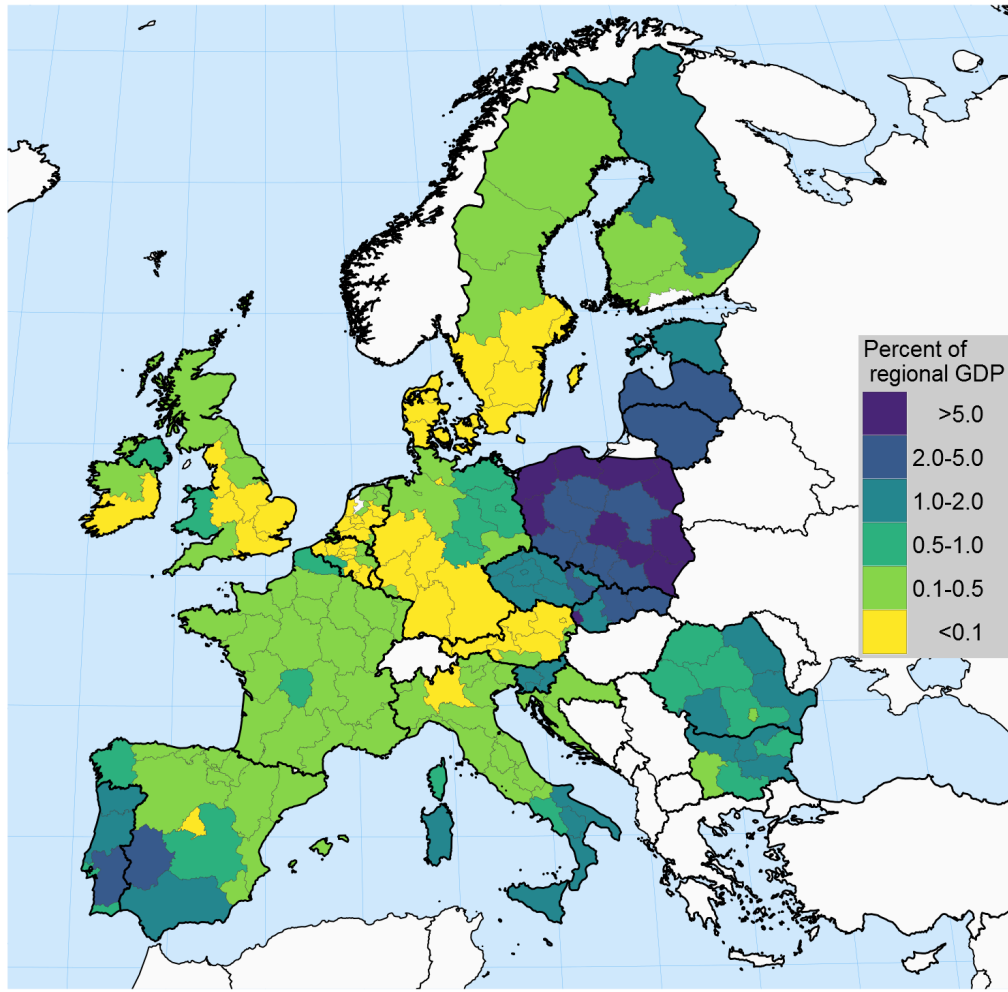
Our dataset consists of over two million projects granted to approximately one million different beneficiaries. That means, on average, every beneficiary receives co-financing for two projects. Only 17% of beneficiaries carry out more than one project, only 3% have more than five and only 1% more than ten. The beneficiary with the most co-financed projects (more than 18,000) is the Spanish *ICEX Espana Exportación e Inversiones*, a governmental institution that promotes (foreign) investments in Spain. The second most (more than 11,000) are carried out in the city of Florence, Italy, the third most (more than 10,000) by the governmental training and orientation section of the region of Tuscany, Italy.



**Figure (5) – Average total project value.**

*Notes: Min.: 0.015, 1st Qu.: 0.222, Median: 0.444, Mean: 1.18, 3rd Qu.: 1.103, Max: 14.100. While the lists of beneficiaries published by the managing authorities do not permit to assign projects to regions Helsinki, Finland, as well as Dutch regions, the figure shows observations for Dutch regions that are matched with ORBIS which contains information on the location of firms.*

Figure 5 shows that regions with fewer projects usually have higher total values per project. There are additional reasons for the regional variation in the average total values such as the distribution of industries, firms of different size and other regional characteristics. Taking those into account simultaneously will be the topic of Section 5.



**Figure (6)** – Sum of total values of projects as share of total regional GDP (sum of 2007-2013).

Notes: Min.: 0.01, 1st Qu.: 0.17, Median: 0.35, Mean: 0.91, 3rd Qu.: 1.10, Max: 8.45. While the lists of beneficiaries published by the managing authorities do not permit to assign projects to regions Helsinki, Finland, as well as Dutch regions, the figure shows observations for Dutch regions that are matched with ORBIS which contains information on the location of firms.

Figure 6 shows the total sum of project values as a share of regional GDP. Since the database covers seven years, we also sum regional GDP over those seven years. This map shows that, in general, convergence regions receive the most relative to their GDP.<sup>23</sup>

In the so-called “Blue Banana”<sup>24</sup> the total value of projects is below 0.1% of regional GDP. In those regions, projects with the highest expenditure tend to be carried out by small and medium-sized firms in the education sector, in public administration as well as in professional, scientific and technical activities. Those projects are mostly aimed at *Regional Competitiveness and Employment* and the themes *Innovation & RTD* as well as *Human Capital*.

<sup>23</sup>We define convergence regions as less developed regions with a GDP per capita below 75% of the EU-25 average.

<sup>24</sup>Populous and usually rich regions that range from England, Belgium and the Netherlands, parts of Germany, Austria to Northern Italy.

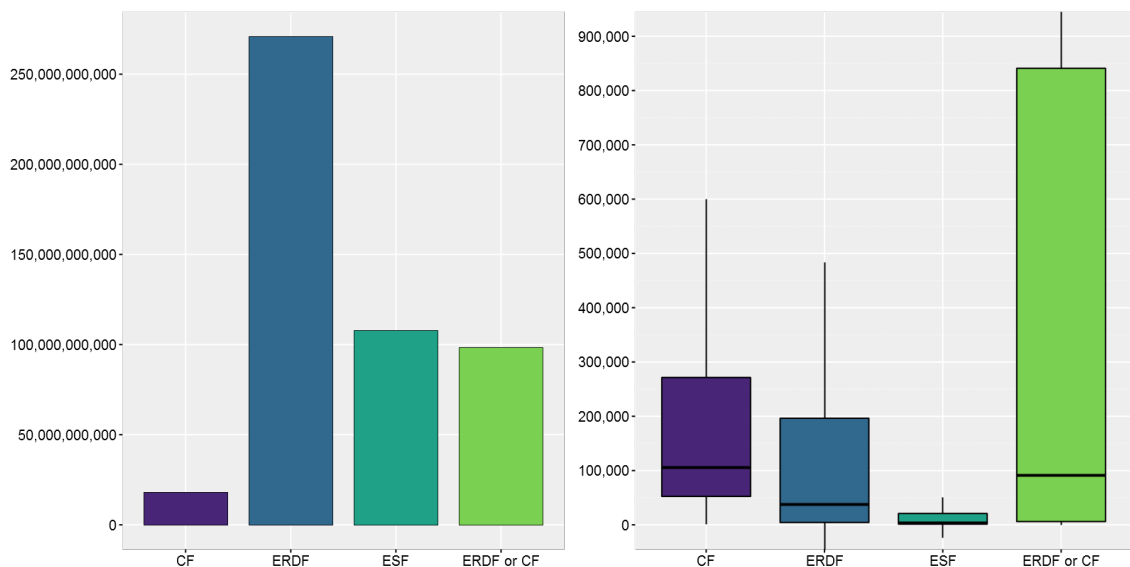
The sum of project values in Scandinavian, French, Northern Italian, Northeastern Spanish regions as well as Scotland and Northern Ireland, UK, ranges between 0.1% and 0.5% of regional GDP. In contrast, regional policy plays a more important role for the Baltics and Poland. There, project values account for up to eight percent of regional GDP. Most of them are *Road* and *Environment* projects carried out by very large firms that operate in public administration and the manufacturing industry. The region with the highest share in regional GDP is Podkarpackie in Southeast Poland with 8.45% or EUR 8.2 billion.

Project values per capita follow a similar overall distribution, where the amount in most regions falls below EUR 500. The region with the highest value per capita is Bratislava, Slovakia, with EUR 12,575 over the course of seven years.

## 4.2 Type of funds, objectives and themes

The size and number of projects varies not only across regions but also across the type of co-financing fund, the corresponding objective and theme. Figures in this subsection show sums of total project values on the left-hand side and distributions of individual project values on the right-hand side. A characteristic of almost all variables in the dataset, is the existence of - sometimes large - outliers at the top.

We first distinguish between the type of fund, i.e. the ERDF, ESF or the CF. Figure 7 shows that project values vary widely between funds.



**Figure (7)** – Sum of total values by fund.

Notes: Left: Sum of total values by type of fund. CF: Cohesion fund, ERDF: European Regional Development Fund, ESF: European Social Fund. Right: Distribution of values per project; dark horizontal line marks the median.

The total values of projects subsidised by the ERDF sum to roughly EUR 270 billion compared to EUR 107 billion from the ESF and EUR 18 billion from the CF. There is a fourth category that cannot be unequivocally assigned to a fund due to missing information in the list of beneficiaries. These are 25,000 projects, which account for EUR 100 billion. A

comparison with official data (see Section 3.3) indicates that those projects are most likely funded by the CF. The right-hand side of Figure 7 shows the median project size across fund types. The median CF project value amounts to about EUR 100,000 compared to EUR 37,000 for the ERDF and EUR 3,200 for the ESF.

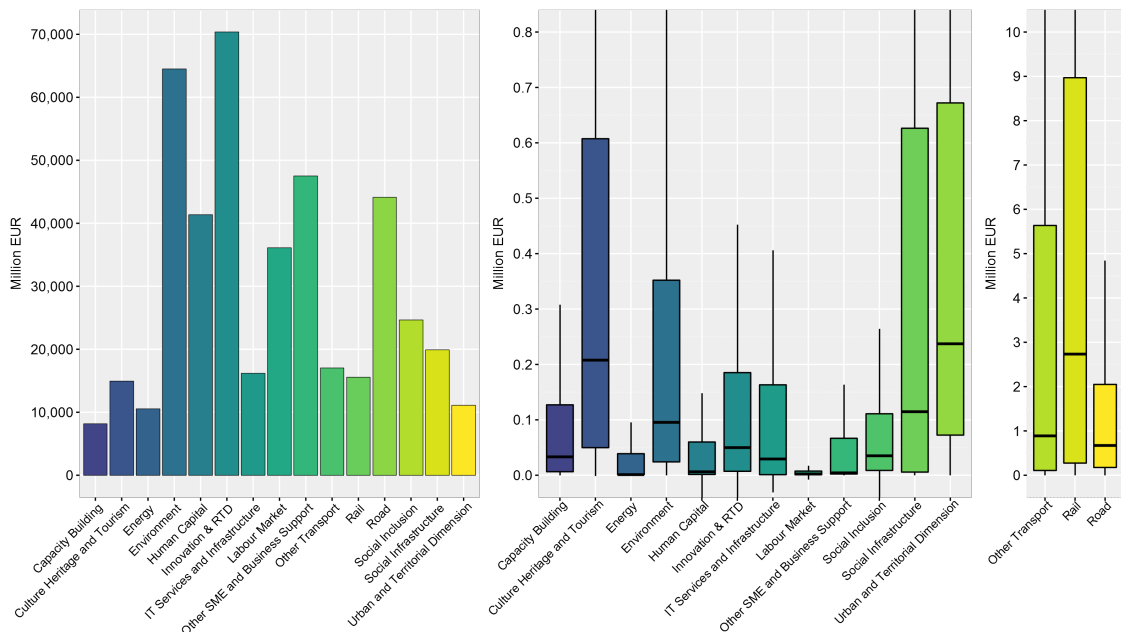
Regarding the objective, Table 6 indicates that while around half of the projects are aimed at *Convergence*, their value is two times higher than *Regional Competitiveness and Employment*. As expected, operational programmes in peripheral regions have a focus on *Convergence* while OPs in richer, more central regions receive more to boost *Regional Competitiveness and Employment*.

**Table (6)** – Objectives

Objective	Total sum of project values in Mio. EUR	Number of projects
Convergence	342,568	723,294
Regional competitiveness and employment	117,721	826,563

Notes: The numbers represent 75% of observations and 93.5% of the total value.

The European Commission defines fifteen themes (see Section 3.2) to classify projects. Most projects related to transportation<sup>25</sup> are characterised by a much larger median project value than the rest, which is why we divide up the graphic for better readability.



**Figure (8)** – Sum of total project values by theme.

Notes: Left: Total project sums by theme. Middle and right: Distribution of values per project; dark horizontal line marks the median. The figure represents 83% of observations and 90% of the sum of total values.

The left-hand side of Figure 8 reveals that the largest sums are committed to projects in categories *Innovation & RTD* (EUR 70 billion), *Environment* (EUR 65 billion) and

<sup>25</sup>We define transportation as *Road, Rail and Other Transportation*.



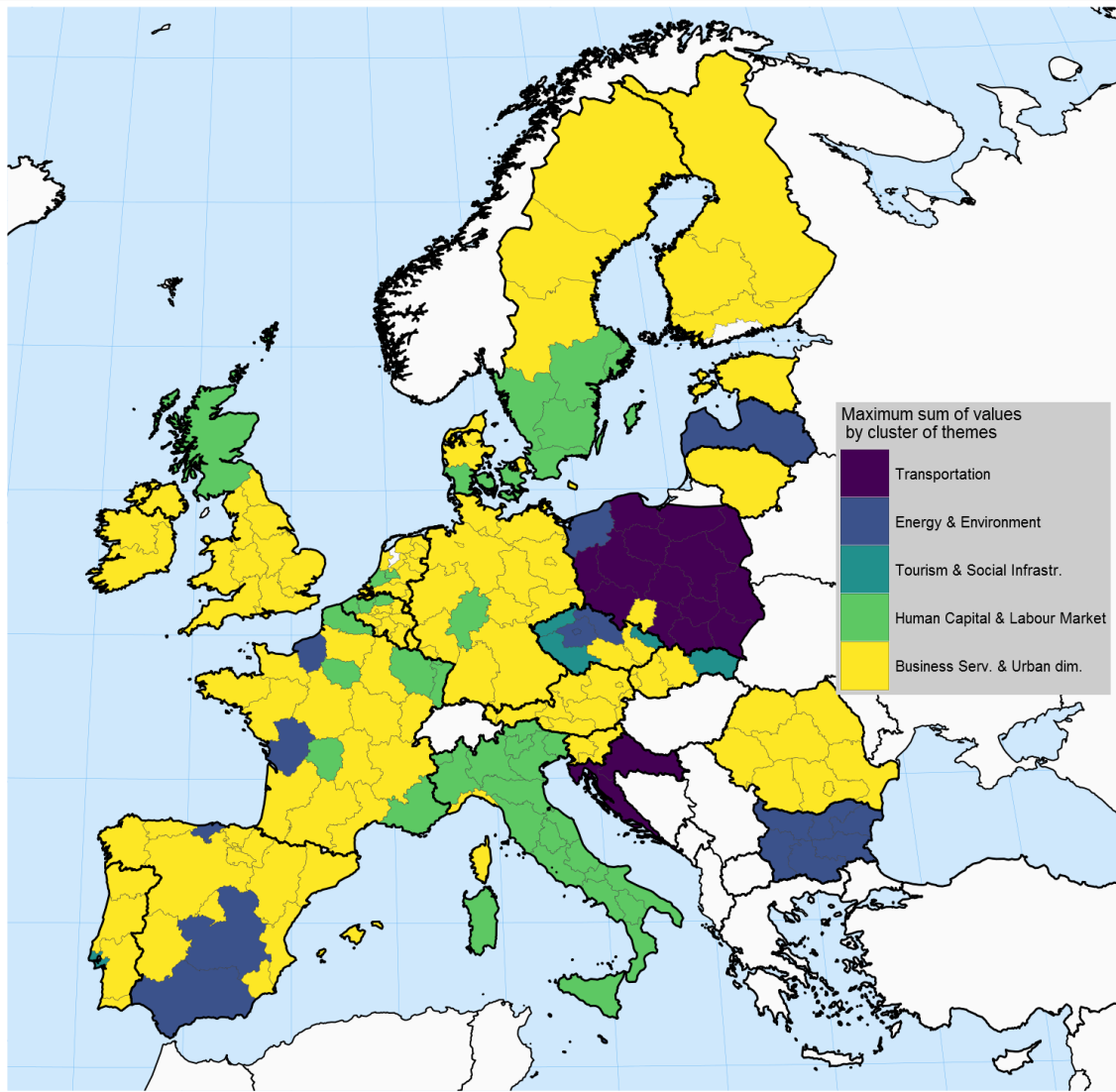
*Other SME and Business Support* (EUR 46 billion), followed by *Labour Market*, *Road* and *Human Capital* projects. The number of observations per theme ranges from 1,058 *Rail* projects to over 600,000 projects related to the *Labour Market*.

The middle and right-hand side of Figure 8 show considerable variation of individual project values across themes. Projects related to *Transportation* (EUR 1-3 million), *Urban and Territorial Dimension* as well as *Culture Heritage and Tourism* (EUR 200,000 each) are the largest. *Human Capital* (EUR 6,000), *Labour Market* (EUR 2,200) and *Energy* (EUR 1,250) projects are the smallest. The word clouds (see Figure 2) give some insights into the type of projects that are associated with each theme.

There is not only large variation in the sums contributed to each theme, also their regional distribution differs. In order to give a first descriptive picture of their distribution, Figure 9 shows the maximum sum of project values per theme in a region. For better readability the map shows five groups of themes instead of all fifteen.<sup>26</sup>

---

<sup>26</sup>The aggregation of the themes is based on a hierarchical clustering, where we use the cosine dissimilarity as distance measure and Ward's method (Ward Jr 1963) for the clustering. More information about the clustering of text data can be found in Feinerer, Hornik and Meyer (2008).



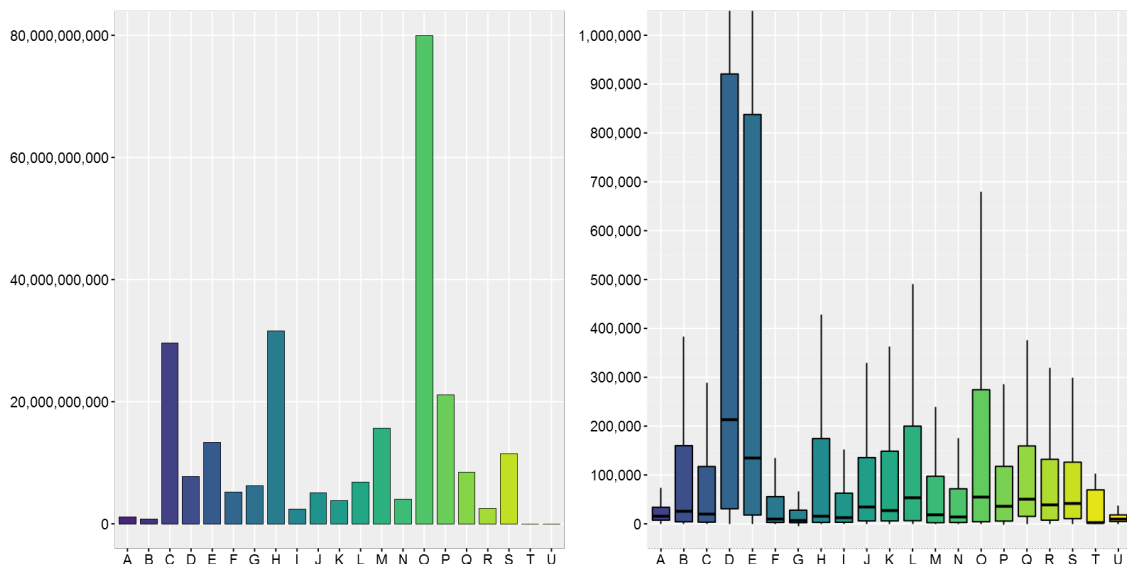
**Figure (9) – Maximum summed total project values by theme group.**

Notes: *Transportation*: Road, Rail, Other Transport; *Social Inclusion* is part of the *Human Capital* group; *Business Services*: Innovation & RTD, Other SME and Business Support, Capacity Building, IT services and infrastructure. While the lists of beneficiaries published by the managing authorities do not permit to assign projects to regions Helsinki, Finland, as well as Dutch regions, the figure shows observations for Dutch regions that are matched with ORBIS which contains information on the location of firms.

In total, the largest sums in Poland and Croatia are related to *Transportation* projects. *Energy* and *Environment* projects are largest in Latvia, parts of the Czech Republic, two French regions, three Spanish regions and a Polish region. Contributions to *Culture*, *Tourism* and *Social Infrastructure* are most pronounced in some Czech regions and East Slovakia. Scotland, South Sweden, almost all of Italy and parts of Germany and France have their largest project sums related to *Human Capital*, the *Labour Market* and *Social Inclusion*. The topmost project sums in the rest of Europe are associated with the fifth category that includes *SME and Business Services*, *IT*, *Capacity Building*, *Innovation & RTD* as well as *Urban and Territorial Dimension*.

### 4.3 Industrial structure and company characteristics

The matched data from ORBIS allows us to go into more detail with respect to characteristics of the beneficiaries. Figure 10 shows the distribution of funds across NACE Rev.2 industries. The matching process and the coverage in ORBIS enables us to assign an industry to one third of observations, i.e., almost 700,000 projects.



**Figure (10)** – Sum of total project values by NACE Rev.2 industry.

Notes: Left: Sum of total project values by NACE Rev.2 industry. Right: Distribution of values per project; dark horizontal line marks the median. A: Agriculture, Forestry, fishing; B: Mining, quarrying; C: Manufacturing; D: Energy; E: Water supply, sewerage, waste management; F: Construction; G: Wholesale; H: Transportation; I: Accommodation and food services; J: Information and communication; K: Financial services; L: Real estate services; M: Professional, scientific and technical activities; N: Administration and support activities; O: Public administration, defence, social security; P: Education; Q: Human health and social work; R: Arts, entertainment, recreation; S: Other services; T: Household services; U: Activities of extraterritorial organisations. The figure represents 33% of observations and 51% of the sum of total values.

Overall, projects with the highest sum of total values are carried out by firms or institutions operating in public administration, defense and social security (EUR 80 billion), transportation and manufacturing (EUR 30 billion each) and education (EUR 20 billion). The right-hand side of Figure 10 shows that firms that operate in the energy and water sector carry out the largest projects on average. Most of those projects have to do with electricity production, air conditioning, water supply as well as waste collection and treatment. When it comes to the number of projects in a specific sector, Table 7 lists the five NACE Rev.2 sectors with the most projects.

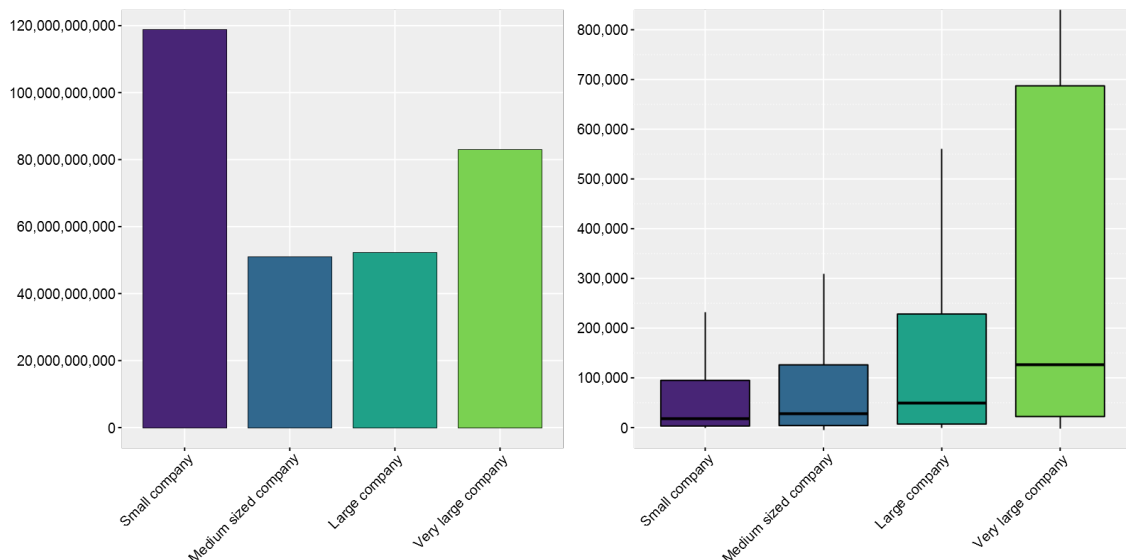
**Table (7)** – Top 5 sectors with the most observations.

NACE industry	NACE 4-digit	Sector description	Number of projects
O	8411	General public administr. activities	83,329
P	8559	Other education n.e.c	49,722
S	9499	Activities of other membership organisations n.e.c.	37,777
P	8542	Tertiary education	23,067
M	7022	Business and other management consultancy activities	17,781

Notes: n.e.c.: not classified elsewhere

Beneficiary firms or institutions also differ in their size. Figure 11 shows the distribution

across four categories of company size defined by ORBIS. That classification is based on the number of employees, revenue, assets and whether they are listed.<sup>27</sup> It considers companies to be small if they are not included in any other category, which could possibly inflate the number of small companies due to missing data. However, labeling the observations ourselves based on the number of employees, we find a very similar distribution of labels.



**Figure (11)** – *Sum of total project values by company size.*  
*Notes: Left: Sum of total project values by company size. Right: Distribution of values per project; dark horizontal line marks the median. The figure represents 39% of observations and 62% of the sum of total values.*

In terms of overall project sums, small firms make up the largest share (EUR 120 billion), followed by very large firms (EUR 80 billion) as well as medium-sized and large firms (EUR 50 billion each). Note that there are fifteen times more small beneficiaries of structural funds than very large ones. They are especially strongly represented in projects co-financed by the ESF. In total, 85% of recipients in the database are small or medium-sized (SME) companies, which reflects the priorities set out in the Community’s strategic guidelines on cohesion. Table 7 seems to be largely representative for different firm sizes too.

## 5 Focus on the single beneficiary

In the following, we investigate the determinants of the total project value of a single project in our database. This sheds light on conditional differences between projects in different regions, with beneficiaries in different industries and of different sizes and supported by different funds with different objectives and themes. To this end, we estimate the following Equation:

$$\ln(TV_i) = \alpha + \psi R_i + \gamma F_i + \delta O_i + \rho T_i + \phi I_i + \beta S_i + \varepsilon \quad (4)$$

<sup>27</sup>See Section 3 for a detailed description.

where vector  $\ln(TV_i)$  represents the logarithm of the total project value of project  $i$ ,  $R_i$  the region in which project  $i$  is located,  $F_i$  the type of fund which supports project  $i$ ,  $O_i$  the objective of the funding for project  $i$ ,  $T_i$  the theme under which project  $i$  is supported,  $I_i$  the industry of the beneficiary and  $S_i$  the size of the beneficiary of project  $i$ . The variables  $R_i$ ,  $F_i$ ,  $O_i$ ,  $T_i$ ,  $I_i$  and  $S_i$  are factor variables and the category with the median coefficient is always the one that is excluded respectively. Thus, the resulting coefficients should be interpreted relative to the project with the coefficient being in the middle of the (conditional) distribution of this variable. We do not include projects with total values which are zero or negative and end up with 544,784 observations for which Equation 4 is estimated.<sup>28</sup>

## 5.1 Regional effects

Comparing the total value of single projects across NUTS-2 regions, while controlling for the type of fund, objective, theme, industry and size of the beneficiary, reveals differences of up to plus and minus 400%. The 20 top and bottom regions are presented in Table 8. The lowest conditional amounts per project can be found in Austria, Estonia, Spain and Germany. While the latter also has one region among the regions with the highest conditional values, the others do not. Apart from Germany, the highest conditional values per project can be observed in the UK, the Netherlands, Malta and Luxembourg.<sup>29</sup> Interestingly, though, the conditional difference is not strongly related to the economic development of the regions, the correlation with GDP per capita is only 0.19 and with GDP -0.07 (averages from 2007-2013) respectively.

**Table (8)** – Regression coefficients with respect to NUTS-2 region

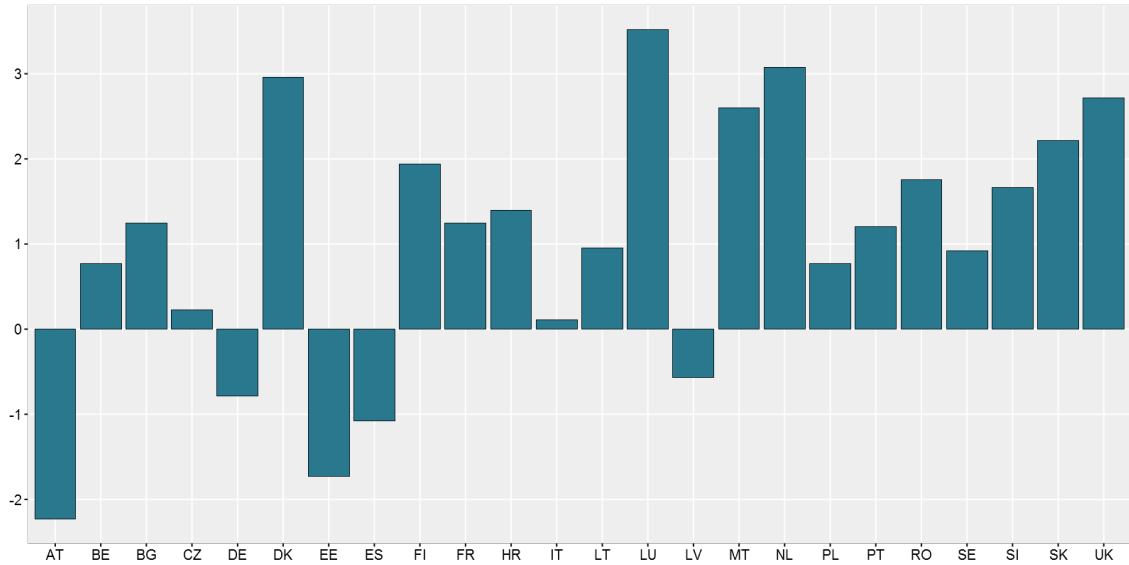
Top 20				Bottom 20			
UKH	+412%	NL12	+228%	ES24	-244%	ES53	-291%
UKE	+381%	UKM	+226%	ES43	-247%	ES42	-294%
UKL	+335%	NL31	+225%	ES21	-253%	ES22	-296%
UKC	+322%	NL32	+219%	AT13	-259%	ES62	-298%
NL13	+313%	MT00	+217%	DED	-259%	ES12	-312%
UKG	+277%	UKD	+217%	ES11	-266%	ES23	-328%
NL11	+273%	NL33	+215%	DEA	-271%	EE00	-329%
NL22	+244%	LU00	+213%	ES64	-281%	ES13	-340%
UKK	+242%	NL21	+208%	ES41	-288%	AT11	-397%
DE6	+230%	ITF3	+201%	DE4	-290%	AT12	-446%

Notes: Coefficients of regressions based on Equation 4 (542,162 degrees of freedom, adj.  $R^2$  of 0.47). Dependent variable: logarithm of the total project value. Other control variables (firm size, industry, fund type and objective, theme) are also included but reported in previous and subsequent tables. Coefficients are relative to median NUTS2-region which is PL32.

Beneficiaries which are similar in size, are in the same industry, receive money from the same fund and within the same theme, but are located in Lower Austria on average have

<sup>28</sup>Only observations matched with ORBIS data can be considered for the econometric exercises in this section.

<sup>29</sup>Curiously, the top six UK regions had a majority vote for “Leave” in the Brexit referendum 2016.



**Figure (12)** – Comparison of residuals of regression of Equation 4 by country.

9.5 times lower project values than beneficiaries located in the East of England. Apparently, in Lower Austria, 90% of projects are smaller than EUR 10,000 and 4,688 projects are smaller than EUR 1,000, whereas in the East of England the smallest project is already EUR 133,168 in size.

Regressing Equation 4 controls for regional fixed effects, thus, any regional influence on the remaining differences between total project values of single projects should be removed. Figure 12 shows the average residuals of a regression similar to Equation 4, but excluding the regional control variables ( $R_i$ ). Hence, the figure shows the (average) unexplained part of the project value when controlling for fund type, objective, theme, industry and firm size split by country. It confirms that the projects, which are similar in many dimensions but the region, with the lowest total values can be found in Austria, Estonia and Spain and the ones with the highest values in Luxembourg, the Netherlands, Denmark and the UK.

## 5.2 Type of fund, objective and themes

Controlling for everything else, the projects with the largest total value can be observed co-funded by the ERDF (left panel of Table 9). They are larger than those co-funded by the ESF by two-thirds and by the CF by one third. This corresponds to the goals of the various funds, as, e.g., the ESF is rather funding smaller projects related to inclusion and adaptability of workers in the labour market and employment, while the ERDF funds are aimed the economic structure and fundamentals of regions (Commission 2017). Furthermore, projects within the *Convergence* objective are larger by around 50% than projects under the *Regional Competitiveness and Employment* objective (right panel of Table 9).

**Table (9)** – Regression coefficients with respect to fund type and objective

<b>Fund type</b>	<b>Coef.</b>	<b>Objective</b>	<b>Coef.</b>
ERDF or CF	+62%	Convergence	+56%
ERDF	+31%	Regional Competitiveness and Employment	0%
CF	0%		
ESF	-35%		

Notes: Coefficients of regressions based on Equation 4 (542,162 degrees of freedom, adj.  $R^2$  of 0.47). Dependent variable: logarithm of the total project value. Other control variables (region, industry, theme, firm size) are also included but reported in previous and subsequent tables. ERDF: European Regional Development Fun, ESF: European Social Fund, CF: Cohesion Fund.

Moreover, large differences in conditional project values arise with respect to the projects' themes. 42% of the highest project values in the database (above EUR 50 million) are associated with the three transport-related themes (*Road, Other Transport, Rail*). It might not be surprising that the projects with themes related to *Labour Market, Human Capital and SME* are, as can also be seen in the unconditional analysis in Section 4, among the projects with the lowest conditional project values, but the *Energy* theme would probably be expected to be among the one with the highest project values.

**Table (10)** – Regression coefficients with respect to theme

<b>Theme</b>	<b>Coef.</b>	<b>Theme</b>	<b>Coef.</b>
Road	+277%	Social Inclusion	-18%
Other Transport	+211%	Capacity Building	-18%
Rail	+199%	Labour Market	-20%
Urban and Territorial Dimension	+104%	Human Capital	-50%
Culture Heritage and Tourism	+73%	Other SME and Business Support	-58%
Innovation & RTD	+35%	Social Infrastructure	-88%
Environment	+33%	Energy	-109%
IT Services and Infrastructure	0%		

Notes: Coefficients of regressions based on Equation 4 (542,162 degrees of freedom, adj.  $R^2$  of 0.47). Dependent variable: logarithm of the total project value. Other control variables (firm size, industry, region, fund type and objective) are also included but reported in previous tables.

However, the conditional differences in project values across themes do not necessarily correspond to the differences across industries. An interesting observation is that the lowest project values can be observed for the *Energy* theme, but the highest project values for the energy industry (Section 5.3). The maximum project value within the *Energy* theme as well as the energy industry is found for the same project which amounts to EUR 906 million (see Table 1- 3). However, the *Energy* theme includes 7,715 projects with a project value of EUR 1,250 each and 64% of the projects have a value below EUR 5,000. Those projects below EUR 5,000 are not conducted by firms in the energy industry but are observed mainly in "Real Estate activities" (59%), "Manufacturing" (11%) and "Wholesale/retail; repair vehicles" (8%).

### 5.3 Industry and company characteristics

The highest conditional project values are attributed to beneficiaries in the energy and water industries (Table 11). The largest three project values of companies in the energy industry are EUR 906 million, EUR 103 million and EUR 100 million and only 37% of the projects have a total value below EUR 100,000. Within the wholesale industry, on the other hand, only 12% of the projects have total values larger than EUR 100,000.

**Table (11)** – Regression coefficients with respect to NACE-industry

Industry NACE Code	Coef.
D - Energy (without Water)	+121%
E - Water supply; sewerage, waste <sup>1</sup>	+70%
Q - Human health and social work activities	+58%
P - Education	+42%
T - Activities of households as employers <sup>2</sup>	+16%
U - Activities of extraterritorial org. and bodies	+13%
S - Other service activities	+12%
M - Professional, scientific and technical activities	+11%
J - Information and communication	+6%
B - Mining and quarrying	0%
R - Arts, entertainment and recreation	-5%
H - Transportation and storage	-11%
O - Public adm. and defence; compulsory social security	-14%
N - Administrative and support service activities	-15%
C - Manufacturing	-15%
K - Financial and insurance activities	-19%
L - Real estate activities	-21%
I - Accommodation and food service activities	-29%
F - Construction	-31%
A - Agriculture, forestry and fishing	-45%
G - Wholesale/ retail; repair vehicles	-51%

Notes: <sup>1</sup> Water supply; sewerage, waste management and remediation activities. <sup>2</sup> undifferentiated goods- and services-producing activities of households for own use. Coefficients of regressions based on Equation 4 (542,162 degrees of freedom, adj. R<sup>2</sup> of 0.47). Dependent variable: logarithm of the total project value. Other control variables (firm size, region, fund type and objective, theme) are also included but reported in previous and subsequent tables.

As expected, the firm size also plays a role for the level of the total project value (Table 12). Overall, we see that the larger the beneficiary is the larger is the total value. However, the difference in project size is surprisingly low, as the size of a project of a very large company is only approximately twice the size of a medium-sized company. However, the average revenue of the very large companies in our sample (EUR 653 million) is approximately 130 times the average revenue of the medium-sized companies (EUR 5 million).

For a smaller subsample of projects, further information from ORBIS is available. Table 13 presents the regression results including the additional data, i.e., estimating the following:

$$\ln(TV_i) = \alpha + \kappa U_i + \lambda E_i + \mu Y_i + \psi R_i + \gamma F_i + \delta O_i + \rho T_i + \phi I_i + \varepsilon \quad (5)$$

where  $U_i$  are the sales of beneficiary of project  $i$ ,  $E_i$  the number of employees of benefi-



**Table (12)** – Regression coefficients with respect to firm size

Size of beneficiary	Coef.	Size of beneficiary	Coef.
Very large company	+98%	Medium-sized company	0%
Large company	+39%	Small company	-27%

Notes: Coefficients of regressions based on Equation 4 (542,162 degrees of freedom, adj.  $R^2$  of 0.47). Dependent variable: logarithm of the total project value. Other control variables (region, industry, theme, objective, fund type) are also included but reported in previous and subsequent tables.

ciary (sales and number of employees corresponds to last available observation in ORBIS) and  $Y_i$  the founding year of beneficiary (in four brackets - before 1950, between 1950 and 1980, between 1980 and 2000, between 2000 and 2010 and after 2010). Furthermore, the regression contains all control variables included in Equation 4, apart from the size of the beneficiary  $S_i$  as this ORBIS variable depends on the number of employees of the beneficiary  $E_i$  or its revenue  $U_i$ . The regression results are based on 208,068 observations.

**Table (13)** – Regression results including additional ORBIS information

Variable	Coefficient	(Std. Error)
Revenues (Mio. EUR)	0.0147***	(0.0032)
No. of Employees (10,000)	0.0122***	(0.0075)
Founding Year (1950-1979)	-0.2529***	(0.0242)
Founding Year (1980-1999)	-0.1184***	(0.0227)
Founding Year (2000-2009)	-0.1701***	(0.0232)
Founding Year (After 2010)	-0.3060***	(0.0257)

Notes: Coefficients of regressions based on Equation 5 (205,792 degrees of freedom, adj.  $R^2$  of 0.51). Dependent variable: logarithm of the total project value. Other control variables (firm size, industry, region, fund type and objective, theme) are also included but not reported. Robust standard errors are in parentheses. \*, \*\*, \*\*\* indicate significance at the 1%, 5% and 10% level.

The results confirm that the size of the beneficiary matters for the value of a single project. Controlling for all the other variables, higher revenues and more employees are associated with significantly higher project values. Furthermore, the age of the company receiving funds does play a role. Younger beneficiaries which were founded after 2000 have projects with significantly lower total values than companies which were founded before that and beneficiaries founded before 1950 receive the highest total values.

## 6 Conclusion

The novel database introduced in this study contains detailed information on over two million projects co-financed by the European Regional Development Fund, the European Social Fund and the Cohesion Fund in 25 EU member states in the programming period 2007-2013. Additional to project information such as project expenditure and a project category (theme), the beneficiaries are matched with the ORBIS business database.

Descriptive analyses show that there are indeed different patterns of regional funds distribution across and within countries, both in terms of project and beneficiary characteristics.

First, we find that the number of projects carried out in a region and the expenditure per project varies significantly across regions. Moreover, we find that most regional funds project expenditure is dedicated to transportation infrastructure, *Energy* and *Environment* projects in lagging regions, whereas the priority in richer regions is given to fostering *Innovation & RTD* and *Other SME and Business Support*.

Conditional on project- and firm-level characteristics, the largest single projects in terms of their total value are co-funded by the ERDF under the *Convergence* objective. They are attributed to infrastructure projects (*Road, Rail and Other Transport*) and, independently of their theme, are conducted in regions of the United Kingdom or the Netherlands. Regarding the beneficiary, firms with higher revenues and more employees carry out projects with higher total value, however, the average project value of (very) large firms is only about twice as high as that of small entities. Moreover, older firms are associated with larger regional projects.

We aim to contribute to the academic and political debate by making a dimension of EU regional policy implementation visible that has not gained much attention until now. The possibility to compare individual project and beneficiary characteristics across heterogeneous regions and countries opens a new strand of research questions and may entail interesting conclusions on more or less effective ways of distributing regional funds. The next step is to use this new database to study and evaluate different dimensions of the EU regional policy in 2007-2013.

## References

- Bachtler, J., Mendez, C. and Oraže, H. (2014), 'From conditionality to Europeanization in Central and Eastern Europe: Administrative performance and capacity in cohesion policy', *European Planning Studies* **22**(4), 735–757.
- Bachtrögler, J. (2016), On the effectiveness of EU structural funds during the Great Recession: Estimates from a heterogeneous local average treatment effects framework. Department of Economics Working Paper Series, 230. WU Vienna University of Economics and Business, Vienna.
- Barone, G., David, F. and de Blasio, G. (2016), 'Boulevard of Broken Dreams. The ends of the EU funding (1997: Abruzzi, Italy)', *Regional Science and Urban Economics* **60**, 31–38.
- Becker, S. O., Egger, P. H. and Von Ehrlich, M. (2012), 'Too much of a good thing? On the growth effects of the EU's regional policy', *European Economic Review* **56**(4), 648–668.
- Becker, S. O., Egger, P. H. and von Ehrlich, M. (2013), '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.
- Becker, S. O., Egger, P. H. and von Ehrlich, M. (2016), Effects of EU Regional Policy: 1989-2013. Working Paper Series, 271. Centre for Competitive Advantage in the Global Economy, University of Warwick.
- Bernini, C. and Pellegrini, G. (2011), 'How are growth and productivity in private firms affected by public subsidy? evidence from a regional policy', *Regional Science and Urban Economics* **41**(3), 253–265.
- Breidenbach, P., Mitze, T. and Schmidt, C. M. (2016), 'EU Structural Funds and Regional Income Convergence—A sobering experience'. CEPR Discussion Paper No. 11210.
- Cappelen, A., Castellacci, F., Fagerberg, J. and Verspagen, B. (2003), 'The impact of eu regional support on growth and convergence in the european union', *JCMS: Journal of Common Market Studies* **41**(4), 621–644.
- Commission, E. (2017), 'European structural and investment funds'. DG Regional Policy. [Accessed: 2017-05-03].  
**URL:** [http://ec.europa.eu/regional\\_policy/en/funding](http://ec.europa.eu/regional_policy/en/funding)
- Dall'Erba, S. and Le Gallo, J. (2008), 'Regional convergence and the impact of European structural funds over 1989–1999: A spatial econometric analysis', *Papers in Regional Science* **87**(2), 219–244.
- De Zwaan, M. and Merlevede, B. (2013), Regional policy and firm productivity. ETSG 2013 - No. 377.
- European Commission (2006), 'Commission Regulation (EC) No 1828/2006 of 8 December 2006, setting out rules for the implementation of Council Regulation (EC) No 1083/2006 laying down general provisions on the European Regional Development Fund, the European Social Fund and the Cohesion Fund and of Regulation (EC) No 1080/2006 of the European Parliament and of the Council on the European Regional Development Fund'. Official Journal of the European Union.
- European Commission (2011), 'Commission Regulation (EC) No 31/2011 of 17 January 2011, amending annexes to Regulation (EC) No 1059/2003 of the European Parliament and of the Council on the establishment of a common classification of territorial units for statistics (NUTS)'. Official Journal of the European Union.
- European Council (2006a), 'Council Decision of 6 October 2006 on Community strategic guidelines on cohesion (2006/702/EC)'.

- European Council (2006*b*), ‘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’.
- Eurostat, N. (2008), ‘Rev. 2–statistical classification of economic activities in the european community’, *Office for Official Publications of the European Communities, Luxembourg* .
- Feinerer, I., Hornik, K. and Meyer, D. (2008), ‘Text mining infrastructure in R’, *Journal of Statistical Software* **25**(1), 1–54.  
**URL:** <https://www.jstatsoft.org/index.php/jss/article/view/v025i05>
- Ferrara, A. R., McCann, P., Pellegrini, G., Stelder, D. and Terribile, F. (2016), ‘Assessing the impacts of Cohesion Policy on EU regions: A non-parametric analysis on interventions promoting research and innovation and transport accessibility’, *Papers in Regional Science* (2016).
- Hagen, T. and Mohl, P. (2009), Econometric evaluation of EU cohesion policy: A survey. ZEW Discussion Paper No. 09-052.
- Hartsenko, J. and Sauga, A. (2012), Does financial support from the EU structural funds has an impact on the firms’ performance: Evidence from Estonia, in ‘Proceedings of 30th International Conference Mathematical Methods in Economics’.
- Joulin, A., Grave, E., Bojanowski, P. and Mikolov, T. (2016), ‘Bag of tricks for efficient text classification’, *arXiv preprint arXiv:1607.01759* .
- Kalemli-Ozcan, S., Sorensen, B., Villegas-Sanchez, C., Volosovych, V. and Yesiltas, S. (2015), How to construct nationally representative firm level data from the ORBIS global database, Technical report, National Bureau of Economic Research.
- Kyriacou, A. P. and Roca-Sagalés, O. (2012), ‘The impact of eu structural funds on regional disparities within member states’, *Environment and Planning C: Government and Policy* **30**(2), 267–281.
- Maynou, L., Saez, M., Kyriacou, A. and Bacaria, J. (2016), ‘The impact of structural and cohesion funds on eurozone convergence, 1990–2010’, *Regional Studies* **50**(7), 1127–1139.
- Pellegrini, G., Terribile, F., Tarola, O., Muccigrosso, T. and Busillo, F. (2013), ‘Measuring the effects of European regional policy on economic growth: A regression discontinuity approach’, *Papers in Regional Science* **92**(1), 217–233.
- Percoco, M. (2013), ‘Strategies of regional development in European regions: Are they efficient?’, *Cambridge Journal of Regions, Economy and Society* pp. 303–318.
- R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.  
**URL:** <http://www.R-project.org>
- Rodríguez-Pose, A. (2013), ‘Do institutions matter for regional development?’, *Regional Studies* **47**(7), 1034–1047.
- Rodríguez-Pose, A. and Fratesi, U. (2004), ‘Between development and social policies: The impact of European structural funds in Objective 1 regions’, *Regional Studies* **38**(1), 97–113.
- Rodríguez-Pose, A. and Garcilazo, E. (2015), ‘Quality of government and the returns of investment: Examining the impact of cohesion expenditure in european regions’, *Regional Studies* **49**(8), 1274–1290.
- Stone, M. (1974), ‘Cross-validators choice and assessment of statistical predictions’, *Journal of the royal statistical society. Series B (Methodological)*, 111–147.

Ward Jr, J. H. (1963), ‘Hierarchical grouping to optimize an objective function’, *Journal of the American statistical association* **58**(301), 236–244.

## A Appendix

### A.1 List of operational programmes and coverage in the novel database

In the following table, we first report the fund co-financing the particular OP and the objective under which it operates as well as the geographical unit concerned (NUTS-2 or NUTS-1 or NUTS-0).<sup>30</sup> Second, Table 14 states the declaration date of each OP, i.e., the reporting date of the list of beneficiaries or, if not provided by the managing authority, the date when we downloaded the list. Finally, Table 14 indicates the degree of detail of reported project sums: From only committed co-financing values (C) to EU and national public co-funding and private ineligible expenditure data (C\_EU + C\_NAT + I), and whether we know the value actually paid-out.

**Table (14)** – List of operational programmes (OP)

OP Code	Fund	Obj.	NUTS	Committed	Paid-out	Declaration
BG 2007BG051PO001	ESF	1	ORBIS	C	no	11.06.2015
BG 2007BG051PO002	ESF	1	-	-	-	-
BG 2007BG161PO001	ERDF	1	-	-	-	-
BG 2007BG161PO002	ERDF	1	-	-	-	-
BG 2007BG161PO003	ERDF	1	-	-	-	-
BG 2007BG161PO004	ERDF & CF	1	-	-	-	-
BG 2007BG161PO005	ERDF & CF	1	0	C	yes	2.07.2013
BE 2007BE051PO001	ESF	1	ORBIS	C_EU + C_NAT + I	yes	10.2014
BE 2007BE052PO001	ESF	2	ORBIS	C_EU + C_NAT + I	yes	10.2014
BE 2007BE052PO002	ESF	2	ORBIS	C_EU + C_NAT + I	yes	10.2014
BE 2007BE052PO003	ESF	2	ORBIS	C_EU + C_NAT + I	yes	10.2014
BE 2007BE052PO004	ESF	2	ORBIS	C_EU + C_NAT + I	yes	10.2014
BE 2007BE052PO005	ESF	2	ORBIS	C_EU + C_NAT + I	yes	10.2014
BE 2007BE161PO001	ERDF	1	2	C_EU + C_NAT + I	yes	10.2014
BE 2007BE162PO001	ERDF	2	2	-	yes	10.2014
BE 2007BE162PO002	ERDF	2	ORBIS	C	yes	09.2014
BE 2007BE162PO003	ERDF	2	ORBIS	C_EU + C_NAT + I	yes	10.2014
CZ 2007CZ052PO001	ESF	2	2	C_EU	yes	04.09.2015
CZ 2007CZ05UPO001	ESF	1 & 2	0	C_EU	yes	04.09.2015
CZ 2007CZ05UPO002	ESF	1 & 2	0	C_EU	yes	04.09.2015
CZ 2007CZ161PO001	ERDF	1	2	C_EU	yes	04.09.2015
CZ 2007CZ161PO002	ERDF	1	2	C_EU	yes	04.09.2015
CZ 2007CZ161PO004	ERDF	1	0	C_EU	yes	04.09.2015
CZ 2007CZ161PO005	ERDF	1	2	C_EU	yes	04.09.2015
CZ 2007CZ161PO006	ERDF & CF	1	0	C_EU	yes	04.09.2015
CZ 2007CZ161PO007	ERDF & CF	1	0	C_EU	yes	04.09.2015
CZ 2007CZ161PO008	ERDF	1	2	C_EU	yes	04.09.2015
CZ 2007CZ161PO009	ERDF	1	2	C_EU	yes	04.09.2015
CZ 2007CZ161PO010	ERDF	1	2	C_EU	yes	04.09.2015
CZ 2007CZ161PO012	ERDF	1	0	C_EU	yes	04.09.2015
CZ 2007CZ161PO013	ERDF	1	2	C_EU	yes	04.09.2015
CZ 2007CZ162PO001	ERDF	2	2	C_EU	yes	04.09.2015
CZ 2007CZ16UPO001	ERDF	1 & 2	0	C_EU	yes	04.09.2015
CZ 2007CZ16UPO002	ERDF	1 & 2	0	C_EU	yes	04.09.2015
DK 2007DK052PO001	ESF	2	2	-	yes	08.09.2015
DK 2007DK162PO001	ERDF	2	2	-	yes	08.09.2015
DE 2007DE051PO001	ESF	1	1	C (or paid-out)	partly yes	31.12.2014
DE 2007DE051PO002	ESF	1	1	C (or paid-out)	partly yes	31.12.2013
DE 2007DE051PO003	ESF	1	1	C (or paid-out)	partly yes	31.03.2015

Continued on next page

<sup>30</sup>For the Netherlands, there are only national, i.e., no region-specific operational programmes and the official beneficiary lists do not contain information on their location. Partly, a NUTS-1 classification would be possible. However, we learn about the NUTS-2 region in which beneficiaries operate from ORBIS in two thirds of Dutch projects. To observations with an ORBIS match and a non-missing value for the postal code in ORBIS, we can assign a NUTS-3 region.

	OP Code	Fund	Obj.	NUTS	Committed	Paid-out	Declaration date
DE	2007DE051PO004	ESF	1	1	C (or paid-out)	partly yes	06.2015
DE	2007DE051PO005	ESF	1	1	C	yes	31.12.2014
DE	2007DE051PO006	ESF	1	1	C (or paid-out)	partly yes	31.12.2014
DE	2007DE052PO001	ESF	2	1	C (or paid-out)	partly yes	30.04.2014
DE	2007DE052PO002	ESF	2	1	C (or paid-out)	partly yes	08.12.2014
DE	2007DE052PO003	ESF	2	1	C	yes	31.12.2014
DE	2007DE052PO004	ESF	2	1	C (or paid-out)	partly yes	06.11.2013
DE	2007DE052PO005	ESF	2	1	C (or paid-out)	partly yes	31.12.2013
DE	2007DE052PO006	ESF	2	1	C	-	28.02.2014
DE	2007DE052PO007	ESF	2	1	C (or paid-out)	partly yes	31.03.2015
DE	2007DE052PO008	ESF	2	1	C	-	31.12.2014
DE	2007DE052PO009	ESF	2	1	C	partly yes	30.06.2014
DE	2007DE052PO010	ESF	2	1	C	-	05.10.2015
DE	2007DE052PO011	ESF	2	1	C (or paid-out)	partly yes	06.2015
DE	2007DE05UPO001	ESF	1 & 2	0	C (or paid-out)	yes	30.04.2015
DE	2007DE161PO001	ERDF	1	1	C (or paid-out)	partly yes	31.12.2014
DE	2007DE161PO002	ERDF	1	1	C	yes	31.12.2014
DE	2007DE161PO003	ERDF	1	1	C (or paid-out)	partly yes	31.12.2013
DE	2007DE161PO004	ERDF	1	1	C (or paid-out)	partly yes	06.2015
DE	2007DE161PO005	ERDF	1	0	C	-	31.12.2014
DE	2007DE161PO006	ERDF	1	1	C (or paid-out)	partly yes	31.03.2015
DE	2007DE161PO007	ERDF	1	1	C	yes	31.12.2014
DE	2007DE162PO001	ERDF	2	1	C (or paid-out)	partly yes	31.12.2014
DE	2007DE162PO002	ERDF	2	1	C	yes	31.12.2013
DE	2007DE162PO003	ERDF	2	1	C (or paid-out)	partly yes	31.12.2014
DE	2007DE162PO004	ERDF	2	1	C	yes	31.12.2014
DE	2007DE162PO005	ERDF	2	1	C (or paid-out)	partly yes	16.04.2015
DE	2007DE162PO006	ERDF	2	1	C (or paid-out)	partly yes	16.09.2015
DE	2007DE162PO007	ERDF	2	1	C	yes	31.12.2014
DE	2007DE162PO008	ERDF	2	1	C (or paid-out)	partly yes	31.12.2014
DE	2007DE162PO009	ERDF	2	1	C (or paid-out)	partly yes	31.12.2014
DE	2007DE162PO010	ERDF	2	1	C (or paid-out)	partly yes	31.03.2015
DE	2007DE162PO011	ERDF	2	1	C (or paid-out)	partly yes	01.07.2015
EE	2007EE051PO001	ESF	1	2	C	yes	31.08.2015
EE	2007EE161PO001	ERDF & CF	1	2	C	yes	31.08.2015
EE	2007EE161PO002	ERDF & CF	1	2	C	yes	31.08.2015
ES	2007ES051PO002	ESF	1	2	C	yes	20.05.2013
ES	2007ES051PO003	ESF	1	2	C	yes	20.05.2013
ES	2007ES051PO004	ESF	1	2	C	yes	20.05.2013
ES	2007ES051PO005	ESF	1	2	C	yes	25.07.2013
ES	2007ES051PO006	ESF	1	2	C	yes	20.05.2013
ES	2007ES051PO007	ESF	1	2	C	yes	20.05.2013
ES	2007ES051PO008	ESF	1	2	C	yes	20.05.2013
ES	2007ES051PO009	ESF	1	2	C	yes	20.05.2013
ES	2007ES052PO001	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO002	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO003	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO004	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO005	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO006	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO007	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO008	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO009	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO010	ESF	2	2	C	yes	20.05.2013
ES	2007ES052PO011	ESF	2	2	C	yes	20.05.2013
ES	2007ES05UPO001	ESF	1 & 2	0	C	yes	25.07.2013
ES	2007ES05UPO002	ESF	1 & 2	0	C	yes	20.05.2013
ES	2007ES05UPO003	ESF	1 & 2	0	C	yes	20.05.2013
ES	2007ES161PO001	ERDF	1	2	C	yes	13.10.2015
ES	2007ES161PO002	ERDF	1	2	C	yes	13.10.2015
ES	2007ES161PO003	ERDF	1	2	C	yes	13.10.2015
ES	2007ES161PO004	ERDF	1	2	C	yes	13.10.2015
ES	2007ES161PO005	ERDF	1	2	C	yes	13.10.2015
ES	2007ES161PO006	ERDF	1	2	C	yes	13.10.2015
ES	2007ES161PO007	ERDF	1	2	C	yes	13.10.2015
ES	2007ES161PO008	ERDF	1	2	C	yes	13.10.2015
ES	2007ES161PO009	ERDF & CF	1	0	C	yes	13.10.2015
ES	2007ES162PO001	ERDF	2	2	C	yes	01.02.2016
ES	2007ES162PO002	ERDF	2	2	C	yes	13.10.2015
ES	2007ES162PO003	ERDF	2	2	C	yes	13.10.2015
ES	2007ES162PO004	ERDF	2	2	C	yes	13.10.2015
ES	2007ES162PO005	ERDF	2	2	C	yes	13.10.2015
ES	2007ES162PO006	ERDF	2	2	C	yes	13.10.2015
ES	2007ES162PO007	ERDF	2	2	C	yes	13.10.2015
ES	2007ES162PO008	ERDF	2	2	C	yes	13.10.2015
ES	2007ES162PO009	ERDF	2	2	C	yes	13.10.2015
ES	2007ES162PO010	ERDF	2	2	C	yes	13.10.2015
ES	2007ES162PO011	ERDF	2	2	C	yes	13.10.2015

Continued on next page

	OP Code	Fund	Obj.	NUTS	Committed	Paid-out	Declaration date
ES	2007ES16UPO001	ERDF	2	0	C	yes	13.10.2015
ES	2007ES16UPO002	ERDF	1 & 2	0	C	yes	31.10.2015
ES	2007ES16UPO003	ERDF	1 & 2	0	C	yes	31.10.2015
FR	2007FR051PO001	ESF	1	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR051PO002	ESF	1	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR051PO003	ESF	1	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR051PO004	ESF	1	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR052PO001	ESF	2	2	C_EU + C_NAT	-	
FR	2007FR161PO001	ERDF	1	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR161PO002	ERDF	1	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR161PO003	ERDF	1	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR161PO004	ERDF	1	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO005	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO006	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO007	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO008	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO009	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO010	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO011	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO012	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO013	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO014	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO015	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO016	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO017	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO018	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO019	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO020	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO021	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
FR	2007FR162PO022	ERDF	2	2	C_EU + C_NAT	-	20.07.2016
HR	2007HR051PO001	ESF	1	ORBIS	C_EU + C_NAT	-	01.02.2015
HR	2007HR161PO001	CF	1	ORBIS	C_EU + C_NAT	-	01.07.2015
HR	2007HR161PO002	ERDF	1	ORBIS	C_EU + C_NAT	-	01.06.2015
HR	2007HR161PO003	ERDF	1	ORBIS	C_EU + C_NAT	-	10.06.2015
IE	2007IE052PO001	ESF	2	0	C or paid-out	-	19.09.2016
IE	2007IE162PO001	ERDF	2	2	C_EU + C_NAT	yes	10.2015
IE	2007IE162PO002	ERDF	2	2	C	yes (2)	09.2016
IT	2007IT051PO001	ESF	1	2	-	-	16.02.2015
IT	2007IT051PO002	ESF	1	2	C	yes	22.10.2015
IT	2007IT051PO003	ESF	1	2	C	yes	27.09.2012
IT	2007IT051PO004	ESF	1	2	C	yes	30.10.2014
	(ITC2 & ITF5)						
IT	2007IT051PO005	ESF	1	2	C	-	30.12.2013
IT	2007IT051PO006	ESF	1	2	C	yes	19.01.2015
IT	2007IT051PO007	ESF	1	-	-	-	-
IT	2007IT052PO001	ESF	2	2	C	yes	31.12.2014
IT	2007IT052PO002	ESF	2	2	C	-	31.12.2014
IT	2007IT052PO003	ESF	2	2	C	yes	31.12.2013
IT	2007IT052PO004	ESF	2	2	C	-	31.12.2014
IT	2007IT052PO005	ESF	2	2	C	yes	31.12.2014
IT	2007IT052PO006	ESF	2	2	C	-	31.09.2015
IT	2007IT052PO007	ESF	2	2	C	yes	31.12.2013
IT	2007IT052PO008	ESF	2	2	C	yes	31.12.2014
IT	2007IT052PO009	ESF	2	2	C	yes	19.05.2015
IT	2007IT052PO010	ESF	2	2	C	yes	31.12.2014
IT	2007IT052PO011	ESF	2	2	C	yes	05.10.2015
IT	2007IT052PO012	ESF	2	2	C	-	14.04.2015
IT	2007IT052PO013	ESF	2	2	C	-	16.01.2015
IT	2007IT052PO014	ESF	2	-	-	-	-
IT	2007IT052PO015	ESF	2	2	C_EU + C_NAT + I	yes	31.08.2015
IT	2007IT052PO016	ESF	2	2	C	yes	31.12.2014
IT	2007IT052PO017	ESF	2	2	C	yes	19.01.2015
IT	2007IT161PO001	ERDF	1	2	C	yes	31.10.2015
IT	2007IT161PO002	ERDF	1	2	C	yes	31.12.2014
IT	2007IT161PO003	ERDF	1	2	C_EU + C_NAT	yes	31.08.2013
IT	2007IT161PO004	ERDF	1	-	-	-	-
IT	2007IT161PO005	ERDF	1	2	C	-	30.12.2015
IT	2007IT161PO006	ERDF	1	-	-	-	-
IT	2007IT161PO007	ERDF	1	2	C	partly yes	14.10.2013
IT	2007IT161PO008	ERDF	1	2	C_EU + C_NAT	yes	12.01.2016
IT	2007IT161PO009	ERDF	1	2	C_EU + C_NAT	yes	11.11.2015
IT	2007IT161PO010	ERDF	1	2	C	yes	31.05.2015
IT	2007IT161PO011	ERDF	1	2	C	yes	28.02.2015
IT	2007IT161PO012	ERDF	1	2	C_EU + C_NAT	yes (2)	31.12.2014
IT	2007IT162PO001	ERDF	2	2	C	yes (2)	13.01.2016
IT	2007IT162PO002	ERDF	2	2	C	partly yes	31.10.2014
IT	2007IT162PO003	ERDF	2	2	C_EU + C_NA	yes (2)	30.09.2015

Continued on next page

	OP Code	Fund	Obj.	NUTS	Committed	Paid-out	Declaration date
IT	2007IT162PO004	ERDF	2	2	C	yes	22.06.2015
IT	2007IT162PO005	ERDF	2	2	C_EU + C_NAT	yes	31.08.2015
IT	2007IT162PO006	ERDF	2	2	C	yes	15.06.2015
IT	2007IT162PO007	ERDF	2	2	C_EU + C_NAT	yes (2)	25.03.2014
IT	2007IT162PO008	ERDF	2	2	C_EU + C_NAT	yes (2)	31.10.2013
IT	2007IT162PO009	ERDF	2	2	C	yes	31.09.2015
IT	2007IT162PO010	ERDF	2	2	C	yes	24.07.2015
IT	2007IT162PO011	ERDF	2	2	C	yes	15.01.2016
IT	2007IT162PO012	ERDF	2	2	C	yes	31.12.2014
IT	2007IT162PO013	ERDF	2	2	C	partly yes	23.10.2015
IT	2007IT162PO014	ERDF	2	2	C	-	30.06.2013
IT	2007IT162PO015	ERDF	2	2	C_EU + C_NAT + I	yes	30.06.2015
IT	2007IT162PO016	ERDF	2	2	C_EU + C_NAT	yes (2)	20.11.2015
LV	2007LV051PO001	ESF	1	2	C_EU + C_NAT + I	yes (2)	09.08.2015
LV	2007LV161PO001	ERDF	1	2	C_EU + C_NAT + I	yes (2)	09.08.2015
LV	2007LV161PO002	ERDF & CF	1	2	C_EU + C_NAT + I	yes (2)	09.08.2015
LT	2007LT051PO001	ESF	1	2	C_EU + C_NAT	yes (2)	10.09.2015
LT	2007LT051PO002	ESF	1	2	C_EU + C_NAT	yes (2)	10.09.2015
LT	2007LT161PO001	ERDF & CF	1	2	C_EU + C_NAT	yes (2)	10.09.2015
LT	2007LT161PO002	ERDF & CF	1	2	C_EU + C_NAT	yes (2)	10.09.2015
LU	2007LU052PO001	ESF	2	2	C_EU + C_NAT	-	12.2013
LU	2007LU162PO001	ERDF	2	2	C_EU + C_NAT	-	12.2014
MT	2007MT051PO001	ESF	1	2	C_EU + C_NAT	yes (2)	09.11.2015
MT	2007MT161PO001	ERDF & CF	1	2	C_EU + C_NAT	yes (2)	30.09.2015
NL	2007NL052PO001	ESF	2	0	C_EU + C_NAT + I	yes	01.08.2016
NL*	2007NL162PO001	ERDF	2	0	C_EU + C_NAT + I	yes	21.11.2016
NL	2007NL162PO002	ERDF	2	0	C_EU + C_NAT + I	yes	21.11.2016
NL	2007NL162PO003	ERDF	2	0	C_EU + C_NAT + I	yes	02.2015
NL	2007NL162PO004	ERDF	2	0	C_EU + C_NAT + I	yes	21.11.2016
AT	2007AT051PO001	ESF	1	2	C	yes	15.08.2015
AT	2007AT052PO001	ESF	2	0	-	yes	20.11.2015
AT	2007AT161PO001	ERDF	1	2	C (or paid-out)	partly yes	15.08.2015
AT	2007AT162PO001	ERDF	2	2	C	yes	13.01.2015
AT	2007AT162PO002	ERDF	2	2	C	yes	13.01.2015
AT	2007AT162PO003	ERDF	2	2	C	yes	13.01.2015
AT	2007AT162PO004	ERDF	2	2	C	yes	13.01.2015
AT	2007AT162PO005	ERDF	2	2	C	yes	13.01.2015
AT	2007AT162PO006	ERDF	2	2	C	yes	13.01.2015
AT	2007AT162PO007	ERDF	2	2	C	yes	13.01.2015
AT	2007AT162PO008	ERDF	2	2	C	yes	13.01.2015
PL	2007PL051PO001	ESF	1	2	C	yes (2)	30.06.2015
PL	2007PL051PO001	ESF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO001	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO002	ERDF & CF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO003	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO004	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO005	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO006	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO007	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO008	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO009	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO010	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO011	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO012	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO013	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO014	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO015	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO016	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO017	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO018	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO019	ERDF	1	2	C	yes (2)	30.06.2015
PL	2007PL161PO020	ERDF	1	2	C	yes (2)	30.06.2015
PT	2007PT051PO001	ESF	1	-	-	-	-
PT	2007PT052PO001	ESF	2	2	C_EU + C_NAT	-	02.11.2015
PT	2007PT05UPO001	ESF	1 & 2	0	C	-	28.01.2014
PT	2007PT05UPO001	ESF	1 & 2	0	C	-	28.01.2014
PT	2007PT05UPO002	ESF	1 & 2	0	1: C_EU + C_NAT, 2: C	-	12.2013
PT	2007PT161PO001	ERDF	1	0	C_EU + C_NAT	-	22.11.2015
PT	2007PT161PO002	ERDF	1	2	C_EU + C_NAT	-	31.07.2015
PT	2007PT161PO003	ERDF	1	2	C_EU + C_NAT + I	-	31.07.2015
PT	2007PT161PO004	ERDF	1	2	C_EU + C_NAT	-	30.06.2015
PT	2007PT161PO005	ERDF	1	2	C_EU + C_NAT	-	30.06.2015
PT	2007PT161PO006	ERDF	1	2	C_EU + C_NAT	-	31.07.2015
PT	2007PT162PO001	ERDF	2	2	C_EU + C_NAT	-	31.07.2015
PT	2007PT162PO002	ERDF	2	2	C_EU + C_NAT + I	-	31.07.2015
PT	2007PT16UPO001	ERDF & CF	1	0	C_EU + C_NAT	-	30.09.2015

Continued on next page



	OP Code	Fund	Obj.	NUTS	Committed	Paid-out	Declaration date
PT	2007PT16UPO002	ERDF	1 & 2	0	C_EU + C_NAT	-	30.11.2014
RO	2007RO051PO001	ESF	1	0	C_EU + C_NAT + I	-	31.07.2015
RO	2007RO051PO002	ESF	1	0	C_EU + C_NAT + I	-	31.07.2015
RO	2007RO161PO001	ERDF	1	0	C_EU + C_NAT + I	-	31.07.2015
RO	2007RO161PO002	ERDF	1	0	C_EU + C_NAT + I	-	31.07.2015
RO	2007RO161PO003	ERDF & CF	1	0	C_EU + C_NAT + I	-	31.07.2015
RO	2007RO161PO004	ERDF & CF	1	0	C_EU + C_NAT + I	-	31.07.2015
RO	2007RO161PO005	ERDF	1	0	C_EU + C_NAT + I	-	31.07.2015
SI	2007SI051PO001	ESF	1	2	C_EU + C_NAT	-	14.10.2015
SI	2007SI161PO001	ERDF	1	2	C_EU + C_NAT	-	14.10.2015
SI	2007SI161PO002	ERDF & CF	1	2	C_EU + C_NAT	-	14.10.2015
SK	2007SK05UPO001	ESF	1 & 2	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK05UPO002	ESF	1 & 2	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK161PO001	ERDF	1	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK161PO002	ERDF & CF	1	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK161PO003	ERDF	1	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK161PO004	ERDF & CF	1	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK161PO005	ERDF	1	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK161PO006	ERDF	1	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK161PO007	ERDF	1	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK162PO001	ERDF	2	2	C_EU + C_NAT + I	yes (2)	30.06.2015
SK	2007SK16UPO001	ERDF	1 & 2	2	C_EU + C_NAT + I	yes (2)	30.06.2015
FI	2007FI052PO001	ESF	2	2	C_EU + C_NAT	yes (2)	12.10.2015
FI	2007FI052PO002	ESF	2	2	C_EU + C_NAT	yes (2)	12.10.2015
FI	2007FI162PO001	ERDF	2	2	C_EU + C_NAT	yes (2)	12.10.2015
FI	2007FI162PO002	ERDF	2	2	C_EU + C_NAT	yes (2)	12.10.2015
FI	2007FI162PO003	ERDF	2	2	C_EU + C_NAT	yes (2)	12.10.2015
FI	2007FI162PO004	ERDF	2	2	C_EU + C_NAT	yes (2)	12.10.2015
FI	2007FI162PO005	ERDF	2	-	-	-	-
SE	2007SE052PO001	ESF	2	2	C or paid-out	-	18.11.2015
SE**	2007SE162PO001	ERDF	2	0	C or paid-out	-	06.11.2015
SE	2007SE162PO002	ERDF	2	0	C or paid-out	-	06.11.2015
SE	2007SE162PO003	ERDF	2	0	C or paid-out	-	06.11.2015
SE	2007SE162PO004	ERDF	2	0	C or paid-out	-	06.11.2015
SE	2007SE162PO005	ERDF	2	0	C or paid-out	-	06.11.2015
SE	2007SE162PO006	ERDF	2	0	C or paid-out	-	06.11.2015
SE	2007SE162PO007	ERDF	2	0	C or paid-out	-	06.11.2015
SE	2007SE162PO008	ERDF	2	0	C or paid-out	-	06.11.2015
UK	2007UK051PO001	ESF	1	1	C_EU + C_NAT	-	30.04.2014
UK	2007UK051PO002	ESF	1	1	C_EU + C_NAT	-	25.08.2016
UK	2007UK052PO001	ESF	2	1	C_EU + C_NAT	-	25.08.2016
UK	2007UK052PO002	ESF	2	1	C_EU + C_NAT	-	30.04.2014
UK	2007UK052PO003	ESF	2	1	C_EU + C_NAT + I	yes	23.11.2016
UK	2007UK05UPO001	ESF	1 & 2	-	-	-	-
UK	2007UK161PO001	ERDF	1	1	C_EU + C_NAT	-	30.04.2014
UK	2007UK161PO002	ERDF	1	1	C_EU + C_NAT	-	25.08.2016
UK	2007UK161PO003	ERDF	1	1	C_EU + C_NAT	-	31.01.2015
UK	2007UK162PO001	ERDF	2	1	C_EU + C_NAT	-	30.04.2014
UK	2007UK162PO002	ERDF	2	1	C	-	01.05.2014
UK	2007UK162PO003	ERDF	2	1	C_EU + C_NAT + I	partly yes (3)	08.10.2015
UK	2007UK162PO004	ERDF	2	1	C_EU + C_NAT + I	yes (2)	01.01.2014
UK	2007UK162PO005	ERDF	2	1	C_EU + C_NAT	-	31.12.2013
UK	2007UK162PO006	ERDF	2	1	C_EU + C_NAT + I	yes	01.10.2015
UK	2007UK162PO007	ERDF	2	1	C_EU + C_NAT	-	29.06.2016
UK	2007UK162PO008	ERDF	2	1	C	-	29.06.2016
UK	2007UK162PO009	ERDF	2	1	C_EU + C_NAT	-	29.06.2016
UK	2007UK162PO010	ERDF	2	1	C_EU + C_NAT	-	29.06.2016
UK	2007UK162PO011	ERDF	2	1	C_EU + C_NAT	-	31.01.2015
UK	2007UK162PO012	ERDF	2	1	C_EU + C_NAT	-	25.08.2016
UK	2007UK162PO012	ERDF	2	-	-	-	-

Notes: This table lists the operational programmes and information found in the official lists of beneficiaries provided by regional managing authorities or other regional or national authorities. The underlying list of operational programmes including their names can be downloaded at [http://ec.europa.eu/regional\\_policy/en/policy/evaluations/data-for-research/](http://ec.europa.eu/regional_policy/en/policy/evaluations/data-for-research/) in the section “EU Budget commitments by fund by year and by programme” when selection the programming period 2007-2013. \* Projects of 2007NL162PO002 and 2007NL162PO004 are part of the list of beneficiary for OP 2007NL162PO001. \*\* Unequivocal assignment of ERDF projects to one of ERDF OPs is not possible. *C\_EU* stands for committed EU co-financing, *C\_NAT* for committed national co-financing, and *I* signifies ineligible cost, i.e., the cost carried by the beneficiary. Hence, *C\_EU + C\_NAT + I* means that the structure of the total project value is known. *C* means the a committed value, without declared partition across national and EU budget, is reported. A “yes” in column *Paid-out* means that we have information on paid-out values. “yes (2)” denotes that we know the partition of the paid-out amount into EU and national public co-financing. If the fifth and consequent columns do not contain any information, the dataset does not cover them as we have not found beneficiaries lists provided by the respective authorities. When the NUTS dimension given is “ORBIS”, it means that we have NUTS-2 information for those beneficiaries of the respective OP we could find and match with the ORBIS database. Objective 1 refers to the *Convergence* objective, Objective 2 to *Regional competitiveness and employment*.