ISIGrowth

INNOVATION-FUELLED, SUSTAINABLE, INCLUSIVE GROWTH

Working Paper

Varieties of European National Innovation Systems

Ariel L. Wirkierman
Science Policy Research Unit (SPRU), University of Sussex, UK

Tommaso Ciarli
Science Policy Research Unit (SPRU), University of Sussex, UK

Maria Savona
Science Policy Research Unit (SPRU), University of Sussex, UK

13/2018 May

This project has received funding from the European Union Horizon 2020 Research and Innovation action under grant agreement No 649186

Varieties of European National Innovation Systems

Ariel L. Wirkierman[*] Tommaso Ciarli[†] Maria Savona[‡]

Abstract This paper provides a novel, empirically grounded map of National Innovation Systems (NIS) in Europe, based on a unique micro level analysis across several EU countries. By focusing on the Eurostat Community Innovation Survey 2014 (CIS2014) micro-aggregated data, we perform an exploratory factor analysis to provide a micro-level grounding to the multi-faceted components of NIS. We relate the structure, innovation strategies and performance of the firm to relevant institutional characteristics of the NIS in which it is embedded, including the nature of public sector support (e.g. cooperation and procurement) and the characteristics of the public-private links (e.g. with universities, foreign institutions and/or other firms), amongst others. We then redesign the map of the European technology 'clubs' by means of a cluster analysis based on our factors/NIS dimensions. Our findings ground the diagnostics of the European NIS, add to the most recent literature on NIS by taking into account the micro–level sources of the European NIS 'clubs', and complement the historical picture provided by Cirillo et al. (2016a).

Note: This paper represents Deliverable DA 1.2 and contributes to ISI-Growth Work Package 1, Task 1.2.

[*]Science Policy Research Unit (SPRU), University of Sussex, UK. Email: A.L.Wirkierman@sussex.ac.uk
[†]Science Policy Research Unit (SPRU), University of Sussex, UK. Email: T.Ciarli@sussex.ac.uk
[‡]Science Policy Research Unit (SPRU), University of Sussex, UK. Email: M.Savona@sussex.ac.uk

1 Introduction and Background

At times of global macro-economic recession and national attempts of secessions, there is a revamped need to understand where Europe as a Science, Technology and Innovation System currently stands against competitors such as the US and Japan, and increasingly China, and whether a 'European Paradox' has ever and is still characterising it (Dosi et al., 2006; Hammadou et al., 2014).

The European Paradox has been described as a relative excellent performance in basic research and scientific infrastructures co-existing with a weaker performance in terms of industrial applications and innovation outputs in many European countries, compared with the US. Some scholars have instead argued that the European Science, Technology and Innovation System lags behind *also* in terms of basic science performance, and that the innovation output returns of R&D investments and public science could be much higher (Dosi et al., 2006; see also Cirillo et al., 2016a for a recent review). However, the evidence is not conclusive about the presence — and importance — of a European Paradox. This depends crucially on the theoretical background explaining its causes as well as the measurement and empirical issues related to the choice of relevant variables.

The innovation and technology gap literatures have attempted to tackle both these aspects over the past decades. Both branches of literature have conspicuously borrowed from the National Innovation System (NIS, hereinafter) framework, which has sparked following seminal contributions that have provided rich historical accounts of emblematic cases of NIS, such as Japan, the (then) Soviet Union and the US (Freeman, 1987; Lundvall, 1992; Nelson, 1993).

The NIS approach has then long informed research and policy makers on the sources and nature of countries' differences in science and innovation performances, public support to science and economic outcomes. It has proven useful to 'appreciatively' complement the growth literature on technology clubs and countries' divergences due to catching-up processes in science and technology performance (see Castellacci, 2008; Castellacci and Archibugi, 2008; Fagerberg and Srholec, 2008, among others).

Over time, the innovation scholarship on NIS has flourished, and has been the object of eminent reflections, even recently (Nelson, 2006; Lundvall, 2007), within the debate on the European Paradox. A common denominator of the original presentation and developments of the NIS approach is the implicit rejection of a 'one size fits all' normative implication of the linear R&D model (Soete et al., 2010; Cirillo et al., 2016b). There is not a single recipe that countries can follow to ensure technological upgrading and catch-up. The very notion of a successful pathway of catching up is nonsensical, as countries' (and firms') idiosyncrasies and heterogeneities are such that it is the complex combination of initial conditions and the evolution of several of them, rather than their arithmetic sum, that lies behind any success story. But what are these factors?

In a nutshell, the NIS approach posits that a wide set of national characteristics, beyond the obvious size, population and GDP per capita, are relevant to explain national differences in Science, Technology and Innovation Systems and ultimately their economic performance. More specifically, the core components of a NIS are:

1. the private organizations responsible for the applications of basic science and creation of knowledge and at firm and sectoral levels;

2. the scientific and technological public infrastructures, such as research centres, universities and higher education institutions;

3. the battery of instruments used by the government to fund and support both of the above, such as public procurement, grants, subsidies to firms and R&D tax credits;

4. the nature and intensity of links between private and public actors aimed at increasing scientific and technological capabilities.

Indeed, the all-encompassing nature of the NIS approach makes it useful, yet quite difficult to capture empirically, in the absence of a rigorous theoretical grounding (Castellacci and Natera, 2013). One of the most comprehensive attempts to map the global variety of NIS is offered by Fagerberg and Srholec (2008), who look at a large variety of variables to empirically ground a typology of technological clubs. They distinguish variables pertaining to the "innovation system" (albeit oddly a component is named as the whole), "governance", "openness" and "political system", all contributing to a multi-dimensional description of national innovation systems. Interestingly for our purposes, in a related contribution one the authors argues that "the most relevant contextual factors cut across the established boundaries between sectors and countries" (Shrolec and Verspagen, 2008, p. 3) and are attributable to firm heterogeneity.

The present paper builds upon, and complements the exercises in Shrolec and Verspagen (2008) and Cirillo et al. (2016b), by providing a novel, micro-level grounded mapping of European National Innovation Systems (NIS).

The aim is to empirically derive the composite dimensions of NIS by reprising the emphasis that the NIS approach has traditionally put on firms' behaviour and performance, albeit embedded in the complex network of actors, which firms interact with and respond to.

We relate the structure, innovation strategy and performance of firms to several institutional characteristics of the NIS, such as the nature of public sector support (e.g. cooperation and procurement) and the characteristics of the public-private links (e.g. with universities, foreign institutions and/or other firms), amongst others. Firms choose to invest, cooperate and benefit from various forms of public support and interactions as a result of their specific characteristics and perception of the local, national and international context they operate in. Arguably, the best way to capture the macro-economic dimension of NIS is to resort to its micro-economic foundation.

To this purpose, we use the Community Innovation Survey (CIS) available in a comparable and 'pseudo-micro-founded' format from Eurostat, in line with some recent works (Frenz and Lambert, 2012; Frenz and Prevezer, 2012), which put forward that the CIS captures *structural* (rather than transient) features of innovative activity.[1]

We reproduce the NIS dimensions along four firm-centered activities: (i) innovation inputs and demand sources; (ii) geography and type of cooperation links; (iii) public sector policies in the form of public procurement and indirect support to firms; and (iv) innovation outputs.[2] In the best NIS tradition, the dimensions above allow us to look at how three categories of *subjects/actors* (private sector, government and public institutions) 'score' in terms of *objects/activities* (inputs, outputs and cooperation links) and how each of these categories of actors affects each others' performance in terms of inputs, outputs and cooperation. For instance, the government might intervene with public procurement, which will affect the amount and direction of innovation investments carried out by firms and possibly the intensity of cooperation with

[1] As it is well-known, and described at length in the next section, the CIS includes: innovation outputs, a range of innovation inputs in addition to R&D, as well as data on sources of information for innovation, cooperation partners and intellectual property rights protection and aligns to international standards in terms of questionnaire and data collection (OECD, 2002, 2005, 2009).

[2] As detailed in our empirical strategy section, it is worth noting that the specificity of our chosen methodological tool — exploratory factor analysis — allows us, within each dimension considered, to avoid grouping variables and assigning weighting schemes in an ad-hoc manner.

public (local, national or international) institutions. Our focus is on a broad set of European countries, and how they rank across the composite dimensions that the data allow us to derive.

With no pretension to establish causal links, which is not the objective of the analysis proposed here, we therefore ask:

1. What are the most relevant *latent* dimensions that characterise the NIS and that emerge from the (observed) micro-level (firm) dimensions (i) to (iv) above?

2. How do countries rank along these latent dimensions? Do comparisons in such ranking allow to unravel characteristics of the NIS that articulate 'club' positions in Europe?[3]

3. Does this newly derived map of European NIS allow to say something on the directions that innovation and industrial policy should take to ensure catching up of peripheral macro-regions in Europe and overcome the 'European Paradox'?

We find a very articulated map of European NIS, that confirms the original spirit of the NIS literature: there is no unique recipe for countries to follow. European countries cluster along several dimensions, some are in line with the extant empirical literature, some others emerge as exceptions to the NIS common sense. Yet, within this variety, we find a common denominator of core dimensions, namely the role of government, public support to innovation, and the cooperation between public and private actors, that suggests a way to build up on countries' idiosyncratic strengths to achieve their own pathway of growth.

The rest of the paper is structured as follows: the next section describes the micro-aggregated database employed and the data reduction technique adopted to obtain the factors within NIS dimensions; Section 3 lays down our empirical strategy to answer the questions posed above. Section 4 empirically illustrates the factors bearing relevant NIS dimensions; Section 5 reports on the comparison of country rankings along these dimensions in detail; Section 6 collages the different pieces of evidence to depict a map of Eropean NIS, while section 7 summarises and concludes.

[3] We also compare our findings with the largely used country ranking obtained from the European Innovation Scoreboard, both as a way of robustness check of our empirical strategy and to unravel potential discrepancies that might emerge (see Appendix B for details).

2 Dataset and Methodology

2.1 Dataset: Community Innovation Survey 2014 (CIS2014)

We use the publicly available micro-aggregated version of Eurostat Community Innovation Survey, 2014 edition (CIS, hereinafter).[4]

The CIS is a firm-level survey executed at a national scale, which collects data on several dimensions of innovative activity and outcomes. The unit of analysis considered is the enterprise with 10 or more employees enrolled (in most cases) in the official statistical business register of each country. To ensure cross-country comparability, the survey is carried out by means of a standard questionnaire, based on the definitions and underlying methodology included in the well-known Oslo manual for collecting and interpreting innovation data (OECD and EUROSTAT, 2005).

The survey is performed every two years, covering the 28 EU member states and some additional countries.[5] Most statistics refer to the 3-year reference period 2012-2014, even though some indicators specifically correspond to 2012 and/or 2014.

Rather than using a firm-level dataset, we use micro-aggregated CIS results (i.e. data that have been aggregated across firms within each country, innovation type, economic activity and size class combination). This choice is dictated by a number of reasons.

First, European innovation statistics generally use aggregated national data.[6] By using micro-aggregated data we provide a novel, and more fine-grained picture than the use of traditional country-level indicators would allow.

Second, in the process of consolidating firm-level observations, national statistical institutes extrapolate collected data, by means of appropriate weighting schemes, in order to get population totals. As a consequence, official micro-aggregated data deal with the issue of sample size heterogeneity across countries.

Third, it should be borne in mind that individual firms cannot be followed from one CIS wave to another, which implies that such data cannot be treated as a panel across sequential CIS editions.

[4] A detailed meta-data description can be found in: `http://ec.europa.eu/eurostat/cache/metadata/en/inn_cis9_esms.htm`

[5] The CIS 2014 has been conducted in the following additional countries: Norway, Iceland, Switzerland, Serbia, Macedonia and Turkey.

[6] See section '3.1. Data description' in Eurostat CIS 2014 meta-data documentation: `http://ec.europa.eu/eurostat/cache/metadata/en/inn_cis9_esms.htm`

Fourth, focusing on *micro-aggregated* results allows us to obtain variables measuring both the *proportion* of firms that engage in innovation activity, co-operation, receive public funding or achieve a certain outcome,[7] as well as the *intensity* with which firms perform those tasks (e.g. the amount of R&D expenditure classified by type). This is crucial as CIS *firm-level* studies mostly rely on binary or Likert-scale variables, as innovative expenditure data by type is aggregated (due to confidentiality issues), preventing its use in empirical studies (Shrolec and Verspagen, 2008).

Eurostat performs no imputation for missing firm-level data. In general, this implies a trade-off between country availability and the breadth of variables considered in empirical analyses (see, for example, the discussion in Shrolec and Verspagen, 2008, p. 12). Given that our aim is to have the widest possible country coverage, we have estimated missing values at the micro-aggregated level using regression techniques.[8]

We considered 26 European countries for which data gaps made the missing-data imputation process parsimonious.[9] As a result, we obtained a working dataset consisting of 33 variables across 26 countries.

The 33 variables considered provide information on the expenditures, ownership structure, knowledge acquisition, sources of cooperation links, public funding/procurement, protection mechanisms (patents/trademarks), persistence and productivity in relation to innovation activities and outcomes.

The CIS covers both inputs/strategies (e.g. implementation, adoption) and outputs/effects (e.g. successful, ongoing or abandoned) of innovative activities. Moreover, the CIS organises data collection according to the type of innovation activity that firms declare to be engaged in (product, process, organisational and marketing innovation). The observed working variables that feed our data reduction procedures are (almost exclusively) limited to product and process innovation (i.e. technological innovation),[10] even though we consider some variables that correspond to the subset of innovative firms,[11] as well as some refer-

[7]Variables of this sort are a "ratio between the selected combination of indicator, type of innovators and — in most cases — the total category of the selected type of innovators", as reported in: `http://ec.europa.eu/eurostat/cache/metadata/en/inn_cis9_esms.htm`

[8]Please see Appendix A for details.

[9]The countries considered (with the corresponding ISO2 code) are: Austria (AT), Belgium (BE), Bulgaria (BG), Cyprus (CY), Czech Republic (CZ), Germany (DE), Denmark (DK), Estonia (EE), Greece (EL), Spain (ES), Finland (FI), France (FR), Croatia (HR), Hungary (HU), Ireland (IE), Italy (IT), Lithuania (LT), Latvia (LV), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Sweden (SE), Slovenia (SI) and Slovakia (SK).

[10]In the CIS these firms are labelled 'INNOACT': product and process innovative enterprises regardless of organisational and marketing innovation.

[11]In the CIS these firms are labelled 'INNO': innovative enterprises.

ring to the total universe of firms.[12] Note that we have chosen the indicators per types of firms that maximises the number of observations across countries, conditioned therefore to data availability.

2.2 Exploratory Factor Analysis

This section contains a brief description of the method, emphasising aspects of relevance for our empirical strategy in Section 3.[13]

Our starting point is a multivariate sample of observations for 33 variables across 26 countries covering a variety of aspects of the innovation process, as captured by the CIS. As mentioned in the previous section, we aim to articulate the four dimensions that characterise a NIS: (i) innovation inputs and demand sources, (ii) the geography and type of cooperation links, (iii) government role and public sector policies, and (iv) innovation outputs.

We use exploratory factor analysis (EFA, hereinafter) to reduce the set of 33 variables to underlying concepts (called factors). EFA is a statistical data reduction technique which allows us to combine and summarise groups of observed variables according to their covariances. Essentially, it uncovers the way in which these variables form coherent subsets. For each of the four dimensions (i)-(iv) we apply EFA in order to identify (latent) common factors that best describe the differences across countries.

Table 1 reports a dictionary of the 33 variables (derived) from the CIS-2014 that have been used in our empirical analysis for each dimension (i)-(iv). Each row corresponds to a variable and includes a code label used throughout the paper, the firm type which it refers to, a short description and its unit of measurement.

[12]In the CIS the label used is 'TOTAL': total enterprises.

[13]For a comprehensive treatment see Timm (2002); Raykov and Marcoulides (2008); Everitt and Hothorn (2011); Rencher and Christensen (2012).

Table 1: CIS-2014 variables considered for the empirical analysis

Selected variables of/derived from the Community Innovation Survey 2014 Ed. (CIS-2014) used in the paper:

# Variable	Factor Analysis	Variable Label	Firm Type	Indicator Short Description	Unit	Indicator full description (derived from EUROSTAT)
1	Firm Innovation Inputs and Demand Sources	INNO_PPANPP_LARMAR_EU	Innovative firms all dimensions	Largest market: EU	%	Enterprises for which the largest market in terms of turnover is: EU/EFTA/EU-candidates
2		INNO_PPANPP_LARMAR_LREG	Innovative firms all dimensions	Largest market: Local/Regional	%	Enterprises for which the largest market in terms of turnover is the local/regional market
3		INNO_PPANPP_LARMAR_NAT	Innovative firms all dimensions	Largest market: National	%	Enterprises for which the largest market in terms of turnover is the national market
4		INNO_PPANPP_ENMRG_YES	Innovative firms all dimensions	Firm merged/took over	%	Enterprises that have merged with/take over another enterprise
5		INNO_PPANPP_GP_YES	Product/Process innovative firms	Firm part of enterprise group	%	Enterprises that are part of an enterprise group
6		INNOACT_EXPTOT14_ENT_POPU14	Product/Process innovative firms	R&D expenditure per firm	EUR/NR	Average total innovation expenditures in 2014 per firm
7		INNOACT_RRDEX14_PC	Product/Process innovative firms	Share of external R&D	% of TIE	Share of expenditures in external R&D in 2014 over total innovation expenditures
8		INNOACT_RRDIN14_PC	Product/Process innovative firms	Share of in-house R&D	% of TIE	Share of expenditures in in-house R&D in 2014 over total innovation expenditures
9		INNOACT_EXPTOT14_C	Product/Process innovative firms	Manufacturing/Aggregate R&D	% of TIE	Share of total innovation expenditures in 2014 in Manufacturing
10		INNOACT_ROEK14_PC	Product/Process innovative firms	Acquisition of external knowledge	% of TIE	Share of expenditures in acquisition of external knowledge in 2014
11	Cooperation Links	INNOACT_C01	Product/Process innovative firms	Cooperation within the enterprise group	%	Enterprises co-operating with other enterprises within the enterprise group
12		INNOACT_COEUR_YES	Product/Process innovative firms	Cooperation with EU partners	%	Enterprises engaged in innovation co-operation with a partner in EU/EFTA/EU-candidates
13		INNOACT_CONAT_YES	Product/Process innovative firms	Cooperation with National partners	%	Enterprises engaged in any type of innovation co-operation with a national partner
14		INNOACT_COCNIN_YES	Product/Process innovative firms	Cooperation with China/India	%	Enterprises engaged in any type of innovation co-operation with a partner in China or India
15		INNOACT_COUS_YES	Product/Process innovative firms	Cooperation with the US	%	Enterprises engaged in any type of innovation co-operation with a partner in United States
16		INNOACT_C02	Product/Process innovative firms	Cooperation with competitors, same sector	%	Enterprises co-operating with competitors or other enterprises of the same sector
17		INNOACT_C031	Product/Process innovative firms	Cooperation with private clients/customers	%	Enterprises co-operating with clients or customers from the private sector
18	Government Role and Public Sector Policies	INNOACT_FUNGMT	Product/Process innovative firms	Funding from Central Government	%	Enterprises that received funding from central government
19		INNOACT_C032	Product/Process innovative firms	Coop. with public sector clients/customers	%	Enterprises co-operating with clients or customers from the public sector
20		INNOACT_C06	Product/Process innovative firms	Cooperation with universities/HEI	%	Enterprises co-operating with universities or other higher education institutions
21		INNOACT_C09	Product/Process innovative firms	Cooperation with Gvt/Research Inst.	%	Enterprises co-operating with Government, public or private research institutes
22		INNOACT_FUNLOC	Product/Process innovative firms	Funding from Local/Regional Auth.	%	Enterprises that received funding from local or regional authorities
23		TOTAL_PUBDOM	Total firms	Domestic Procurement	%	Enterprises with procurement contract for domestic public sector
24		TOTAL_PUBFINRQ	Total firms	Foreign proc. req. innovation activities	%	Enterprises with procurement contract for foreign public sector/innovation activities required
25		TOTAL_PUBFOR	Total firms	Foreign Procurement	%	Enterprises with procurement contract for foreign public sector
26	Firm Innovation Outputs	INNO_PROPAT	Innovative firms	Application for a patent	%	Enterprises that applied for a patent
27		INNO_PROTM	Innovative firms	Registration of a trademark	%	Enterprises that registered a trademark
28		INPDT_NEWFRM_YES	Product innovative firms	Turnover from porducts new to firm	%	Enterprises introduced new or significantly improved products that were only new to the firm
29		INPDT_NEWMAR_YES	Product innovative firms	Turnover from porducts new to market	%	Enterprises introduced new or significantly improved products that were new to the market
30		INPCS_INPSNM0	Process innovative firms	Process innovation new to firm	%	Enterprises that have introduced process innovation not new to the market
31		INPCS_INPSNM1	Process innovative firms	Process innovation new to market	%	Enterprises that have introduced process innovation new to the market
32		INONG_ENT_POPU14	Firms with ongoing innovation	Ongoing innovation activities	%	Enterprises with on-going innovation activities only
33		INNO_TURN_EMP	Innovative firms	Turnover per employee	EUR/EMP	Total turnover in 2014 per employee

References:
% (percentages) are expressed in relation to the total of firms of the corresponding firm type
NR: number; EUR: euros at current prices; EMP: employees; % of TIE: percentage of Total Innovation Expenditure; % of Turnover: percentage of total firm turnover
Innovative firms all dimensions corresponds to firm type INNO_PPANPP: Product and/or process innative enterprises and organisation and/or marketing innovative enterprises
Source: Own elaboration based on EUROSTAT CIS 2014 Database

The underlying rationale behind EFA is to formulate a linear probability model with specific moment constraints such that the observed covariances *between* the observed variables can be explained by the relationship of these variables with the (common) latent factors. Essentially, the k-factor model for q observed variables and k latent factors can be formulated as:

$$x_i = c_i + u_i, \qquad \forall i = 1, \ldots, q \tag{1}$$

$$c_i = \lambda_{i1} f_1 + \cdots + \lambda_{ik} f_k, \qquad \forall i = 1, \ldots, q \tag{2}$$

where, in our context, the variable x_i, which measures an observable characteristic of innovative activity (e.g. share of in-house R&D expenditure), is linked to a linear combination of (unobserved) latent factors c_i and randomly disturbed by the term u_i.

By assuming that:

1. Random disturbances u_i are uncorrelated with each other: $\mathrm{Cov}(u_i, u_s) = 0, \quad \forall i, s = 1, \ldots, q$;
2. Random disturbances u_i are uncorrelated with latent factors f_j: $\mathrm{Cov}(u_i, f_j) = 0, \quad \forall i = 1, \ldots, q$ and $\forall j = 1, \ldots, k$;
3. Factors f_j are uncorrelated with each other:[14] $\mathrm{Cov}(f_j, f_r) = 0, \quad \forall j, r = 1, \ldots, k$;
4. Factors are standardised:[15] $\mathrm{E}(f_j) = 0, \mathrm{V}(f_j) = 1, \quad \forall j = 1, \ldots, k$.

we obtain the essential result that:

$$\mathrm{Cov}(x_i, x_s) = \mathrm{E}(x_i x_s) = \lambda_{i1}\lambda_{s1} + \cdots + \lambda_{ik}\lambda_{sk}, \quad \forall i, s = 1, \ldots, q, \quad i \neq s$$

i.e. the covariance amongst observed variables x_i and x_s depends exclusively on the connection between the variables and the k common factors (coefficients $\lambda_{i1}, \ldots, \lambda_{ik}$ for x_i and $\lambda_{s1}, \ldots, \lambda_{sk}$ for x_s).

The formulation of the problem (1)-(2) under assumptions 1-4 implies that coefficients $\lambda_{i1}, \ldots, \lambda_{ik}$ are regression coefficients of x_i on the factors $f_1, \ldots, f_k$.

[14] This latter constraint on the cross-moments between factors will be relaxed in our implementation of the setting.

[15] Due to their being unobserved, the scales and locations of factors can be fixed arbitrarily (Everitt and Hothorn, 2011, p. 137).

Such coefficients are labelled *factor loadings* and quantify the correlations between the observed variables and the factors, i.e. coefficient λ_{ij} quantifies the correlation between variable x_i and factor f_j. When jointly considered, the k-factor model may be compactly expressed as:

$$\boldsymbol{x} = \boldsymbol{\Lambda} \boldsymbol{f} + \boldsymbol{u} \tag{3}$$

where:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_q \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11} & \dots & \lambda_{1k} \\ \vdots & \ddots & \vdots \\ \lambda_{q1} & \dots & \lambda_{qk} \end{bmatrix}, \quad \boldsymbol{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_k \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_q \end{bmatrix}$$

Crucially, the assumptions above imply that the population covariance matrix of the original variables is given by:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda} \boldsymbol{\Lambda}^{T} + \boldsymbol{\sigma}_u \tag{4}$$

where $\boldsymbol{\sigma}_u = \text{diag}[\text{V}(u_i)]$ is a diagonal matrix with the variances of the variable-specific random disturbances u_i.

Thus, the estimation problem of interest is to find point estimates $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\boldsymbol{\sigma}}_u$ such that the sample covariance matrix $\boldsymbol{S}$ of the (manifest) variables can be approximately written as:

$$\boldsymbol{S} \approx \widehat{\boldsymbol{\Lambda}} \widehat{\boldsymbol{\Lambda}}^{T} + \widehat{\boldsymbol{\sigma}}_u$$

i.e. to obtain a predicted covariance matrix that resembles the sample covariance matrix of the manifest variables.[16]

Two estimation methods are normally used: an eigenproblem technique known as principal factor analysis and a maximum likelihood approach (Everitt and Hothorn, 2011, pp.141-2). We adopt the latter, as it has an associated inferential procedure to test the null hypothesis that k common factors are sufficient to describe the data against the alternative that the population covariance matrix has no constraints (as imposed by the k-factor model). The test statistic is distributed χ^2 under the null hypothesis. Thus, we may start with $k = 1$ and successively increase k until we do not reject the null hypothesis.[17]

Note, however, that solutions with k and $k+1$ factors will produce a different

[16] Note that "factor analysis is essentially unaffected by the rescaling of the variables" (Everitt and Hothorn, 2011, p. 139), so it is essentially equivalent to work with the covariance or correlation matrix.

[17] In practice, we may consider the value of k for which we do not reject the null hypothesis as an upper bound on the number of factors relevant in practice (Everitt and Hothorn, 2011, p. 155).

set of factor loadings *altogether*. In fact, a solution with not enough factors will have too many high factor loadings associated to each of them, whereas a solution with an excess of factors may render difficult the conceptual interpretation (i.e. finding a meaning through combining subsets of the original variables).

A further element to be considered is that factor analysis accounts only for the variation in the observed variables *shared through* the common factors. The focus is on the estimates $\hat{\lambda}_{ij}$ of regression coefficients λ_{ij}.[18] We are *not* accounting for the entire variance of the observed variables.[19]

EFA has been criticised for suffering from two non-uniqueness problems. First, the non-uniqueness of the factor loadings matrix: alternative factor *rotations* alter the *description* of the solution obtained (though not its *structure*).[20]

Secondly, the non-uniqueness of the prediction of the factor scores $\widehat{\boldsymbol{f}}$ based on the point estimates $\widehat{\boldsymbol{\Lambda}}$, the sample of multivariate observations $\boldsymbol{x}$ and its associated sample covariance matrix $\boldsymbol{S}$. Essentially, the k-factor model in (1)-(2) postulates a relationship in which variables are dependent on factors, the latter remaining *unobserved*. However, computing factor scores for each observation of the multivariate sample provides a useful summarising device of individual performance (as we will see in the forthcoming sections). By assuming normality for the conditional distribution of factors, given the observed variables, we may predict factor scores for each country in the original dataset by computing $\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{\Lambda}}\boldsymbol{S}^{-1}\boldsymbol{x}$. This notwithstanding, different methods for obtaining $\widehat{\boldsymbol{f}}$ are available which lead to alternative results (for a detailed discussion, see Everitt and Hothorn, 2011, p. 148).

To sum up, EFA is the main tool we have used in this paper to analyse our CIS dataset. Section 3 below describes the empirical strategy adopted to articulate EFA and its results in order to answer the research questions posed in Section 1.

---

[18] In fact, the estimate for the variance of the variable-specific disturbance term $\hat{V}(u_i)$ is obtained as a residual. This may give rise to *Heywood* cases: the point estimate of the diagonal terms in $\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Lambda}}^T$ may exceed the sample variance of the manifest variable resulting in a negative estimate for $\hat{V}(u_i)$ (for details, see Everitt and Hothorn, 2011).

[19] These two latter features, i.e. number of factors and share of variance accounted for, should be taken into consideration when interpreting results, especially when comparing EFA with other data reduction techniques, such as Principal Component Analysis (PCA).

[20] Given an *orthogonal* matrix $\boldsymbol{M}$, a factor rotation is a linear transformation applied to factor loadings $\boldsymbol{\Lambda}$ that leaves the covariance matrix $\boldsymbol{\Sigma}$ in (4) unaltered: $(\boldsymbol{\Lambda M})(\boldsymbol{\Lambda M})^T = \boldsymbol{\Lambda}(\boldsymbol{MM}^T)\boldsymbol{\Lambda}^T = \boldsymbol{\Lambda}(\boldsymbol{MM}^{-1})\boldsymbol{\Lambda}^T = \boldsymbol{\Lambda\Lambda}^T$.

3 Empirical strategy

We adopted an empirical strategy consisting in four steps.

3.1 Exploratory Factor Analysis (EFA)

We perform an EFA for each of the following four dimensions (the corresponding number of CIS variables in parenthesis): (i) innovation inputs and demand sources (10), (ii) the geography and type of cooperation links (7), (iii) the role of the government and public sector policies (8), and (iv) innovation outputs (8). For each (i)-(iv) we fit a k-factor model — as specified in (3) — to a sample of multivariate observations for 26 countries.[21]

As mentioned above, to obtain the point estimates of the matrix of factor loadings $\mathbf{\Lambda}$ in (3) we apply maximum likelihood (ML), which is a scale-free estimation method (Timm, 2002, p. 504) that allows to successively increase the number of factors k (starting from $k = 1$), according to the statistical significance of the statistic associated to the hypothesis that k common factors are sufficient to describe the structure of correlations observed in the data.[22]

As a data preparation procedure, we standardise all data points by subtracting the sample mean and dividing by the standard deviation for each original variable.[23]

Once obtained the point estimates, we adjust factor loadings applying the *oblimin* 'rotation', which is an oblique transformation that allows for correlation between factors (rather than imposing an orthogonal rotation).[24] Adopting this transformation implies that our solution in each case (i)-(iv) now consists of three matrices:

$$\underset{(q\times k)}{\widehat{\mathbf{\Gamma}}} = \underset{(q\times k)}{\widehat{\mathbf{\Lambda}}^{*}} \times \underset{(k\times k)}{\widehat{\mathbf{\Phi}}} \tag{5}$$

[21] It has to be borne in mind that including variables that are implicitly contained in other variables should be avoided in factor analysis. For example, consider including a set of variables measuring the percentage of firms engaged in alternative types of innovation cooperation, *as well as* a variable quantifying firms engaged in *any type* of cooperation. The latter variable should be excluded, otherwise factors that load highly on cooperation measures will be artificially higher (see e.g. Shrolec and Verspagen, 2008).

[22] Usually, studies using firm-level CIS data avoid the recourse to maximum likelihood factor analysis, due to the fact that binary and Likert-type variables do not conform to the hypothesis of multivariate normality of the underlying data (e.g. Shrolec and Verspagen, 2008). However, unlike in most of the extant literature applying EFA to CIS-like data, we consider continuous variables, making this estimation method particularly fit for our purposes.

[23] Recall that factor analysis is unaffected by the rescaling of the original variables.

[24] The oblimin transformation is particularly apt for solutions obtained with ML. ML imposes a restriction on the diagonal character of $\mathbf{\Lambda}^{T}\boldsymbol{\sigma}_u^{-1}\mathbf{\Lambda}$, so an oblique transformation improves the description of the results (Raykov and Marcoulides, 2008, p. 268). Moreover, it has been noted that orthogonal rotations may often lead to biased results (Shrolec and Verspagen, 2008).

where $\widehat{\boldsymbol{\Gamma}}$ is the *structure* matrix, $\widehat{\boldsymbol{\Lambda}}^*$ the *pattern* (loadings) matrix, and $\widehat{\boldsymbol{\Phi}}$ the *factor intercorrelation* matrix. Essentially, elements of $\widehat{\boldsymbol{\Gamma}}$ provide the correlation coefficients between the latent factors and the observed variables, elements of $\widehat{\boldsymbol{\Lambda}}^*$ are the regression coefficients that, multiplied by (transformed) factors, give us the observed variables, and elements of $\widehat{\boldsymbol{\Phi}}$ quantify the correlation between factors.[25]

We interpret the fitted model results on the basis of matrix $\widehat{\boldsymbol{\Lambda}}^* = [\hat{\lambda}^*_{ij}]$. A high factor loading coefficient $\hat{\lambda}^*_{ij}$ indicates that, for a given correlation structure between factors, the observed variable x_i has a high (linear) association with factor f_j, so we say that variable x_i 'shapes' factor f_j. We group variables $i = 1, \ldots, q$ into subsets according to how their corresponding factor loading coefficients shape different factors. The oblimin transformation produces a simple pattern matrix that allows to unambiguously allocate each observed variable to one of the factors identified (in most cases). This way, factor definitions within each dimension are described on the basis of their constituting elements. The label attributed to each factor mirrors our interpretation of the relative importance of the variables that shape it.

Once estimated the parameters of the structural model and the rotation of the factors, we predict factor scores for each country in the data set. To do so, we fit a linear regression model using the point estimate of the structural matrix $\widehat{\boldsymbol{\Gamma}}$ and the sample covariance matrix $\boldsymbol{S}$.[26] This way, we obtained for each dimension (i)-(iv) a set of factor scores that summarises the performance of each country in terms of the subset of variables composing each factor.

3.2 Country-rankings based on factor scores

Some studies applying EFA on firm-level CIS data perform a two-stage analysis: they first obtain a set of factors on the basis of observed variables, and then fit a k-factor model on the *first-stage* estimates of factor scores to obtain a new (reduced) set of common (latent) factors (e.g. Shrolec and Verspagen, 2008).

Instead, given the initial conceptual distinction we made amongst the four dimensions (i)-(iv) defined in the previous subsection, we proceed differently.

[25] The oblimin 'rotation' procedure consists in applying a nonsingular transformation matrix $\boldsymbol{T}$ such that $\boldsymbol{f}^* = \boldsymbol{T}\boldsymbol{f}$ and $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\boldsymbol{T}^{-1}$ in (3). Moreover, the population covariance matrix implied by the model in (4) becomes: $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + \boldsymbol{\sigma}_u$, where $\boldsymbol{\Phi}$ is the population factor inter-correlation matrix. For details see Timm (2002).

[26] Note that the least squares point estimate of *factors* coincides with the maximum likelihood estimation. Therefore, predicting factor scores applying OLS to a linear regression model "is consistent with finding ML estimates of the parameters of the EFA model" (Timm, 2002, p. 510).

For each factor we rank (decreasingly) the country scores, averaged across factors within each dimension — using the proportion of variance explained by each factor as weight — to obtain an average country ranking with its summary performance for each dimension (i)-(iv).

Then, we qualitatively analyse differences in country rankings across dimensions. For example, we jointly consider country rankings for dimensions (i) innovation inputs and demand sources and (iv) innovation outputs. Our interest lies in understanding how countries that have a (relatively) high/low ranking position in terms of innovation inputs perform in terms of innovation outputs.

Proceeding in a similar manner, we study the relationships between all the relevant combinations between dimensions: government policies and innovation output, cooperation links and innovation output, government policies and innovation inputs, government policies and cooperation links and, finally, cooperation links and innovation inputs.

3.3 Correlation of factor scores

The third step of our analysis consists in analysing the correlations between factor scores (pooling all dimensions) across countries. Starting from a matrix with the countries in rows and factor scores in columns, we computed the correlation matrix between factor scores across countries. We selected those off-diagonal values greater than 0.65 and visualised the strength of (linear) associations between factors. On the basis of this visualization we provide some interpretation on the relationship between the different factors identified with the EFA.

3.4 Country Clustering

Finally, the fourth step in our analysis consists in applying a K-means clustering algorithm (Hartigan and Wong, 1979) to a matrix with the countries in rows and the factor-specific ranking position in columns. In this case we aim at obtaining a partition of 26 countries into a set of (mutually exclusive) clusters.

4 Exploratory Factor Analysis: Results

Our exploratory factor analysis sheds light on four dimensions of national innovation systems in Europe: (i) innovation inputs and demand sources, (ii) the geography and type of cooperation links, (iii) government role and public sector policies, and (iv) innovation outputs.[27]

4.1 Firm innovation inputs and demand sources

For the analysis of firm innovation inputs we have considered variables related to firm R&D expenditure (intensity, composition in terms of external/in-house and manufacturing component), knowledge acquisition, key sources of demand (local/regional, national and EU markets) and ownership structure (whether the firm merged with/took over other firms and/or if it is part of a group).

Table 2 reports the results on factor analysis of firm innovation inputs and demand sources. The table is composed of four panels (A)-(D) with the pattern (loadings) matrix, the factor intercorrelation matrix, the inferential procedure to test the adequacy of the number of factors identified and the country ranking for each factor (and across factors), respectively.[28]

The pattern matrix $\widehat{\boldsymbol{\Lambda}}$ of equation (5) in panel (A) displays the factor loading (in columns) for each manifest variable (in rows).[29] We have identified 3 factors (columns of the panel) that jointly explain 61% of the total variance in the correlation structure between variables. The last row of the panel — `Cumulative Var` — shows the cumulative proportion of variance explained by each of the factors identified.

The first factor identified, `iMarket`, loads high on the geographic origin of the largest demand source for each firm: EU, local/regional and national customers, respectively. The second factor, `iFirmStr`, has high loadings on two aspects of firms' ownership structure: whether the firm merged or took over other firms and whether it is part of an enterprise group. The third factor, `iRD`, loads high on variables related to the intensity, composition and sectoral structure of R&D expenditure. Finally, the observed variable 'Acquisition of external knowledge' has not been allocated to any of the three factors above-mentioned.

In panel (C) we look at the result of testing the (null) hypothesis that 3 factors are sufficient to describe the correlation structure between the manifest

[27] For this and the forthcoming subsections see Table 1 for a detailed description of the variables considered.

[28] The tables for each of the remaining exploratory factor analyses have the same layout as Table 2.

[29] We avoid displaying negligible values in the pattern matrix to ease reading of the results.

variables. With a p-value of 0.868 we do not reject the null hypothesis.

Note that each of the factors accounts for a similar proportion of explained variance: 21.8%, 20.1% and 19.2%, respectively. Thus, the variability in firm innovative inputs can be described by three latent factors of similar weight: the geographical source of demand, firm structure and R&D-related variables.

One important aspect of the oblique transformation chosen to describe the results consists in quantifying to what extent the three factors identified are correlated with each other. As can be seen from the factor inter-correlation matrix $\widehat{\mathbf{\Phi}}$ in panel (B), factor correlations are always below 0.5, being relatively higher when considering the correlation between firm structure and the two other factors.

Demand, firm structure and R&D expenditure are three key components of a NIS. The structure matrix $\widehat{\mathbf{\Gamma}}$ — obtained by multiplying the pattern matrix $\widehat{\mathbf{\Lambda}}$ and the factor inter-correlation matrix $\widehat{\mathbf{\Phi}}$ in equation (5) — may be used to predict factor scores, which are employed as a summarising device to make international comparisons. As reported on panel (D), on the basis of countries' scores for each factor, we compute factor-specific country-rankings and an overall average country-ranking using the proportion of total variance explained by each factor to weight factor-specific country-rankings.

If we focus on the first 15 positions of the average ranking (column `Rank` of the panel), it emerges that only four countries are in the highest 10 positions for all the three factors identified: Sweden (SE), Germany (DE), France (FR) and Belgium (BE). A second group of countries is amongst the first 10 positions in two out of three factors: Norway (NO), Ireland (IE) and Austria (AT) in factors `iMarket` and `iFirmStr`, whereas Denmark (DK) and Finland (FI) in factors `iFirmStr` and `iRD`. Finally, a third group of countries features prominently in only one of the identified factors: Netherlands (NL), Czech Republic (CZ) and Spain (ES) in `iRD`, whereas Italy (IT), Slovenia (SI) and Greece (EL) in `iMarket`.

This is a first piece of evidence that will be considered, together with the following country rankings along the remaining factors, in composing a picture that positions subsets of countries in terms of strength in different NIS dimensions. For instance, when considering the mid-positioned countries in terms of innovation inputs, we observe that some of them might rank high in terms of R&D intensity but locate much lower in terms of destination markets. The top four enjoy strong private expenditures in R&D as well dynamic destination markets.

Table 2: Factor Analysis: Firm innovation inputs and demand sources

Panel (A) Firm innovation inputs & demand conditions

Pattern (loadings) matrix:

Code	Label	iMarket	iFirmStr	iRD
INNO_PPANPP_LARMAR_EU	Largest market: EU	**0.972**	-0.186	
INNO_PPANPP_LARMAR_LREG	Largest market: Local/Regional	**0.843**	0.255	
INNO_PPANPP_LARMAR_NAT	Largest market: National	**0.690**	0.224	0.104
INNO_PPANPP_ENMRG_YES	Firm merged/took over		**0.977**	
INNO_PPANPP_GP_YES	Firm part of enterprise group		**0.770**	0.168
INNOACT_EXPTOT14_ENT_POPU14	R&D expenditure per firm		0.446	**0.621**
INNOACT_RRDEX14_PC	Share of external R&D			**0.923**
INNOACT_RRDIN14_PC	Share of in-house R&D	0.120	0.264	**0.565**
INNOACT_EXPTOT14_C	Manufacturing/Aggregate R&D	-0.133		**0.555**
INNOACT_ROEK14_PC	Acquisition of external knowledge		0.158	
SS loadings		2.180	2.005	1.916
Proportion Var		0.218	0.201	0.192
Cumulative Var		0.218	0.418	**0.610**

Panel (B) Factor dictionary and intercorrelation matrix:

		iMarket	iFirmStr	iRD
iMarket	Main source of demand	1.000	0.492	0.164
iFirmStr	Firm ownership structure	0.492	1.000	0.450
iRD	R&D intensity and composition	0.164	0.450	1.000

Panel (C) Hypothesis testing on the number of factors

Test of the hypothesis that 3 factors are sufficient.

The chi square statistic is 11.58 on 18 degrees of freedom.

The p-value is 0.868

We do not reject the null hypothesis that 3 factors are sufficient to describe the correlation structure between manifest variables

Panel (D) Country Ranking across factors:

	iMarket	iFirmStr	iRD	Mean	**Rank**
SE	10	1	2	4.4	1
NO	2	2	11	4.8	2
DE	4	4	7	4.9	3
FR	7	6	3	5.4	4
BE	3	10	4	5.7	5
IE	1	5	14	6.4	6
AT	5	3	12	6.5	7
DK	13	9	1	7.9	8
FI	12	7	5	8.1	9
IT	8	11	16	11.5	10
NL	14	13	9	12.1	11
SI	9	15	13	12.3	12
CZ	16	16	10	14.1	13
EL	6	14	24	14.3	14
ES	17	20	6	14.6	15
HR	11	12	22	14.8	16
CY	18	8	25	16.7	17
HU	23	21	8	17.7	18
PT	19	18	20	19.0	19
LT	15	17	26	19.1	20
SK	20	22	15	19.1	21
EE	22	19	17	19.4	22
BG	21	25	18	21.5	23
PL	24	23	19	22.1	24
LV	25	24	23	24.0	25
RO	26	26	21	24.5	26

4.2 Firm cooperation links

For the analysis of firm cooperation links, we have considered the geographic boundaries of cooperation links (with actors within the country, EU, China, India and the US) as well as with other private actors (other firms within the enterprise group, competitors in the same sector, private clients/customers).[30]

Table 3 reports the results on factor analysis of firm cooperation links. The pattern matrix in panel (A) shows that we have identified 3 factors (columns of the panel) that jointly explain 73.5% of the total variance in the correlation structure between variables.

The first factor identified, `cEURNAT`, loads high on cooperation with national and EU partners as well as with firms within the enterprise group. The second factor, `cUSCNIN`, has high loadings on cooperation with the US, China (CN) and India (IN). The third factor, `cCCC`, loads high on cooperation links with competitors, (private) clients and customers.

In panel (C) we look at the result of testing the (null) hypothesis that 3 factors are sufficient to describe the correlation structure between the manifest variables. With a p-value of 0.302 we do not reject the null hypothesis.

Note that there is a marked *asymmetry* of (around) 10 percentage points in the proportion of variance explained by each factor: `cEURNAT` accounts for 35.3%, `cUSCNIN` for 24.5% and, finally, `cCCC` for 13.7%. Thus, the variability in firm cooperation links can be described by three latent factors of decreasing (relative) importance: links with national/EU and enterprise group partners feature prominently, cooperation with global players outside the EU are less relevant (to explain cross-country variability) and a minor role can be ascribed to cooperation links forged with competitors and customers.

The oblique transformation chosen to describe the results allows to mediate the correlations between observed variables and identified factors by distinguishing the inter-factor correlations. In this case, as can be seen from panel (B), factor correlations are always above 0.6, being particularly high between national/EU and outer-EU countries (factors `cEURNAT` and `cUSCNIN`, respectively). Thus, in view of the fact that these latter two factors also account for most of the explained variance, the geographical domain of cooperation links seems to be a prominent aspect of the national innovation systems analysed.

[30]Cooperation links involving public institutions are considered in the factor analysis concerning government role and public sector policies.

Table 3: Factor Analysis: Firm cooperation links

Panel (A) Firm cooperation links

Pattern (loadings) matrix:

Code	Label	cEURNAT	cUSCNIN	cCCC
INNOACT_C01	Cooperation within the enterprise group	**0.876**	0.259	-0.146
INNOACT_COEUR_YES	Cooperation with EU partners	**0.836**		
INNOACT_CONAT_YES	Cooperation with National partners	**0.958**	-0.163	0.177
INNOACT_COCNIN_YES	Cooperation with China/India		**0.806**	0.138
INNOACT_COUS_YES	Cooperation with the US		**0.907**	
INNOACT_C02	Cooperation with competitors, same sector		0.152	**0.794**
INNOACT_C031	Cooperation with private clients/customers	0.271	0.338	**0.502**
SS loadings		2.469	1.712	0.961
Proportion Var		0.353	0.245	0.137
Cumulative Var		0.353	0.597	**0.735**

Panel (B) Factor dictionary and intercorrelation matrix:

		cEURNAT	cUSCNIN	cCCC
cEURNAT	Links to EU and national partners	1.000	0.768	0.707
cUSCNIN	Links to China, India, US	0.768	1.000	0.629
cCCC	Links to competitors, clients, customers	0.707	0.629	1.000

Panel (C) Hypothesis testing on the number of factors

Test of the hypothesis that 3 factors are sufficient.

The chi square statistic is 3.65 on 3 degrees of freedom.

The p-value is 0.302

We do not reject the null hypothesis that 3 factors are sufficient to describe the correlation structure between manifest variables

Panel (D) Country Ranking across factors:

	cEURNAT	cUSCNIN	cCCC	Mean	**Rank**
FI	4	2	1	2.7	1
BE	1	5	7	3.6	2
NO	2	6	4	3.8	3
SE	6	1	5	4.1	4
NL	5	7	3	5.3	5
AT	3	8	6	5.3	6
DK	9	3	9	7.0	7
IE	8	4	20	9.2	8
SI	10	10	8	9.6	9
FR	7	9	21	10.7	10
EE	12	13	11	12.1	11
EL	13	18	2	12.3	12
CY	16	11	10	13.0	13
LT	11	16	16	13.8	14
SK	15	12	15	14.0	15
CZ	14	14	18	14.9	16
DE	17	15	12	15.3	17
HR	19	17	14	17.3	18
PT	18	20	13	17.6	19
HU	21	19	17	19.5	20
ES	20	22	19	20.5	21
LV	23	21	23	22.3	22
IT	22	25	22	23.0	23
BG	25	23	24	24.1	24
PL	24	24	26	24.4	25
RO	26	26	25	25.8	26

As reported on panel (D), if we focus on the first 15 positions of the average ranking (column `Rank` of the panel), 8 countries are in the highest 10 positions for all three factors identified: Finland (FI), Belgium (BE), Norway (NO), Sweden (SE), Netherlands (NL), Austria (AT), Denmark (DK) and Slovenia (SI). This is related to the relatively high (and positive) inter-factor correlations, so that above-average performance in one factor tends to correspond with a similar outcome for other factors as well. Instead, Ireland (IE) and France (FR) are amongst the first 10 countries as regards the two factors that account for the greatest share of the explained variance (`cEURNAT` and `cUSCNIN`), though have a relatively lower score in factor `cCCC`.

This second piece of evidence focusing on the international openness in cooperation of firms shows a higher homogeneity in countries' ranking along the different factors, with a prominent position of Scandinavian countries in the scope and intensity of cooperation at the national/EU and outer-EU levels.

4.3 Government innovation policies

For the analysis of government innovation policies, we have considered the public sector actors which firms cooperate with (Government, Higher Education Institutions/Research Institutes and public sector clients/customers), public funding sources (Local/Regional Authorities and Central Government) as well as source of public procurement, i.e. domestic or foreign. Within the latter, we have also considered whether foreign procurement required innovation activities as part of the contract.

Table 4 reports the results on factor analysis of the role of government and public innovation policies, as captured by the CIS-2014. The pattern matrix in panel (A) shows that we have identified 3 factors (columns of the panel) that jointly explain 69.4% of the total variance in the correlation structure between variables.

The first factor identified, `gGvtFCo`, loads high on cooperation links between firms and government agencies, research institutes, universities, other higher education institutions as well as public sector clients/customers. Moreover, this first factor also loads high on funding received from the central government. The second factor, `gLRFDoPr`, has high loadings on funding received by firms from local/regional authorities and procurement contracts coming from domestic sources. Finally, the third factor, `gForPr`, loads high on variables related to foreign procurement contracts involving domestic firms.

Table 4: Factor Analysis: Government innovation policies

Panel (A) Government innovation policies

Pattern (loadings) matrix:

Code	Label	gGvtFCo	gLRFDoPr	gForPr
INNOACT_FUNGMT	Funding from Central Government	**0.781**		
INNOACT_C032	Coop. with public sector clients/customers	**0.572**		0.261
INNOACT_C06	Cooperation with universities/HEI	**0.846**		0.137
INNOACT_C09	Cooperation with Gvt/Research Inst.	**0.957**		
INNOACT_FUNLOC	Funding from Local/Regional Auth.	0.202	**0.775**	-0.183
TOTAL_PUBDOM	Domestic Procurement		**0.932**	0.162
TOTAL_PUBFINRQ	Foreign proc. req. innovation activities	0.35		**0.625**
TOTAL_PUBFOR	Foreign Procurement		0.119	**0.886**
SS loadings		2.738	1.491	1.326
Proportion Var		0.342	0.186	0.166
Cumulative Var		0.342	0.529	**0.694**

Panel (B) Factor dictionary and intercorrelation matrix:

		gGvtFCo	gLRFDoPr	gForPr
gGvtFCo	Fund. Gvt /Coop. HEI/Research Inst.	1.000	0.663	0.696
gLRFDoPr	Local/Reg. Fund. / Dom. Procurement	0.663	1.000	0.607
gForPr	Foreign Procurement	0.696	0.607	1.000

Panel (C) Hypothesis testing on the number of factors

Test of the hypothesis that 3 factors are sufficient.

The chi square statistic is 7.75 on 7 degrees of freedom.

The p-value is 0.355

We do not reject the null hypothesis that 3 factors are sufficient to describe the correlation structure between manifest variables

Panel (D) Country Ranking across factors:

	gGvtFCo	gLRFDoPr	gForPr	Mean	**Rank**
BE	2	2	1	1.7	1
FI	1	3	3	2.0	2
AT	3	4	2	3.0	3
NO	4	5	6	4.8	4
SE	8	9	5	7.5	5
IE	9	1	12	7.7	6
SI	7	13	7	8.6	7
DE	5	6	22	9.7	8
NL	6	21	9	10.7	9
DK	11	11	10	10.7	10
FR	10	7	16	10.8	11
LT	18	12	4	12.8	12
EL	12	10	18	13.0	13
EE	16	18	8	14.5	14
CZ	13	16	17	14.8	15
PT	14	19	15	15.6	16
HR	21	8	14	15.8	17
CY	20	14	11	16.1	18
ES	15	15	23	17.1	19
SK	19	22	13	18.2	20
IT	17	20	24	19.6	21
HU	23	17	19	20.4	22
LV	24	23	21	23.0	23
PL	22	25	26	23.8	24
BG	26	24	20	23.9	25
RO	25	26	25	25.3	26

In panel (C) we look at the result of testing the (null) hypothesis that 3 factors are sufficient to describe the correlation structure between the manifest variables. With a p-value of 0.355 we do not reject the null hypothesis.

Note that there is a marked *asymmetry* of (around) 16-18 percentage points in the proportion of variance explained by the first factor in relation to the remaining two: `gGvtFCo` accounts for 34.2%, `gLRFDoPr` for 18.6% and, finally, `gForPr` for 16.6%. Thus, the first-order cross-country variability in the role of government and public sector policies may be described by one latent factor that involves innovation cooperation links with public sector (and related) agents as well as funding from central government sources. The two remaining factors involving local/regional funding and public procurement describe a relatively smaller, though still relevant, share of covariance between the manifest variables.

As can be seen from panel (B), the oblique transformation chosen leads to inter-factor correlations that are between 0.6 and 0.7 in all cases and, in particular, correlation is relatively higher between the factor that accounts for the majority of explained variance (`gGvtFCo`) and the other two factors (`gLRFDoPr` and `gForPr`).

As reported on panel (D), if we focus on the first 15 positions of the average ranking (column `Rank` of the panel), only five countries are in the highest 10 positions for all three factors identified: Belgium (BE), Finland (FI), Austria (AT), Norway (NO) and Sweden (SE). Instead, five countries are within the first 10 positions in only one other factor beyond `gGvtFCo`: Ireland (IE), Germany (DE) and France (FR) for factor `gLRFDoPr`, whereas the Netherlands (NL) and Slovenia (SI) for factor `gForPr`. Most of the rest of countries within the top 15 average ranking positions feature amongst the highest 10 ranking positions in only one of the factors identified.

4.4 Firm innovation outputs

For the analysis of innovation outputs we have considered firm-level average productivity, patent applications and registration of trademarks, innovation persistence, as well as the introduction of radical/incremental product and process innovations.

Table 5 reports the results on factor analysis of innovation outputs. The pattern matrix in panel (A) shows that we have identified 4 factors (columns of the panel) that jointly explain 73% of the total variance in the correlation structure between variables.

Table 5: Factor Analysis: Firm innovation outputs

Panel (A) Firm innovation ouptuts

Pattern (loadings) matrix:

Code	Label	oRadPat	oIncrPcs	oOng	oPtvty
INNO_PROPAT	Application for a patent	**0.672**		0.243	
INNO_PROTM	Registration of a trademark	**0.876**		0.133	
INPDT_NEWMAR_YES	Turnover from porducts new to market	**0.855**	0.113	-0.204	0.16
INPDT_NEWFRM_YES	Turnover from porducts new to firm		**0.923**	0.142	
INPCS_INPSNM0	Process innovation new to firm		**0.756**		
INPCS_INPSNM1	Process innovation new to market		**0.686**	-0.163	0.299
INONG_ENT_POPU14	Ongoing innovation activities			**0.924**	0.111
INNO_TURN_EMP	Turnover per employee			0.118	**0.886**
SS loadings		1.970	1.912	1.035	0.924
Proportion Var		0.246	0.239	0.129	0.116
Cumulative Var		0.246	0.485	0.615	**0.730**

Panel (B) Factor dictionary and intercorrelation matrix:

		oRadPat	oIncrPcs	oOng	oPtvty
oRadPat	Radical Prod. Innov. / Patent App.	1.000	0.792	0.441	0.639
oIncrPcs	Incr. Prod. / Rad. Proc. Innov.	0.792	1.000	0.162	0.570
oOng	Ongoing innovation	0.441	0.162	1.000	0.259
oPtvty	Productivity	0.639	0.570	0.259	1.000

Panel (C) Hypothesis testing on the number of factors

Test of the hypothesis that 4 factors are sufficient.

The chi square statistic is 3.73 on 2 degrees of freedom.

The p-value is 0.155

We do not reject the null hypothesis that 4 factors are sufficient to describe the correlation structure between manifest variables

Panel (D) Country Ranking across factors:

	oRadPat	oIncrPcs	oOng	oPtvty	Mean	**Rank**
BE	1	5	3	3	3.0	1
NO	4	7	4	2	4.6	2
DE	8	2	1	12	5.6	3
FI	5	4	11	7	6.0	4
SE	3	15	2	5	7.0	5
AT	2	6	17	10	7.2	6
IE	9	1	23	1	7.4	7
NL	6	11	6	8	7.9	8
FR	7	12	9	6	8.8	9
IT	11	17	5	9	11.6	10
PT	13	8	16	15	12.3	11
CY	16	3	26	13	13.0	12
SI	10	9	21	19	13.1	13
DK	15	18	10	4	13.2	14
CZ	12	14	8	21	13.5	15
EL	14	10	20	14	13.7	16
ES	21	20	7	11	16.6	17
LT	17	13	19	25	17.4	18
SK	18	23	13	16	18.4	19
HU	19	21	14	17	18.4	20
HR	20	16	25	24	20.3	21
BG	22	19	15	26	20.5	22
EE	25	26	12	20	22.3	23
PL	24	24	22	18	22.6	24
LV	23	22	24	22	22.7	25
RO	26	25	18	23	23.8	26

The first factor identified, `oRadPat`, loads high on patent applications, registration of trademarks and turnover from radical product innovations.[31] The second factor, `oIncrPcs` has high loadings on incremental product innovations and incremental/radical process innovations. The last two factors, load high each on a single manifest variable: factor `oOng` on the persistence of innovation activities, whereas factor `oPtvty` on turnover per employee (i.e. average gross output labour productivity).

In panel (C) we look at the result of testing the (null) hypothesis that 4 factors are sufficient to describe the correlation structure between the manifest variables. With a p-value of 0.155 we do not reject the null hypothesis.

Interestingly, there is a marked *asymmetry* in the proportion of variance explained by the first two factors (`oRadPat` and `oIncrPcs`) with respect to the remaining two factors (`oOng` and `oPtvty`). While the first two factors account for 48.5% of the 73% of variance explained (each almost equally contributing with 24 percentage points), the remaining two factors account for only 24.5% of the variance explained (`oOng` accounting for 12.9% and `oPtvty` for 11.6%, respectively). Thus, patents, trademarks and radical/incremental product/process innovations account for a higher share of the observed cross-country variability than the ongoing character of innovation and gross output labour productivity.

As panel (B) shows, the oblique transformation chosen gives rise to inter-factor correlations which are relatively higher between factors `oIncrPcs` and `oRadPat`, but *also* between the latter factor and `oPtvty`. Thus, the factor composed of radical product innovation and patent applications is strongly positively correlated with labour productivity.

As reported on panel (D), if we focus on the first 15 positions of the average ranking (column `Rank` of the panel), 9 countries are within the highest 10 positions for all four factors identified: Belgium (BE), Norway (NO), Germany (DE), Finland (FI), Sweden (SE), Austria (AT), Ireland (IE), Netherlands (NL) and France (FR). Instead, two countries are within the first 10 positions in only two factors (`oOng` and `oPtvty`): Italy (IT) and Denmark (DK), whereas the remainder of top 15 countries scores relatively high in only one of the factors identified.

Once again, this last piece of evidence on the dimension representing innovation (and economic) performance of firms in countries shows that the same core countries ranking high on the first three NIS dimensions described above

[31] In what follows, we assume that the introduction of a product (process) new to the market corresponds to a radical product (process) innovation, whereas the introduction of a product (process) new to the firm (not new to the market) corresponds to an incremental product (process) innovation.

also rank high in terms of outcomes. We go more in depth into unpacking this evidence by looking at how the different NIS dimensions are related to each other in the next section.

5 Mapping the relations among different dimensions of European NIS

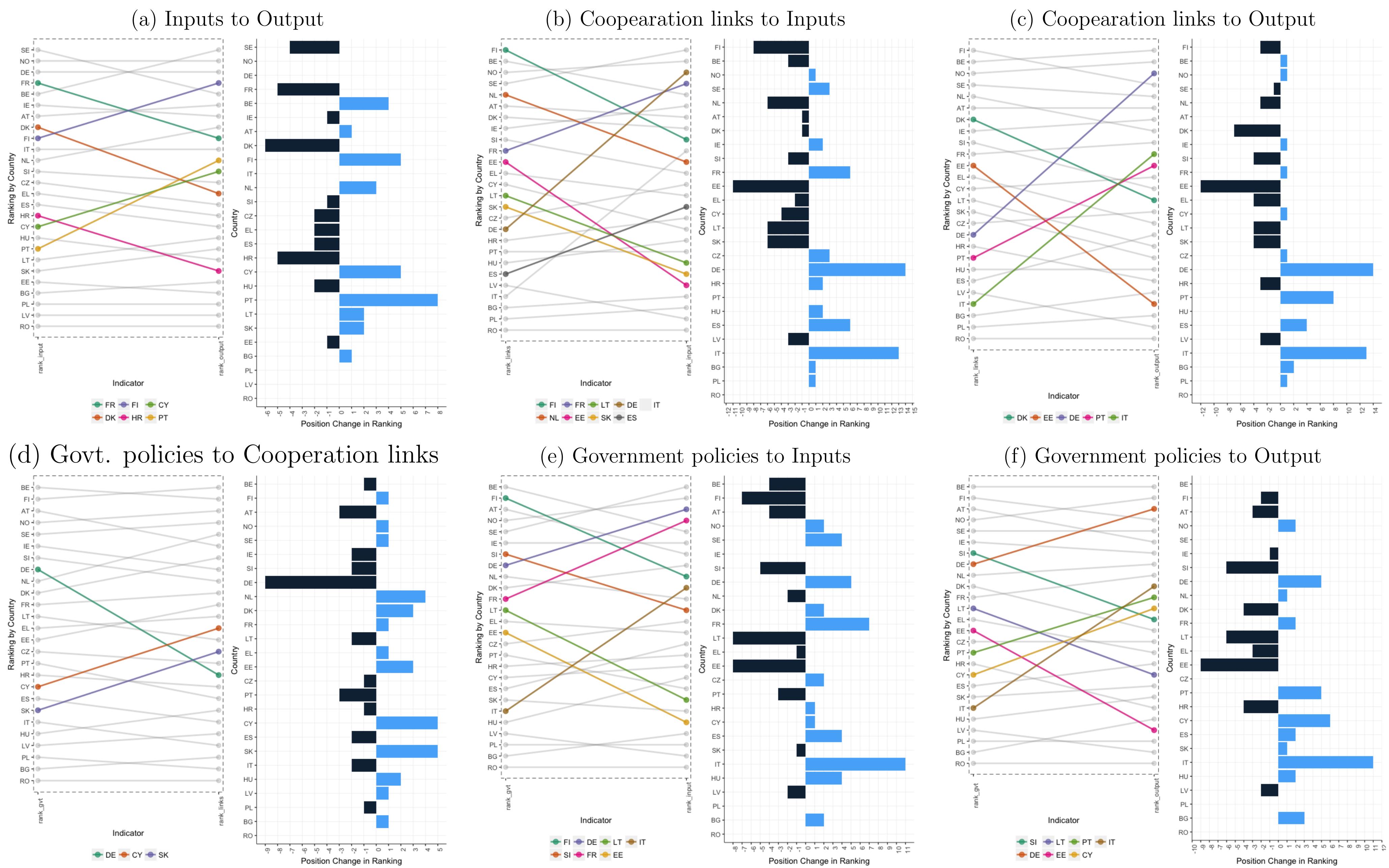
5.1 Comparing Country Rankings across Dimensions

We have shown that country rankings alongside the four NIS dimensions empirically derived in the previous section are relatively similar across countries, although a few important differences exist. In Figure 6, country rankings are shown for pairs of dimensions. For each pair, the graph on the left visualises the ranking of each country for each of the two dimensions. The graph on the right plots the difference (gain or loss) in ranking positions between the first and the second dimension. In the present subsection we look at these differences to describe and discuss the relations between different NIS dimensions. In the subsection that follows we propose a more systematic analysis of the correlations between factors.

Innovation Inputs and Outputs

The first question that we ask is: do countries whose firms are top ranked in terms of innovation inputs (R&D expenditure, demand, and firm organisation) also achieve the best performance in terms of innovation output (radical product innovation, process innovations, persistent innovation, and labour productivity)? Figure 6 panel (a) shows that most countries that rank low on innovation inputs rank low also on innovation output, as expected. As we move to countries in the middle of the ranking distribution, a few countries (e.g. Portugal and Cyprus) seem to gain more than others from relatively low innovation inputs. At the top of the distribution, input/output rankings are quite stable, with a couple of remarkable exceptions: Finland and Belgium score significantly better in terms of innovation output than in terms of inputs, whereas Denmark, France, and Sweden fare relatively low in relation to their innovation inputs.

Figure 6: Ranking differences between dimensions



(a) Inputs to Output

(b) Coopearation links to Inputs

(c) Coopearation links to Output

(d) Govt. policies to Cooperation links

(e) Government policies to Inputs

(f) Government policies to Output

Firm Cooperation, Inputs, and Outputs

Next, we investigate if country differences in terms of firms' innovation input and output may be related to how firms cooperate with other firms and where they tend to cooperate. The second question that we ask is: do countries that rank high with respect to firms' innovation linkages (cooperation with other firms in the same enterprise group, country, EU, US, emerging countries, competitors, and clients) also perform well with respect to innovation inputs and outputs? The relation between collaboration and innovation inputs/outputs is more unstable than the relation between innovation inputs and outputs.

Figure 6 panel (b) shows country rankings on collaborations and innovation inputs (and country differences between the two). On the one hand, in some of the top ranked countries for innovation inputs (such as Germany and France) firms declare below average collaborations with other firms. In the case of Germany (and Italy), for example, this is the case for all collaborations, whereas in the case of France, it is with competitors and clients that firms reach close to the lowest scores. On the other hand, some countries in which there is a large proportion of firms with strong collaboration ties, have a weaker performance in terms of innovation inputs, e.g. Finland and the Netherlands. Surprisingly the main factor in which they do not excel in innovation inputs is international demand: firms in these countries tend to collaborate more than average with clients and competitors across the globe, but are in an average ranking position with respect to demand links.

The difference in the rankings between innovation inputs and collaborations suggests a story of substitution between the two. In several countries firms either invest in collaborations, or in R&D and demand. Only in a few countries (especially at the top and the bottom of the distribution) innovation inputs and collaborations are similarly high or low. This trade-off may be related to the shift from producer to open innovation (Baldwin and von Hippel, 2011), or to the fact that firms and countries may choose to follow the tide or specialise in different innovative activities (Adams, 2012), while these results seem to put in a different perspective those emerging from earlier research on the complementarity between internal and external knowledge (Caloghirou et al., 2004). However, Love and Roper (2001) find very similar results on complementarity for UK, Irish and German firms. The question is open for further micro and macro research.

For what concerns the complementarity between collaborations and innovation outputs, Figure 6 panel (c) shows country rankings for collaborations and innovation outputs (and country differences between the two). Unlike the case of inputs, here we find strong complementarity for the large majority of countries, but also a few cases of stark differences between the extent to which firms tend to collaborate and their innovation output.

Some examples correspond to the same countries discussed above: German and Italian firms have a reduced proportion of firms engaged in innovation cooperation (of all types and geographies) with respect to how they score in terms of innovation outputs. On the contrary, Estonian and Danish firms collaborate more than average, but their output performance is below average. In both cases this is especially related to radical product innovation and process innovation – more than ongoing innovation and labour productivity.

Our results complement firm-level evidence based on subsets of countries which finds that, in some cases, open innovation leads to better firm performance (Powell et al., 1996; Nieto and Santamara, 2007), only for a certain degree of openness (Berchicci, 2013), whereas in other cases there is no relation between cooperation ties (external links) and firm innovative performance (Love and Roper, 2001), providing a broader view at the European level.

Public Investment, Firm Cooperation, Inputs and Outputs

The final question that we ask is: how do public institutions (collaboration with public sector customers, universities and governmental organisations) and policies (funding from the central and local governments, national and foreign procurement) co-vary with innovation inputs, outputs, and collaborations? The answers differ, so we will take one by one.

Figure 6 panel (d) shows country rankings for public innovation policies and collaborations (and country differences between the two). The rankings are rather stable: there seems to be very little crowding out between public and private collaborations. For most countries in which firms enjoy public support to innovation and strong collaboration with public bodies, firms also have above-average collaborations with the private sector (any type and geography of collaboration).[32] The main exception to this finding is Germany where, as we have already observed, firms tend to collaborate with other firms less than on average. Whether this is because of the high level of public support, or because of the the low reliance on external knowledge, is left for further research

[32]This finding is supported by micro evidence. For instance, dne Cappelen et al. (2012) find that Norwegian firms that benefited from tax incentives to R&D innovated more, especially those who collaborated with other firms.

at the micro level.

Figure 6 panel (e) shows country rankings for public innovation policies and innovation inputs (and country differences between the two). The overall picture is not very different from the one discussed for private collaborations – unsurprisingly, given the stable ranking between public policies and private collaborations just examined, but the differences in ranking are smaller. With the exception of Italy, in the lower half of the distribution, countries with average or lower than average public support for firms' innovation also experience low innovation inputs. Italy is the only country in which firms experience quite low public support, and above average innovation input.

Symmetrically, all countries in which firms experience above average government support (in one form of another), also show above-average innovation inputs. Two Baltic countries on the fringe (Lithuania and Estonia) rank in the middle of the distribution with respect to public support, but are well below average with respect to innovation inputs.

Overall, the results seem to indicate that, from an aggregate perspective, and considering different public interventions – such as central and local funding, public private collaboration, and procurement — we observe little additionality, as well as little crowding out.

Countries in which additionality may operate are: Germany, France, Sweden, Norway, and Denmark (ordered by the difference between the ranking in government support and innovation inputs). Countries in which public support may crowd out innovation inputs are: Finland, Belgium, Austria, Slovenia and the Netherlands (ordered by the difference between the ranking in government support and innovation inputs).[33] However, to reiterate, differences in rankings are quite small, and given the micro-aggregated nature of these results, it is safer to argue that in countries where firms benefit from above-average public support, this in some cases leads to even higher performance in terms of innovation inputs, whereas in other cases the performance in innovation inputs does not match the extent of the public contribution.

[33] The micro empirical evidence partly support the macro picture. Aerts and Schmidt (2008) find no crowding out of public R&D subsidies on Flemish and German firms (similarly Czarnitzki and Licht (2006), focusing on German firms); Lööf and Hesmati (2004) find additionality for small firms in Sweden; Griffith et al. (2006) find a positive effect of central and local policies on R&D in Germany, France, Spain, and the UK; in Norway, Clausen (2009) find a positive effect of public support on firm research, but a negative results on firms development activities; Bloch and Graversen (2012) find a positive effect of R&D public funding on private R&D spending. However, positive effect of public spending on firm R&D was also found in the case of Belgium (Aerts and Czarnitzki, 2004), Austria (Falk, 2004), Slovenia (Jakli et al., 2013), and the Netherlands (e.g. Lokshin and Mohnen, 2012, although only for small firms). Overall, Ziga-Vicente et al. (2014) find that over 46 studies investigating additionality across EU countries, 33 find a positive effect, seven find no significant effect and six find a negative effect.

Finally, is there any relation between public support and innovation outputs? Figure 6 panel (f) shows country rankings for public innovation policies and innovation outputs (and country differences between the two). Also in this case, at the two extremes of the ranking distribution, innovation policies and outcomes are quite well aligned: countries in which firms benefit from relatively high public support also tend to score high in terms of innovation outputs. Conversely, countries were firms receive little public support for innovation, also tend to be on the low end of innovation performance. The main exception is again Italy: despite the relatively limited role of innovation policies the country ranks above average in terms of innovation outputs (as well as inputs). A few more countries (with stronger role for government policies) perform better in terms of output than innovation policies, including Germany. This is in line with evidence at the micro level (e.g. Czarnitzki and Licht, 2006) that public support of R&D has a significant positive effect on firms' patenting activity.

However, the same countries that rank quite lower in terms of innovation inputs than in terms of government policies also see a substantial difference in ranking positions with respect to outputs: for these countries (Slovenia, Lithuania and Estonia) innovation policies do not seem to lead to relevant firm innovation effort *nor* outcomes.

Overall, the countries ranking comparisons provide an interesting picture of homogeneity in ranking across the different NIS dimensions at the top and low ends of the distributions, with an even more interesting story of *exceptions* to this trend, which would deserve a more in-depth exploration, which we attempt by means of a cluster analysis in the section that follows. These are: Italy, which scores relatively low in terms of both public support and cooperation links, albeit it enjoys an above average-ranking both in private innovation inputs and outputs. German firms also make the country score relatively low in terms of cooperation, although they benefit from higher than average public support, but they score similarly high both in terms of input and output, where the presence of dynamic demand might be prominent in determining the ranking of the inputs. On the contrary, Denmark for instance, seems to score lower in innovation outputs than its ranking positions in innovation inputs and cooperation links would suggest.

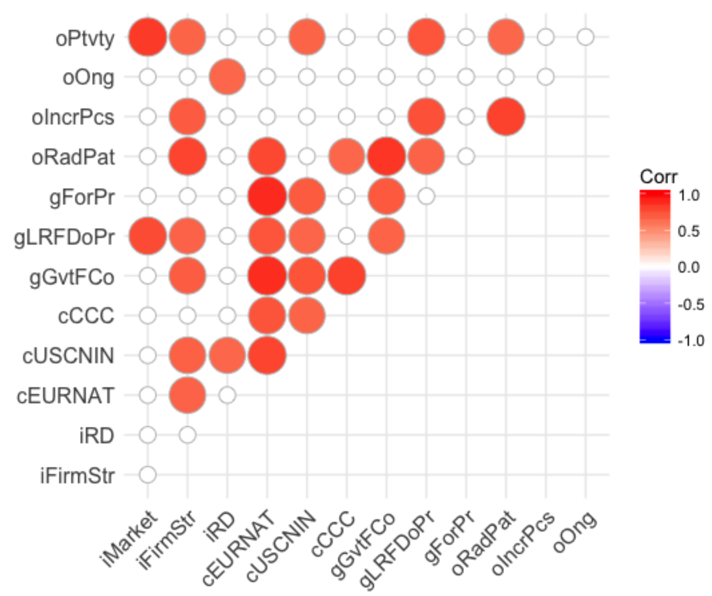
It seems opportune therefore to look at how the *factors* — synthesising the core dimensions of a NIS as derived above — are correlated with each other, which would help making sense of the comparison amongst country rankings just discussed.

5.2 Factor Correlations

To get a better understanding of which factors might explain the relation between innovation policies, firm cooperation, innovation inputs and outputs, we study the correlation amongst all factors, based on country rankings. Two factors are strongly correlated when all countries tend to have a similar ranking on both factors, and weakly correlated otherwise. Figure 7 plots the correlation matrix between every factor pair on a heat colour scale: the darker the circle, the strongest the correlation. To keep the figure readable we plot only the circles representing significant and positive correlations above 0.65.

First, we find that only some of the factors are significantly and positively correlated (the matrix is relatively sparse). Second, some factors are correlated with many other factors (lines and/or columns with many coloured circles), whereas others are related to only one or two other factors (sparse lines and/or rows).

Figure 7: Correlations between factors across countries



In particular, the factors that are mostly correlated with other dimensions of the innovation system are firm cooperation links, whether distant (with US, India and China), or close (with firms from the same enterprise group, country, or within the EU). However, firm cooperation is mainly correlated with other measures of cooperation or with government policies.

This is in line with the discussion on the comparison between country rankings, and suggests that countries in which firms have strong innovation links with other firms also invest substantially in supporting firm innovation. In other words, two crucial aspects of the innovation system seem to be strongly correlated: private and public interactions. More research is required to establish the causal links: does government support provide firms with the means to collaborate? Or do firms which invest in collaborations with private firms, also do so with governments and public research centres?

With respect to innovation inputs and outputs, distant and close collaborations differ. As expected, cooperation is correlated with innovation inputs pertaining to the same group. More interestingly, close collaborations, unlike distant ones, are correlated with radical innovations; distant collaborations, unlike close ones, are correlated with productivity and R&D expenditure.

Close and distant collaborations seem therefore to be complementary: in countries where firms are more often than average able to introduce radical innovations, they are also more likely to establish close collaborations; in countries where firms are more productive than average and spend more than average in R&D, are also more likely to gain from distant collaborations.

Instead, the third factor characterising the cooperation dimension of NIS — the 'inward looking' one with competitors and customers — is not correlated with the other dimensions of the innovation system. However, like close collaborations, it is correlated with radical innovations.

Public innovation policies are expectedly correlated with a sizeable number of factors, in particular central government funding, cooperation with public bodies, local government funding, and domestic procurement.

Besides the (already discussed) correlation with private actors, the figure shows that, on the one hand, central government funding and collaboration with public bodies is particularly correlated with radical innovation output. On the other hand, local government support to innovation and public (domestic) procurement seem to be more effective, being correlated with firms' success in radical innovation, incremental and process innovation, firm productivity, and firms' ability to access large shares of the market. This evidence surely deserves some attention.

To summarise, we have looked at how the different dimensions of NIS, represented by our empirically derived factors of innovation inputs, collaborations with other firms, public support and innovation inputs, are related to each other.

The main factors related to firms' ability to produce *radical innovations* are close collaborations with private firms (relatively close, and especially with customers and competitors) and with public bodies, which benefit from government support (central and local) as well as public procurement.

Process and incremental product innovation are less related to the rest of the innovation system. Countries in which firms introduce new to the firm products and new processes (incremental and radical) more than the average, also seem to benefit from local Government support and national procurement more than average. All the other aspects of the innovation system are not correlated.

Firm performance, proxied here by *labour productivity*, is found to be above-average in countries in which firms' collaboration with distant partners is above-average, and countries in which local government support and domestic public procurement are also above-average.

Finally, *ongoing innovation* does not emerge as a relevant factor, uncorrelated to all other aspects of the NIS.

In sum, the main aspects of the innovation system related to innovation outputs are, in order: (i) local government support and domestic public procurement (radical innovation, incremental innovation and productivity); (ii) interactions with close firms (radical innovations); (iii) interactions with distant firms (productivity); (iv) interaction with customers and competitors (radical innovations); and (v) national government support (radical innovations).

Interestingly, a limited number of innovation inputs are correlated to innovation outputs. The factor related to a firm's ownership structure is correlated with all outputs except for ongoing innovation activities. This factor is in turn correlated with a higher propensity to invest in intramural and extramural R&D. A higher frequency in accessing markets, instead, is correlated with a higher than average firm labour productivity. Overall, in our analysis innovation inputs are scarcely correlated with other aspects of the innovation system, and significantly less than innovation outputs.

It is important at this stage to bear in mind that the articulated, descriptive mapping of the NIS dimensions and their correlations are derived from (micro-aggregated) firm-level variables. As such, the emerging results represent a snapshot of how country-specific firms perceive their national, local and international environment in which they operate and devise their strategy. In the best of NIS traditions, the discussion above challenges to a great extent the traditional linear model of R&D. The fact that radical innovation, patenting and innovation-led productivity are more strongly correlated to public support in any form and cooperation with distant and close partners, than the simple amount of resources devoted internally to R&D, is emblematic.

6 Varieties of European Innovation Systems?

We now turn to gather all the pieces of evidence into a coherent whole, with the aim of unravel the presence of regular patterns and different varieties of European NIS. On the basis of this, we will then attempt a brief discussion on what are the directions that innovation and industrial policy should take to ensure catching up of peripheral macro-regions in Europe and overcome the 'European Paradox'.

We perform a cluster analysis of countries with respect to their ranking in each of the factors. From the 26 countries we obtained seven clusters (see Table 8 below). Three clusters with high-ranking countries in one or more of the factors, which we label, in order, `FrontierSmall`, `NorthSmall`, and `G7+IE`; an intermediate cluster, of countries in the middle of the ranking distribution with respect to innovation inputs and outputs, which we label `LargeMed+CZ`; and three clusters with low-ranking countries on most factors, which we label, in order, `SmallMed+LT`, `CE+EE` and `CEE`.

We examine these innovation systems in turn, and briefly discuss their properties.

The Top-notch NIS. The `FrontierSmall` cluster is composed of countries that are ranked the highest with respect to a large number of factors (and rank above average in the remaining factors): Austria (AT), Belgium (BE), Finland (FI), and Norway (NO). These are all relatively small countries (two of which are Nordic countries), with the highest ranking in terms of Government support and public private collaboration, all private collaborations (except those with outer-EU in the US, India, and China), and relevant output indicators such as radical product innovations and labour productivity.

All countries are quite homogeneous, expect for a few exceptions. For instance, Belgian firms are less likely to engage in cooperation with customers and competitors compared with the other three countries; Austrian and Belgian firms receive more funding from local Governments and less from the central Government with respect to Norwegian firms, which are also less likely to collaborate with public bodies than the other three countries; with respect to innovation outputs, Norwegian firms perform better with respect to labour productivity, whereas firms from the other three countries are more likely to produce a radical innovation. Overall this cluster of (relatively small) countries shows a typical NIS virtuous pathway: high public support which is complemented by high public-private links, mainly local and national and associated high innovation and economic performance of firms.

The Demand-pulled NIS. The `G7+IE` cluster is complementary to `FrontierSmall`. It is formed of two of the largest, core EU countries – Germany (DE) and France (FR) – plus Ireland (IE). These countries are particularly strong in terms of innovation inputs, though interestingly especially demand and firm organisation, but also process innovation and incremental innovations. Their ranking position is above-average in all other factors, except for collaborations with customers/competitors and foreign procurement.

Although similar in many respects, these countries are less homogeneous than the previous group: as already noted German firms rely less than average on collaborations with other firms (this is not the case for France and Ireland) and compensate with a larger collaboration with universities and public research centres; Irish firms also benefit more from public support (central and local), have a higher demand in all markets, and invest less in R&D than German and French firms. This evidence suggests that these countries have a model of NIS pulled by demand (foreign and national), which is associated to high public (national) support to innovation in firms and direct public sector procurement.

The linear R&D-based NIS. The `NorthSmall` cluster is composed of three small countries in Northern Europe – Denmark (DK), the Netherlands (NL), and Sweden (SE) – where firms tend to invest more than in other clusters in R&D and knowledge acquisition, engage more than others in cooperation with non-EU firms, show persistence in their innovation activities, and rank above-average in all other factors. The weakest point of this group of countries are the local government support and domestic public procurement (particularly low in the Netherlands) and process and incremental innovation.

Although the group is relatively homogeneous, the main driver in terms of R&D investment and ongoing innovation is Sweden. Other differences are in public support, with Dutch firms receiving more support from the central government and Danish and Swedish firms receiving more support via local governments and domestic procurement. Danish firms also innovate less with respect to Dutch and Swedish firms, especially in new products. This NIS cluster is the closest to a 'linear' R&D model, where private investment in research ensures persistent innovation, and is coupled with systematic public (national) support. Being driven by R&D efforts of firms, innovation (global) cooperation also seems to be important, small NIS are 'outward-looking' in terms of private cooperation.

Even though these three innovation systems are rather different, they are quite successful, for different reasons. In the `FrontierSmall` cluster firms are at the innovation frontier, and are backed by strong government support and relations with other firms. In the `G7+IE` cluster innovation system firms' are probably more heterogeneous, but on average enjoy large and sustained demand, and innovate more than average, particularly in relation to processes. However, they receive substantially less support from the government (although there are consistent differences between countries). Finally, in the `NorthSmall` cluster firms invest considerably in R&D, are well connected, and are relatively well supported by the government, although here as well there are differences in innovation policies, ranging from central to local government support and public procurement.

The Coping NIS. The `LargeMed+CZ` cluster is formed by large Southern countries where firms receive fairly small support from the government (in any form), collaborate less with other firms than in most other countries (close or distant), but are still well above-average with respect to investment in R&D, radical innovation and productivity – although less performing than firms in the first three groups. These are Italy (IT), Spain (ES), and the Czech Republic (CZ).

Although relatively homogeneous, there a number of differences in this group of countries, particularly with respect to the Czech Republic. First, Czech firms relay substantially more on government support and procurement and they collaborate substantially more than firms in Spain and Italy; second, Italian firms are those that invest the least in R&D, especially external R&D; finally, despite other similarities, Spanish firms are well below-average with respect to innovation output indicators, whereas Czech and Italian firms are above-average.

Even though it is less the case for Spain, this is a quite outstanding cluster, from the point of view of coping with a relatively poor public support and opportunities for innovation cooperation (private and public), and yet where firms (especially in Italy) manage to get close to the level of top EU performers.

The 'Spoiled' under-performing NIS. The `SmallMed+LT` is almost symmetric with respect to the `LargeMed+CZ` group. It is formed of small Mediterranean and Baltic countries, where firms receive above-average support from the government, and collaborate above-average with other firms, but rank quite low in relation to innovation output: Croatia (HR), Cyprus (CY), Greece (EL), Lithuania (LT), Portugal (PT), and Slovenia (SI). In the case of Slovenia, for instance, it seems that the cumulative effect of public funding of firm R%D is slowly positively affecting firm incentives to invest (Jakli et al., 2013).

The difference between large (`LargeMed+CZ`) and small (`SmallMed+LT`) Mediterranean innovation systems is quite interesting and raises a number of questions. Why do firms that, on average, receive more government support, are more connected, even than the average firm in `G7+IE` group, lag behind the other clusters of countries (part from Slovenia, with respect to innovation outputs)? By far the weakest point seems to be firm investment in R&D, which despite all the support and linkages is, on average, well below even that of the last two groups of countries (`CE+EE` and `CEE`). Whether this is due to crowding out, or aspects of the innovation systems that we do not capture here, is an interesting question that is left for further research at the micro level and country-specific.

The Embryonic NIS. The last two groups, `CE+EE` and in particular `CEE`, include countries that rank below-average with respect to all firm indicators: Estonia (EE), Hungary (HU) and Slovakia (SK) in the `CE+EE` group, and Bulgaria (BG), Lithuania (LT), Poland (PL), and Romania (RO) in `CEE`. The main exception are the private investment in R&D of firms in Estonia and the relevance of foreign procurement in Estonia and Slovakia, which, however, do not seem to have had the time to have an impact on the system so far.

Overall, this articulated map of European NIS contributes to unpack further the two main clusters of 'Leading Elite' and 'Catching-up' countries empirically derived in Cirillo et al. (2016b), by relying on a larger number of factors/dimensions of NIS and a cluster analysis based on country rankings of several factors' scores. Emerging results have different, interesting nuances that we have represented by the NIS clusters' labelling.

Albeit providing a static picture here, the main take-home message is very much in line with the original NIS approach, i.e. private R&D expenditures by firms are certainly not enough to make countries successful. Interestingly, public support alone, either in the form of public procurement or support in terms of tax credits, is not enough either. The challenge is to identify what sorts of initial conditions and public-private interactions and demand conditions need to be steered together to get to a Top-Notch class. We reprise some of these issues in the concluding section.

Table 8: Country clustering according to factor rankings

Country clustering according to factor rankings

		Cluster means by factor												
		iMarket	iFirmStr	iRD	cEURNAT	cUSCNIN	cCCC	gGvtFCo	gLRFDoPr	gForPr	oRadPat	oIncrPcs	oOng	oPtvty
FrontierSmall	7	5.5	5.5	8.0	2.5	5.3	4.5	2.5	3.5	3.0	3.0	5.5	8.8	5.5
NorthSmall	2	12.3	7.7	4.0	6.7	3.7	5.7	8.3	13.7	8.0	8.0	14.7	6.0	5.7
G7+IE	3	4.0	5.0	8.0	10.7	9.3	17.7	8.0	4.7	16.7	8.0	5.0	11.0	6.3
LargeMed+CZ	1	13.7	15.7	10.7	18.7	20.3	19.7	15.0	17.0	21.3	14.7	17.0	6.7	13.7
SmallMed+LT	4	13.0	14.0	21.7	14.5	15.3	10.5	15.3	12.7	11.5	15.0	9.8	21.2	18.3
CE+EE	5	21.7	20.7	13.3	16.0	14.7	14.3	19.3	19.0	13.3	20.7	23.3	13.0	17.7
CEE	6	24.0	24.5	20.3	24.5	23.5	24.5	24.3	24.5	23.0	23.8	22.5	19.8	22.3

Frontier Small	AT	7
	BE	7
	FI	7
	NO	7

NorthS mall	DK	2
	NL	2
	SE	2

G7+IE	DE	3
	FR	3
	IE	3

LargeMed +CZ	CZ	1
	ES	1
	IT	1

SmallMed +LT	CY	4
	EL	4
	HR	4
	LT	4
	PT	4
	SI	4

CE+EE	EE	5
	HU	5
	SK	5

CEE	BG	6
	LV	6
	PL	6
	RO	6

7 Final Remarks

This paper has built upon the literatures on National Innovation System(s) (NIS) and complemented the extant, few, empirically grounded attempts to quantify the complexity of dimensions that characterise NIS. Namely, it has added to the exercises in Shrolec and Verspagen (2008) and Cirillo et al. (2016b), by providing a novel, micro-level grounded mapping of European National Innovation Systems.

We have sacrificed the longitudinal perspective, that others in the literature have focused on (see, among others, Castellacci and Natera, 2013), to offer a fine-grained, firm-level grounded picture of the composite dimensions of NIS. We have highlighted in the introduction that we have chosen to privilege the emphasis that the NIS approach has traditionally put on firms behaviour and performance, and looked at the complex network of actors, which firms interact with and respond to. We have taken into account the structure, innovation strategy and performance of the firms and related them to several institutional characteristics of the NIS, such as the nature of public sector support (e.g. cooperation and procurement) and the characteristics of the public-private links (e.g. with universities, foreign institutions and/or other firms).

Our main aim has been, essentially, one of empirically systematising and mapping not only the complex set of dimensions of NIS but also their complementarity and substitutability, as well as how European countries score in terms of the different (latent) dimensions. We have been able to unravel and obtain characteristics of the NIS that position European countries in several clusters, some of which have the features of Top Notch technological clubs, others of typical latecomers (the Embryonic NIS). However, we have also identified national and cluster stories that are at odds with the traditional NIS prescriptions, as the examples of countries that perform well despite (public) adversity (the Coping NIS) and others that perform badly despite a substantial public support (the Spoiled and under-performing NIS).

The nuances of the European NIS are therefore several, as expected from a high heterogeneity of country size, industrial structures, firm behaviour, outward strategies, and nature of public support. The common denominator is that NIS pathways are never attributable to a single dimension, say investment in R&D in firms or public procurement. Not only there is not a 'one size fit all' recipe, but the recipe needs to be tailored to the complexity of conditions and it is not usually only pumping public funds into R&D, which is obviously a necessary but not sufficient condition.

The challenges for industrial and innovation policies are therefore at several levels: one of speeding up the process of moving away from the initial conditions; one of timing of public intervention with respect to the extent that absorptive capacity of firms allows to benefit from it; one of identification of the technological opportunities that best fit the industrial structure; one of appropriate steering of technological upgrading and structural change; one of ensuring macro-level policies that create favourable demand conditions. After all, one of the most virtuous examples of NIS are the demand-pulled ones, where a sustained domestic and international demand require persistent innovation efforts by firms and public organisations.

This paper also opens more questions than it does offer in terms of answers. By mapping the crucial relations between innovation inputs, cooperation links, public support and outputs, we have found a number of non-trivial patterns in European countries' innovation systems. However, because we derive these by merging results from average firms within the innovation survey, we are not able to explain many of these regularities. On the contrary, we leave a number of open questions that we hope inspire much needed future research to provide a theoretical grounding to the role of different dimensions of innovation systems.

We summarise some of these questions below.

Are firm collaborations more important for innovation inputs or output?

To what extent can we safely argue that (at the micro level) private-private collaborations are complementary to public-private collaborations and support? More research is required to establish the causal links: does government support provide firms with the means to collaborate? Or do firms that invest in collaborations with private firms, also do so with governments and public research centres?

In the few cases in which they seem to be substitutes (Germany), is it because the public support and procurement satisfies the access to external knowledge of firms?

Is additionality a cross-country phenomenon? We need micro level comparative/ble studies that analyse the role of public support to firm innovation (inputs) across different innovation systems.[34]

What is missing in the Coping NIS (the small Mediterranean innovation

[34] Griffith et al. (2006) and Freitas et al. (2017) are two relevant examples: the first focus on innovation inputs show limited differences between France, Germany, Spain and the UK; the second, instead, shows that firm innovation behaviour differ significantly across Norway, Italy and France, especially when considering both innovation inputs and outputs in relation to sectoral differences.

systems `SmallMed+LT`) which feature strong public support and linkages, but relatively low innovation output and quite small levels of private R&D?

What would be the effect of increased support in countries in the Spoiled and under-performing (`LargeMed+CZ`) NIS, where firms manage to innovate close to the frontier, but the support is extremely low?

We trust the present contribution sparks the need to investigate further some of these questions, and helps informing policies that are able to go beyond a simplistic Lisbon Strategy '3% narrative' and devise public intervention that supports innovation and upgrading for the European periphery to catch up.

Acknowledgements

T. Ciarli, M. Savona and A. L. Wirkierman gratefully acknowledge the support by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 649186 — ISIGrowth.

References

Adams, J. (2012). Collaborations: The rise of research networks. *Nature*, 490(7420):335–336.

Aerts, K. and Czarnitzki, D. (2004). Using innovation survey data to evaluate r&d policy: The case of belgium. ZEW Discussion Paper 04-055, ZEW - Centre for European Economic Research, Mannheim.

Aerts, K. and Schmidt, T. (2008). Two for the price of one? *Research Policy*, 37(5):806 – 822.

Baldwin, C. and von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization Science*, 22(6):1399–1417.

Berchicci, L. (2013). Towards an open r&d system: Internal r&d investment, external knowledge acquisition and innovative performance. *Research Policy*, 42(1):117 – 127.

Bloch, C. and Graversen, E. K. (2012). Additionality of public r&d funding for business r&d a dynamic panel data analysis. *World Review of Science, Technology and Sustainable Development*, 9(2-4):204–220. PMID: 47688.

Caloghirou, Y., Kastelli, I., and Tsakanikas, A. (2004). Internal capabilities and external knowledge sources: complements or substitutes for innovative performance? *Technovation*, 24(1):29 – 39.

Castellacci, F. (2008). Technology clubs, technology gaps and growth trajectories. *Structural Change and Economic Dynamics*, 19(4):301 – 314.

Castellacci, F. and Archibugi, D. (2008). The technology clubs: The distribution of knowledge across nations. *Research Policy*, 37(10):1659 – 1673. Special Section Knowledge Dynamics out of Balance: Knowledge Biased, Skewed and Unmatched.

Castellacci, F. and Natera, J. M. (2013). The dynamics of national innovation systems: A panel cointegration analysis of the coevolution between innovative capability and absorptive capacity. *Research Policy*, 42(3):579 – 594.

Cirillo, V., Martinelli, A., Nuvolari, A., and Tranchero, M. (2016a). How it all began: The long term evolution of scientific and technological performance. Technical report, Institute of Economics, Scuola Superiore SantAnna, Pisa, Italy.

Cirillo, V., Martinelli, A., Nuvolari, A., and Tranchero, M. (2016b). Only one way to skin a cat? national innovation systems in the xxi century. Technical report, Institute of Economics, Scuola Superiore SantAnna, Pisa, Italy.

Clausen, T. H. (2009). Do subsidies have positive impacts on r&d and innovation activities at the firm level? *Structural Change and Economic Dynamics*, 20(4):239 – 253.

Czarnitzki, D. and Licht, G. (2006). Additionality of public r&d grants in a transition economy. *Economics of Transition*, 14(1):101–131.

Dosi, G., Llerena, P., and Labini, M. S. (2006). The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called european paradox. *Research Policy*, 35(10):1450 – 1464. Triple helix Indicators of Knowledge-Based Innovation Systems.

Everitt, B. and Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*. Springer.

Fagerberg, J. and Srholec, M. (2008). National innovation systems, capabilities and economic development. *Research Policy*, 37:14171435.

Falk, R. (2004). Behavioural additionality effects of r&d subsidies: empirical evidence from austria. Tip working paper, TIP, Vienna.

Freeman, C. (1987). *Technology Policy and Economic Performance: Lessons from Japan.* Pinter, London.

Freitas, I. B., Castellacci, F., Fontana, R., Malerba, F., and Vezzulli, A. (2017). Sectors and the additionality effects of r&d tax credits: A cross-country microeconometric analysis. *Research Policy*, 46(1):57 – 72.

Frenz, M. and Lambert, R. (2012). Mixed modes of innovation: an empiric approach to capturing firms' innovation behaviour. *OECD Working Paper*, DSTI/DOC(2012)6.

Frenz, M. and Prevezer, M. (2012). What Can CIS Data Tell Us about Technological Regimes and Persistence of Innovation? *Industry and Innovation*, 19(4):285–306.

Griffith, R., Huergo, E., Mairesse, J., and Peters, B. (2006). Innovation and productivity across four european countries. *Oxford Review of Economic Policy*, 22(4):483–498.

Hammadou, H., Paty, S., and Savona, M. (2014). Strategic interactions in public r&d across european countries: A spatial econometric analysis. *Research Policy*, 43(7):1217 – 1226.

Hartigan, J. A. and Wong, M. A. (1979). A K-means clustering algorithm. *Applied Statistics*, 28:100–108.

Jakli, A., Burger, A., and Rojec, M. (2013). The quest for more efficient r&d subsidies. *Eastern European Economics*, 51(4):5–25.

Lokshin, B. and Mohnen, P. (2012). How effective are level-based r&d tax credits? evidence from the netherlands. *Applied Economics*, 44(12):1527–1538.

Lööf, H. and Hesmati, A. (2004). The impact of public funding on private r&d investment. new evidence from a firm level innovation study (additionality or crowding out? on the effectiveness of r&d subsidies). (06):26. QC 20120208.

Love, J. H. and Roper, S. (2001). Location and network effects on innovation success: evidence for uk, german and irish manufacturing plants. *Research Policy*, 30(4):643 – 661.

Lundvall, B. A., editor (1992). *National Systems of Innovation: Toward a Theory of Innovation and Interactive Learning*. Pinter, London.

Lundvall, B. A. (2007). National Innovation Systems — Analytical Concept and Development Tool. *Industry and Innovation*, 14(1):95–119.

Nelson, R., editor (1993). *National Innovation Systems: A comparative analysis*. Oxford University Press, Oxford.

Nelson, R. R. (2006). Reflections on the simple economics of basic scientific research: looking back and looking forward. *Industrial and Corporate Change*, 15(6):903–917.

Nieto, M. J. and Santamara, L. (2007). The importance of diverse collaborative networks for the novelty of product innovation. *Technovation*, 27(6):367 – 377.

OECD (2002). *Dynamising National Innovation Systems*. OECD Publishing.

OECD (2005). *Innovation Policy and Performance*. OECD Publishing.

OECD (2009). *Innovation in Firms*. OECD Publishing.

OECD and EUROSTAT (2005). *Oslo Manual: Guidelines for collecting and interpreting innovation data, 3Ed*. OECD Publishing.

Powell, W. W., Koput, K. W., and Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41(1):116–145.

Rafols, I., Ciarli, T., van Zwanenberg, P., and Stirling, A. (2012a). Towards indicators for 'opening up' science and technology policy. In *Science and Technology Indicators Conference*, Montreal.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., and Stirling, A. (2012b). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy*, 41(7):1262 – 1282.

Raykov, T. and Marcoulides, G. A. (2008). *An Introduction to Applied Multivariate Analysis*. Routledge.

Rencher, A. C. and Christensen, W. F. (2012). *Methods of Multivariate Analysis, 3rd. Edition*. Wiley.

Shrolec, M. and Verspagen, B. (2008). The Voyage of the Beagle in Innovation Systems Land. Explorations on Sectors, Innovation, Heterogeneity and Selection. *UNU-MERIT, Working Paper Series*, 2008-008.

Soete, L., Verspagen, B., and Weel, B. T. (2010). *Systems of innovation in Handbook of the Economics of Innovation, ed. by*, page 11591180. Elsevier.

Timm, N. H. (2002). *Applied Multivariate Analysis*. Springer.

van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1):133–143.

Ziga-Vicente, J. ., Alonso-Borrego, C., Forcadell, F. J., and Galn, J. I. (2014). Assessing the effect of public subsidies on firm r&d investment: A survey. *Journal of Economic Surveys*, 28(1):36–67.

dne Cappelen, Raknerud, A., and Rybalka, M. (2012). The effects of r&d tax credits on patenting and innovations. *Research Policy*, 41(2):334 – 345.

Appendixes

A Consolitadtion of dataset and imputation of missing values

Eurostat's publicly available micro-aggregated CIS 2014 database is presented as a series of data files covering different aspects of the CIS questionnaire. In particular we considered the following files:

Table 1: Eurostat CIS-2014 files

File	Label	Description
1	`bas`	Basic economic information on the enterprises
2	`gen`	General information on the enterprises
3	`type`	Enterprises by main types of innovation
4	`spec`	Enterprises by specific types of innovation
5	`prod`	Product and process innovative enterprises
6	`exp`	Innovation activities and expenditures in the enterprises
7	`pub`	Public funding in the enterprises
8	`coop`	Types of co-operation of the enterprises
9	`proc`	Public sector procurement and innovation in the enterprises
10	`ipr`	Intellectual property rights and licensing in the enterprises

As reported in panel (A) of Table 9, 21 out of the 33 variables considered had missing values for, at least, one of the 26 countries included in the analysis. Thus, an estimation procedure to obtain within-sample predictions for the missing values had to be devised.

We proceeded as follows. First, we identified the subset of variables $\{X_j\}_{j=1}^{q_F}$ for which all countries have full data coverage (i.e. panel (B) of Table 9). Observations for those q_F variables are available for the C_F countries with full data as well as for the C_M countries with some missing data.

Second, we considered each one of the variables in panel (A) of Table 9 at a time. Assume we label it Z. Observations for variable Z are available *only* for the set of C_F countries.

Third, we identified amongst $\{X_j\}_{j=1}^{q_F}$ those variables j for which:

$$\rho(Z, X_j) > 0.95 \times \max\{\rho(Z, X_r), r = 1, \dots, q_F\} \tag{6}$$

i.e. the correlation with variable Z was higher than 95% of the maximum correlation value.

Table 9: Dataset variables: imputation of missing values for selected countries

Panel (A) EUROSTAT CIS-2014 variables with missing values for some countries, for which it was necessary to provide an estimate

#Variable	#File	File	Firm Type Code	Firm Type Description	Indicator Code	Indicator Description	Imputation of missing values
1	10	ipr	INNO	Innovative firms	PROPAT	Enterprises that applied for a patent	BE, DK, FI, IE, NL, SI
2	10	ipr	INNO	Innovative firms	PROTM	Enterprises that registered a trademark	BE, CZ, DK, FI, IE, NL, PL, SI
3	2	gen	INNO_PPANPP	Innovative firms all dimensions	ENMRG_YES	Enterprises that have merged with/take over another enterprise	BE, DK, FI, NL
4	2	gen	INNO_PPANPP	Innovative firms all dimensions	GP_YES	Enterprises that are part of an enterprise group	DK, NL
5	2	gen	INNO_PPANPP	Innovative firms all dimensions	LARMAR_EU	Largest market in terms of turnover is: EU/EFTA/EU-candidates	BE, DK, ES, FI
6	2	gen	INNO_PPANPP	Innovative firms all dimensions	LARMAR_LREG	Largest market in terms of turnover is the local/regional market	BE, DK, ES, FI
7	2	gen	INNO_PPANPP	Innovative firms all dimensions	LARMAR_NAT	Largest market in terms of turnover is the national market	BE, DK, ES, FI
8	8	coop	INNOACT	Product/Process innovative firms	C031	Enterprises co-operating with clients or customers from the private sector	IE
9	8	coop	INNOACT	Product/Process innovative firms	C032	Enterprises co-operating with clients or customers from the public sector	IE, NL
10	8	coop	INNOACT	Product/Process innovative firms	C09	Enterprises co-operating with Government, public or private research institutes	SE
11	7	pub	INNOACT	Product/Process innovative firms	FUNGMT	Enterprises that received funding from central government	AT, DK, IE
12	7	pub	INNOACT	Product/Process innovative firms	FUNLOC	Enterprises that received funding from local or regional authorities	AT, DK, IE, SE
13	6	exp	INNOACT	Product/Process innovative firms	ROEK14_PC	Share of expenditures in acquisition of external knowledge in 2014	BG
14	6	exp	INNOACT	Product/Process innovative firms	RRDEX14_PC	Share of expenditures in external R&D in 2014 over total innovation expenditures	BG, NL
15	6	exp	INNOACT	Product/Process innovative firms	RRDIN14_PC	Share of expenditures in in-house R&D in 2014 over total innovation expenditures	NL
16	3	type	INONG	Firms with ongoing innovation	ENT_POPU14	Enterprises with on-going innovation activities only	SE
17	5	prod	INPCS	Process innovative firms	INPSNM0	Enterprises that have introduced process innovation not new to the market	AT, BE, CZ, DE, DK, ES, FI, IE
18	5	prod	INPCS	Process innovative firms	INPSNM1	Enterprises that have introduced process innovation new to the market	AT, BE, CZ, DE, DK, ES, FI, IE
19	9	proc	TOTAL	Total firms	PUBDOM	Enterprises with procurement contract for domestic public sector	BE, DE, DK, ES, FR, IE
20	9	proc	TOTAL	Total firms	PUBFINRQ	Procurement contract for foreign public sector/innovation activities required	BE, DE, DK, ES, FR, IE, LT, SE
21	9	proc	TOTAL	Total firms	PUBFOR	Enterprises with procurement contract for foreign public sector	BE, DE, DK, ES, FR, IE

Panel (B) EUROSTAT CIS-2014 variables with values for all countries considered

#Variable	#File	File	Firm Type Code	Firm Type Description	Indicator Code	Indicator Description
1	1	bas	INNO	Innovative firms	TURN_EMP	Total turnover in 2014 per employee
2	8	coop	INNOACT	Product/Process innovative firms	C01	Enterprises co-operating with other enterprises within the enterprise group
3	8	coop	INNOACT	Product/Process innovative firms	C02	Enterprises co-operating with competitors or other enterprises of the same sector
4	8	coop	INNOACT	Product/Process innovative firms	C06	Enterprises co-operating with universities or other higher education institutions
5	8	coop	INNOACT	Product/Process innovative firms	COCNIN_YES	Engaged in any type of innovation co-operation with a partner in China or India
6	8	coop	INNOACT	Product/Process innovative firms	COEUR_YES	Engaged in innovation co-operation with a partner in EU/EFTA/EU-candidates
7	8	coop	INNOACT	Product/Process innovative firms	CONAT_YES	Engaged in any type of innovation co-operation with a national partner
8	8	coop	INNOACT	Product/Process innovative firms	COUS_YES	Engaged in any type of innovation co-operation with a partner in United States
9	6	exp	INNOACT	Product/Process innovative firms	EXPTOT14_C	Share of total innovation expenditures in 2014 in Manufacturing
10	6	exp	INNOACT	Product/Process innovative firms	EXPTOT14_ENT_POPU14	Average total innovation expenditures in 2014 per firm
11	5	prod	INPDT	Product innovative firms	NEWFRM_YES	Introduced new or significantly improved products only new to the firm
12	5	prod	INPDT	Product innovative firms	NEWMAR_YES	Introduced new or significantly improved products new to the market

References:
All variables refer to the EUROSTAT size class: TOTAL; and NACE Rev. 2 group of industries: B-M73_INN
Source: Own elaboration based on EUROSTAT CIS 2014 Database

In this way, we aimed at singling out variables with observations available for *all* countries that are strongly (and positively) correlated with variable Z, available to only *some* of the countries. We denote by q_r the number of variables that satisfy condition (6).

In the fourth place, we postulated the following linear probability model for the C_F countries with full data availability:

$$Z_i = \beta_0 + \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i, \quad i = 1, \ldots, C_F \tag{7}$$

where $\boldsymbol{x}_i$ is an $q_r \times 1$ vector of observations for the q_r variables in country i. We then estimated (7) using ordinary least squares (OLS, hereinafter), obtaining point estimates $(\hat{\beta}_0, \hat{\boldsymbol{\beta}})$.

Then, we predicted Z_i on the basis of $(\hat{\beta}_0, \hat{\boldsymbol{\beta}})$ and $\boldsymbol{x}_i$ for those C_M countries with *missing* values of Z:

$$\hat{Z}_i = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i, \quad i = 1, \ldots, C_M \tag{8}$$

imputing $\hat{Z}_i$ as an estimate of Z_i for the exploratory factor analysis.

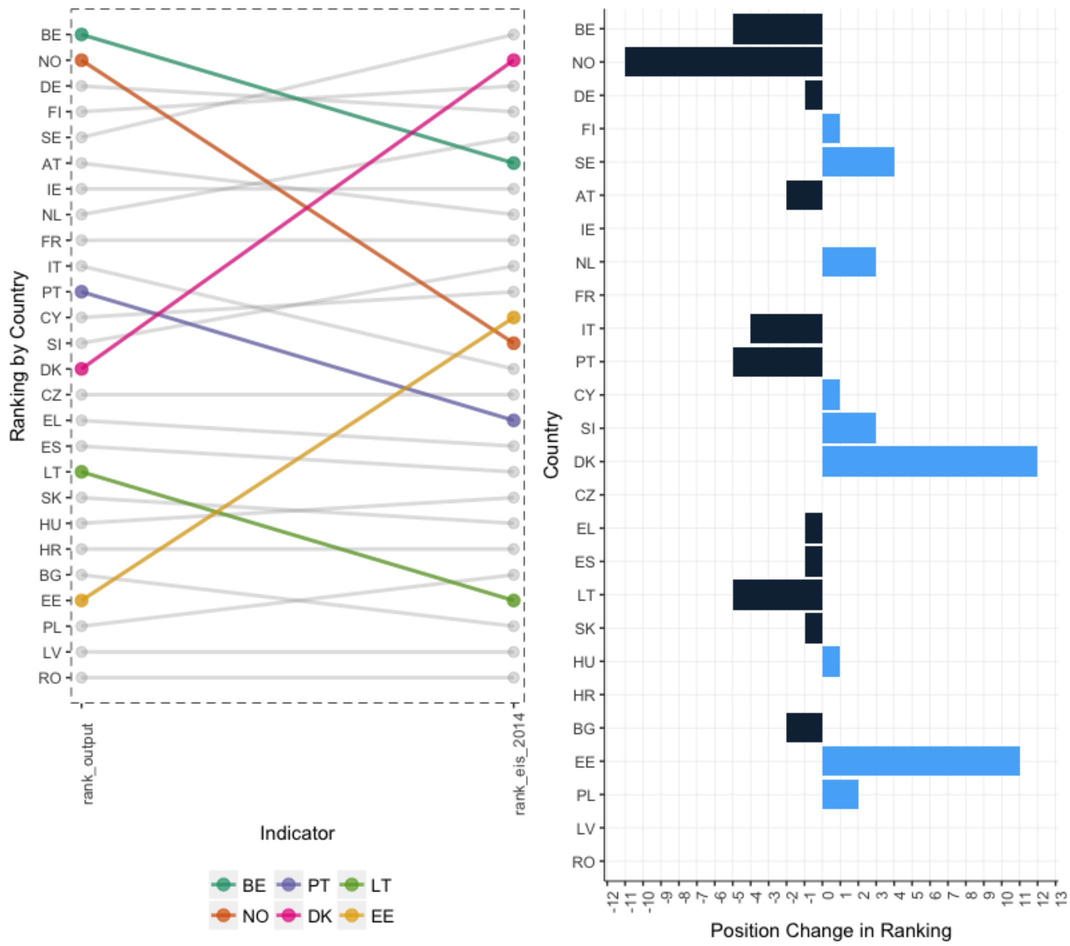
B Comparison with the European Innovation Scoreboard (EIS 2014)

Indicators have pros and cons and can lead to quite different outcomes in assessing the innovation performance of countries, organisations or individuals (e.g. Rafols et al., 2012b,a; van Raan, 2005). In the paper we discuss a number of efforts undertaken by innovation scholars to study countries' innovation performance. All these indicators bring together several measures which may or may not be related to innovation, such as education, economic performance, entrepreneurship, etc.

Instead, the indicator that we propose is based exclusively on firm information: firm experience in terms of innovation input, outputs, as well as support and linkages. The proposed indicator should allow to better capture *innovation*, because it does not refer, for instance, to a general figure of public investment, but to which extent that investment is used by firms.

In the analysis of EU countries' innovative performance, one of the most diffused indicators is the European Innovation Scoreboard (EIS, hereinafter). In order to assess the firm-based indicator of innovative output proposed in this paper we compare the country-ranking between the two indicators, for those countries available in both of them. Figure 10 plots the relative ranking of each country (left panel) for our indicator of innovation output (left column) and the EIS for 2014 (right column), as well as the difference between the two (right panel).

Figure 10: Ranking differences between Innovation Outputs and EIS-2014



For most countries the ranking is remarkably similar. There are a few notable exceptions: we rank Belgium (BE), Norway (NO), Portugal (PT) and Lithuania (LT) higher than in the EIS, whereas we rank Denmark (DK) and Estonia (EE) lower than in the EIS. These differences boil down to the information considered. For instance, we do not control for educational attainment, research, and economic outputs. As discussed, we focus only on innovation, and in particular on firm innovative experience. These missing variables are likely to be the main cause of the differences between the two indicators. However, our interest in this paper is in the firm level perspective on innovation, which is better captured by our proposed synthetic measure of innovation outputs.

C Additional Table: average indicator values by cluster

Table 11: Average indicator values by cluster

Average cluster values of variables from the Community Innovation Survey 2014 Ed. (CIS-2014) used in the paper:

variable	Factor Analysis	Factor	Variable	Description	unit	mean	FrontierSmall 7	NorthSmall 2	G7+IE 3	LargeMed+CZ 1	SmallMed+LT 4	CE+EE 5	CEE 6
1	Innovation Inputs	iMarket	input_INNO_PPANPP_LARMAR_EU	Largest market: EU	%	5.08	7.08	4.40	12.14	4.85	4.27	2.79	1.38
2			input_INNO_PPANPP_LARMAR_LREG	Largest market: Local/Regional	%	6.68	9.53	6.34	17.93	5.26	6.12	1.42	1.49
3			input_INNO_PPANPP_LARMAR_NAT	Largest market: National	%	10.08	12.18	13.28	21.79	8.79	8.78	5.54	3.13
4		iFirmStr	input_INNO_PPANPP_ENMRG_YES	Firm merged/took over	%	2.08	3.79	3.05	3.66	1.45	1.84	0.70	0.35
5			input_INNO_PPANPP_GP_YES	Firm part of enterprise group	%	10.08	19.72	14.98	13.98	8.04	7.73	5.41	2.42
6		iRD	input_INNOACT_EXPTOT14_ENT_POPU14	R&D expenditure per firm	EUR/NR	90747.24	133075.30	175872.96	164663.34	64896.03	35712.41	66580.43	49203.55
7			input_INNOACT_RRDEX14_PC	Share of external R&D	% of TIE	10.50	13.46	17.27	14.12	14.63	5.00	11.00	4.49
8			input_INNOACT_RRDIN14_PC	Share of in-house R&D	% of TIE	38.98	59.41	54.83	51.08	38.46	26.75	28.41	24.27
9			input_INNOACT_EXPTOT14_C	Manufacturing/Aggregate R&D	% of TIE	54.64	55.80	63.04	56.42	63.89	47.36	54.80	49.71
10			input_INNOACT_ROEK14_PC	Acquisition of external knowledge	% of TIE	4.19	2.91	6.03	3.06	4.80	3.44	7.28	3.33
11	Cooperation Links	cEURNAT	links_INNOACT_C01	Cooperation within the enterprise group	%	5.02	10.28	7.40	6.32	2.90	3.69	4.10	1.26
12			links_INNOACT_COEUR_YES	Cooperation with EU partners	%	6.46	13.04	8.27	5.75	3.21	6.42	6.14	1.76
13			links_INNOACT_CONAT_YES	Cooperation with National partners	%	11.04	20.51	14.70	11.58	7.98	11.18	7.97	2.82
14		cUSCNIN	links_INNOACT_COCNIN_YES	Cooperation with China/India	%	1.08	1.88	2.24	1.12	0.48	0.87	0.93	0.24
15			links_INNOACT_COUS_YES	Cooperation with the US	%	1.68	3.23	3.42	2.27	0.77	1.15	1.13	0.29
16		cCCC	links_INNOACT_C02	Cooperation with competitors, same sector	%	3.02	6.23	4.56	2.46	1.58	3.14	2.02	0.70
17			links_INNOACT_C031	Cooperation with private clients/customers	%	5.28	10.88	8.80	4.77	2.62	4.72	4.00	1.24
18	Government	gGvtFCo	gvt_INNOACT_FUNGMT	Funding from Central Government	%	6.47	11.72	9.66	8.24	5.24	6.35	2.61	1.50
19			gvt_INNOACT_C032	Coop. with public sector clients/customers	%	2.18	4.76	3.68	2.35	0.83	1.83	1.45	0.44
20			gvt_INNOACT_C06	Cooperation with universities/HEI	%	4.58	9.67	6.50	5.92	3.17	3.60	2.62	1.03
21			gvt_INNOACT_C09	Cooperation with Gvt/Research Inst.	%	2.96	6.93	3.40	3.84	2.21	2.31	1.34	0.74
22		gLRFDoPr	gvt_INNOACT_FUNLOC	Funding from Local/Regional Auth.	%	3.28	6.60	4.51	6.50	4.35	1.66	0.53	0.32
23			gvt_TOTAL_PUBDOM	Domestic Procurement	%	22.61	35.34	22.78	34.31	16.51	24.26	16.01	8.00
24		gForPr	gvt_TOTAL_PUBFINRQ	Foreign proc. req. innovation activities	%	0.39	0.93	0.58	0.38	0.10	0.30	0.37	0.06
25			gvt_TOTAL_PUBFOR	Foreign Procurement	%	2.83	5.83	3.72	2.50	1.22	3.05	2.31	0.65
26	Innovation Outputs	oRadPat	output_INNO_PROPAT	Application for a patent	%	3.69	8.26	5.51	6.47	2.10	2.10	1.24	1.10
27			output_INNO_PROTM	Application for a trademark	%	6.53	8.74	8.99	8.92	6.85	6.22	3.30	3.34
28			output_INPDT_NEWMAR_YES	Turnover from porducts new to market	%	12.93	21.55	16.06	17.99	11.59	13.15	5.22	4.63
29		oIncrPcs	output_INPDT_NEWFRM_YES	Turnover from porducts new to firm	%	15.91	23.25	16.12	25.34	14.38	18.73	5.69	5.92
30			output_INPCS_INPSNM0	Process innovation new to firm	%	10.15	14.82	11.74	9.74	8.44	13.17	6.67	3.98
31			output_INPCS_INPSNM1	Process innovation new to market	%	6.06	8.00	6.45	9.72	4.69	7.69	2.74	2.18
32		oOng	output_INONG_ENT_POPU14	Ongoing innovation activities	%	1.86	2.48	2.98	3.04	2.54	0.80	1.60	0.82
33		oPtvty	output_INNO_TURN_EMP	Turnover per employee	EUR/EMP	31054.16	48385.69	43056.31	54966.21	31695.90	19863.30	18821.44	12266.51

References:
% (percentages) are expressed in relation to the total of firms of the corresponding firm type
NR: number; EUR: euros at current prices; EMP: employees; % of TIE: percentage of Total Innovation Expenditure
Clusters' composition: FrontierSmall: AT BE FI NO; NorthSmall: DK NL SE; G7+IE: FR, DE, IE; LargeMed+CE: CZ ES IT; SmallMed+LT: CY EL HR LT PT SI;
CE+EE: EE HU SK, CEE: BG LV PL RO
Source: Own elaboration based on EUROSTAT CIS 2014 Database