WORKING PAPER | ISSUE 07 | 27 NOVEMBER 2020

UNDERSTANDING THE EUROPEAN UNION'S REGIONAL POTENTIAL IN LOW-CARBON TECHNOLOGIES

ENRICO BERGAMINI AND GEORG ZACHMANN

Our research identifies existing and potential specialisation in green technologies in European Union regions, and proposes an approach to identify policies that can help to realise this potential. Using the Organisation for Economic Cooperation and Development's REGPAT database for regionalised patent data, we estimate the potential advantage European NUTS2 regions could have in 14 green technologies. We use network proximity between technologies and between regions to understand technological/regional clusters, and build the regressors for estimating potential regional advantage in specific technologies via zero-inflated beta regressions. We construct a dataset of lagged potentials and labour market, economic and demographic variables, and perform an elastic net regularisation to understand the association with current revealed advantages. Our approach indicates an association between technological advantage in green technologies in EU regions and participation rates in labour markets, sectoral employment in science and technology, general higher education, duration of employment, percentage of GDP spent on research and development (public and private), and other expenditure on R&D. If confirmed by causality tests, the established associations could help in designing horizontal economic policies to enable specific regions to realise their specialisation potential in specific green technologies.

Enrico Bergamini is a PhD candidate at Collegio Carlo Alberto, Turin
Georg Zachmann (georg.zachmann@bruegel.org) is a Senior Fellow at Bruegel

The authors would like to thank Donovan Tokuyama for excellent research assistance, and Robert Kalcik and Catarina Midões for their fundamental inputs and suggestions. We would also like to thank Antoine Mathieu-Collin, Alexander Roth, Claudia Ghisetti and Reinhilde Veugelers for their valuable comments.

Recommended citation:
Bergamini, E. and G. Zachmann (2020) 'Understanding the European Union's regional potential in low-carbon technologies', *Working Paper* 07/2020, Bruegel

bruegel

1 Introduction and literature review

Keeping the global temperature increase below 2°C above pre-industrial levels will require the almost complete decarbonisation of our energy system early in the second half of this century. This will require the increasing and diverse deployment of green technologies (including vehicles, power plants, appliances and batteries), which will replace the existing stock of high-carbon technologies.

Regions with strong climate policies – such as the EU – will become pilot markets for low-carbon technologies. But being a pilot market does not automatically translate into a competitive edge in the new technologies, as the limited success of photovoltaic cell production in Europe shows. To enable domestic companies to flourish in these new sectors, policymakers seek to complement the creation of early markets for decarbonisation technologies with some form of industrial policy.

In this paper, we argue that not all regions have the potential to excel in all green technologies, and that a fundamental characteristic of green industrial policies should be the consideration and inclusion of a tailored regional aspect. Our analysis relies on systematic evidence originating from the complexity-based literature triggered by Hidalgo and Hausmann (2009), and builds on analysis of green technologies patenting and the current advantages of similar regions. Hidalgo and Hausmann (2009) also employed the concept of relatedness between different technologies.

The literature that builds on these theoretical frameworks and empirical approaches has advanced the European efforts to pursue so-called Smart Specialisation Strategies (Foray *et al*, 2011). Based on empirical work on geographically granular data, these analyses identify competitive advantages in terms of regions' knowledge bases, labour markets, geographical characteristics or industrial structures, and provide guidelines for diversification opportunities at regional level. The idea is that regions can build on local characteristics in order to diversify into technologies related to their existing structures (Balland *et al*, 2019). Van den Berge and Weterings (2014) explored the potential for regions to diversify in eco-technologies. They found that in regions in which the knowledge base was previously characterised by the presence of green innovations, the likelihood of developing new technologies is greater. Montresor and Quatraro (2018) also explored the role of relatedness in green technology in a regional context, finding that the relatedness of the existing knowledge base can facilitate the entry of green technologies. They also found that the process of innovation makes green innovation desirable also for regions with stocks of related non-green knowledge.

Our empirical work builds on the idea of related diversification (Boschma and Frenken, 2011): comparative regional advantages in specific green technologies and sectors can be built on top of existing strengths in related technologies and sectors. Accordingly, the first step in our analysis is to identify regional green technology potential based on the relatedness to green technologies of current regional comparative advantages. In order to identify potential advantages, we refer to the framework conceptualised in Hausmann *et al* (2019), building on the regression-based forecasting technique in Zachmann and Roth (2018), to estimate future potential advantage at the regional level, and we focus specifically on green technologies.

The second step in our analysis focuses on identifying policies that can help to realise regional potential in low-carbon technologies. A growing body of empirical literature is studying the policy relevance of Smart Specialisation, and regional innovation. Boschma and Gianelle (2014) argued that an experienced entrepreneurial base, labour mobility across related industries, and inter-regional collaboration are factors for success in the diversification process. Santoalha and Boschma (2020) confirmed that the presence of green-related capabilities in a region, and political support for green development at the regional level, can foster innovation. Steen *et al* (2016) found that lack of political support for green objectives, also at the national level, is a barrier to regional innovation. Crespo *et al* (2017), explained how *"developing new growth paths in related industries or technological domains increases the probability of regional competitive advantage because the shorter cognitive distance enhances mutual learning, knowledge spillovers and actors' redeployment of skills from one domain to another."*

In the second stage of our research we investigate observed regional characteristics, from labour markets to policy and institutional aspects, which might lead regions to create, realise and exploit this potential. We use an innovative methodology and a novel dataset to do so. We essentially ask: which labour market, economic and demographic conditions are associated, together with potential specialisation, with stronger relative technological advantage?

Hence, our analysis is based on a two-stage approach in which we first estimate potential and revealed green specialisation, and subsequently select labour market, demographic and economic variables that are associated with it.

The paper proceeds as follows: section 2 explains the data sources used, technological definitions and regionalisation of patents. In section 3, we build an empirical strategy to estimate regional technological advantage, and use a regression-based technique to estimate potential advantage.

Section 4 explains our methodology for the second stage regressions in which we perform a data-driven selection of variables. In section 5, we discuss limitations and results. Section 6 concludes and discusses possible policy implications.

2 Data

In the field of innovation economics, the most widely-used source of data is patents statistics. Patents are not only an indicator of technological specialisation in innovation activities, but also a proxy for regional economies' sectoral specialisations. The PATSTAT database contains information gathered from statistical offices worldwide, and is commonly analysed as a proxy for innovation activity. The advantage of this data is that it contains very granular information, covering full patents texts, as well as inventor-level data.

However, a well-known issue in the use of PATSTAT for geographical analysis is missing information, especially at sub-national level. We exploit the Organisation for Economic Co-operation and Development's REGPAT database, a plugin for the PATSTAT database with enhanced geocoding, providing information consistently geocoded at the regional level. The number of patents attributed to a region is based on the location of patent inventors that applied to the European Patent Office or for international patents under the Patent Cooperation Treaty (PCT). The earliest application of individual patent families is used and attributed in fractions to all inventor countries and technology codes. REGPAT contains patents listed under the Patent Cooperation Treaty and the European Patent Office. We combine the patents from both sources, preferring EPO to PCT, by keeping the PCT entries only where the patent is not filed under both.

In terms of technological definitions, patents are classified under different technological classes: under the Cooperative Patent Classification (CPC) scheme and the International Patent Classification (IPC). These schemes provide a very granular product identification, which we can aggregate in a tailored technological definition. We selected the definition of low-carbon technologies based on the Joint Research Centre's definition, as in Fiorini *et al* (2017). CPC codes are grouped for 14 technologies, namely solar panels, hydrogen-related technologies, solar and thermal energy, wind energy, hydro energy, energy management, efficient lighting, efficient heating and cooling, combustion, residential insulation, bio-fuels, batteries, electric cars, efficient rail transport and nuclear energy. The list of relevant CPC-Y codes is in table A.1 of the Appendix. We use IPC definitions for all the technologies and we use only the Y class CPC codes to identify low-carbon technologies.

In the following sections, we provide details about the estimations of revealed technological advantage and potential technological advantage, which will be included respectively in the left and right-hand sides of our final dataset. Meanwhile, the economic, demographic and labour market variables included in the second part of the analysis are based on the full Eurostat and Urban Data Platform[1] databases, and cover a very wind range of fields. We discuss these sources in detail in section 4.

3 Empirical strategy

3.1 Estimation of potential advantage

First, we define the technological advantage of regions in specific technologies. In this section, we largely build on the previous work of Zachmann and Roth (2018), which offered a detailed description of the steps involved. The main difference is in the geographical scope of the estimations. While Zachmann and Roth (2018) worked at the national level, we work here on the NUTS2[2] level.

We start from the definition of Revealed Technological Advantage (RTA), calculating the relative specialisation in patents of a region. This measure is based on the same concept as the Balassa-inspired measures of Revealed Comparative Advantage (RCA), which are built on export, rather than patent, data. These measures 'standardise' patent counts, indicating the relative specialisation of a region in a particular technology. Formally, the revealed technological advantage in for a region is a fraction of two shares:

(1)

$$RTA = \frac{\frac{x_{il}}{\sum_i x_{il}}}{\frac{\sum_l x_{il}}{\sum_{i,l} x_{il}}}$$

where:

x_{il} is the number of patents of technology *i* in region *l*

x_{il} is the sum of patents of technology *i* across all regions

x_{il} is the sum of all patents across all regions

[1] The Urban Data Platform was created jointly by the European's Commission Joint Research Center and the Directorate General for Regional and Urban Policy (DG REGIO). Available at: https://urban.jrc.ec.europa.eu.

[2] Each of the 244 NUTS2 regions has between 800,000 and 3 million inhabitants.

For the subsequent estimations, all the RTAs are generated while excluding green technologies from the sample. RTAs are, in turn, standardised in the following way:

(2)

$$sRTA = \frac{\frac{RTA-1}{RTA+1}+1}{2}$$

where RTA is derived as in equation (1).

In order to estimate the potential revealed technological advantage of regions (pRTA), we apply the methodology inspired by Hausmann *et al* (2019) and thus bring the work of Zachmann and Roth (2018) to a sub-national level.

This methodology assumes a relationship between the comparative advantage of different products or technologies. For instance, a region's comparative strength in one product can imply a potential strength in another product. This is because there can be a link, a similarity, either between the pair of products or between the pair of regions. The intuition behind potential advantage is the attempt to estimate correlations between regions and technologies that are based on latent factors that are unknown *a priori*. For example, the latent factors that make regions similar could be factor costs, infrastructures, geography or domestic market sizes. These correlations could also be based on technological links (eg similar value chains, technological spill-overs, degree of complexity).

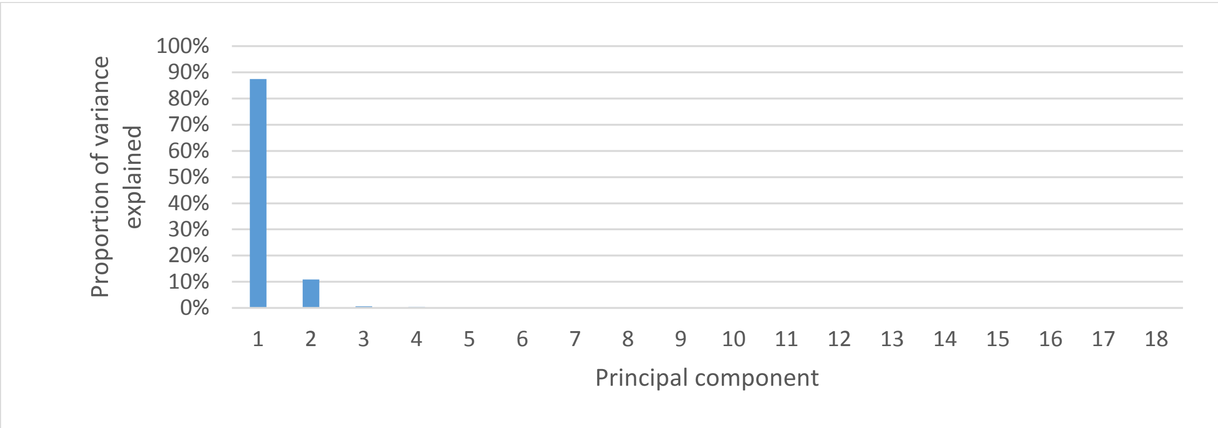
We structure the dataset in three non-overlapping five-year sums of patent counts, from 2001 to 2016, in order to smooth out the volatility in patent activity. We build location-technology cross-tables at four geographical levels (country, NUTS2 region, inventor, application). We apply equation (1) and (2) to the cross-tables, and obtain RTAs.

As in Roth and Zachmann (2018), we construct 18 different region-technology proximity networks. We borrow the definitions of the technological networks from two papers (Yan and Luo, 2017; Stellner, 2014). The methods applied include simple correlations, minimum pairwise conditional probabilities, class-to-class cosine similarity, class-to-patent cosine similarity, co-classification, co-occurrence to generate the networks on the four different aggregations, geographic (regions and countries) and personal (inventors and applicants).

At this stage, we obtain 18 different matrices of technology-region proximity measures, of 258 NUTS2 regions and 637 technology codes. In a final step, all the matrices that contain the weighted product and region densities are stacked vertically, in order to construct two column vectors of IPC class-regions pairs. The column vectors represent 18 weighted networks of technologies and regions (Hidalgo *et al*, 2007). These 18 networks are very collinear, and contain largely similar information, and summarise the latent and observed characteristics of the networks. We apply a Principal Component Analysis and reduce the dimensionality from 18 to 2 principal components. Figure 1 illustrates the portion of variance explained by the principal components, based on the 18 technology-region networks. The correlation between the networks and the first principal component is on average higher than 0.93. The procedure of bringing the dataset from patent counts to principal components is applied separately for all the different, non-overlapping time stacks, in which we originally divided REGPAT data.

At this stage, we can proceed with the estimation of potential advantage, using a regression-based technique as in Roth and Zachmann (2018). For our estimates, we make use of a zero-inflated beta regression. The use of a zero-inflated model is necessary in order to model this information, which has been aggregated from largely sparse matrices (Ospina and Ferrari, 2012). First, we regress the principal components in t_1 on t_2 RTA values. Once the parameters of this model are estimated, we subsequently fit them on the matrices at t_2 and obtain the predicted values for technological advantage for t_3.

Figure 1: Principal components of 18 technology-aggregation networks



Source: Bruegel.

As per the implementation, we rely on the R package *GAMLSS* and its function *BEZI*. We repeat the same approach using linear regressions in order to have a baseline evaluation of our model. The zero-inflated beta regressions show a mean squared error of 0.05 compared to the baseline, and an average R^2, for all the stacks, not higher than 0.35. The statistics, if compared to Roth and Zachmann

(2018), which performed a country-level analysis, exhibit poorer performance of our regional models. We discuss the implications of this and possible ways to improve the models in section 5.1.

The predicted data resulting from fitting the regressions, will be at t_3, and represents our measure of potential technological advantage (pRTA). In the following section, we present some of our pRTA estimations, and give examples of proximity network between regions and technologies.

3.2 The geographical dimension of potential advantage

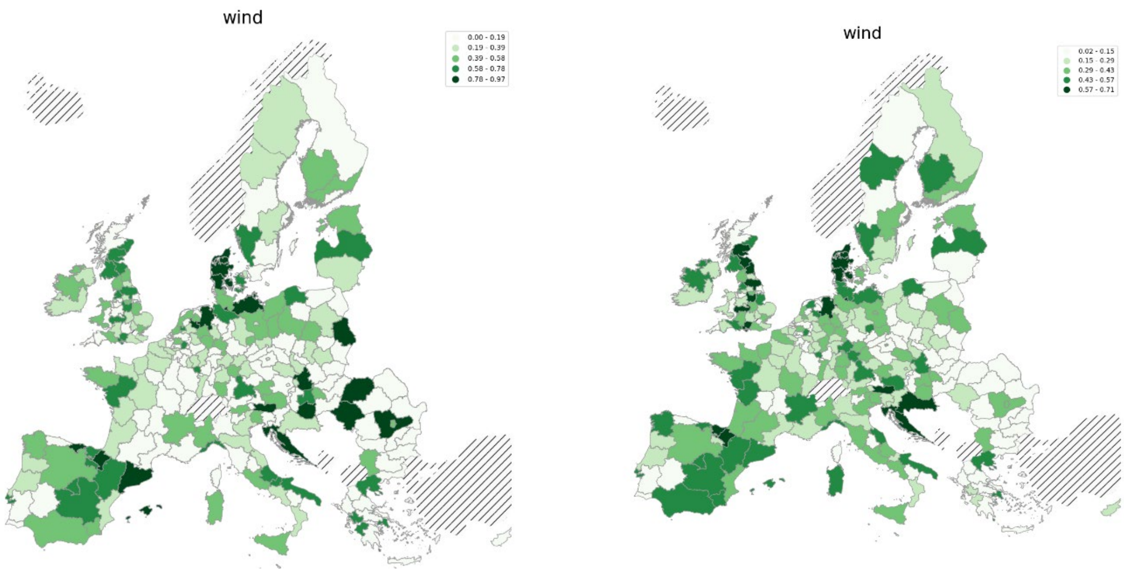
In this section, we explore and visually present the estimated pRTAs, which will be used as regressors in the next part of the paper. After calculating the revealed technological advantage and estimating the potential technological advantage, we observe that certain low-carbon products show a pattern of strong concentration in few regions, such as Rhône-Alpes in France, Dresden and Stuttgart in Germany and Lombardy in Italy: well-known industrial districts and technological hubs. This phenomenon is even more evident when looking at patent counts themselves.

Over time, we observe a general increase in low-carbon technological advantage across Europe, although our measure of RTA seems to be quite volatile, despite the five-year smoothing already mentioned. At this stage, in fact, the trade-off is between the granularity of patent data at the regional level, the time dimension necessary to create a panel, and our technological definitions. The necessity of defining peculiar groups of patents as 'technologies', combined with the sliced time dimension, spurs the volatility in patent activity, especially when observed as RTA. In terms of innovation specialisation, certain technologies, such as nuclear, remain exclusive to a smaller number of regions that are already strong in nuclear technology innovation, as shown in Figure 3. Other technologies, such as wind and hydro power, appear to be promising for many regions.

In Figure 2, it is possible to observe how wind-related technologies had similar geographical distribution for RTAs in 2013 as for potential RTAs for the successive period, 2019. Countries including Denmark, Germany and Spain have at least one region with some degree of specialisation in wind that resulted in a country-level advantage in 2013. Some regions, instead, exhibit strong potential future advantage in wind, despite only modest actual advantage in 2013. This is the case across Scotland, or the Pays de la Loire in the north-west of France.

This, most likely, has to do with the technological complexity involved in producing these products. As mentioned, the other effect at play is the trade-off between technology definitions, time and categories. Because of this effect, nuclear technologies shows much less patenting activity compared to patenting, for example, in solar panels. While the production of products for nuclear power plants involves many sophisticated technologies, the entry barrier for companies is high. Other low-carbon technologies allow easier access for newcomers and thus a wider spread over several countries. Industrial hubs have an advantage or a potential advantage in many low-carbon technologies, as these regions' strengths in various technology areas provide a lot of points of contact also for low-carbon technologies.

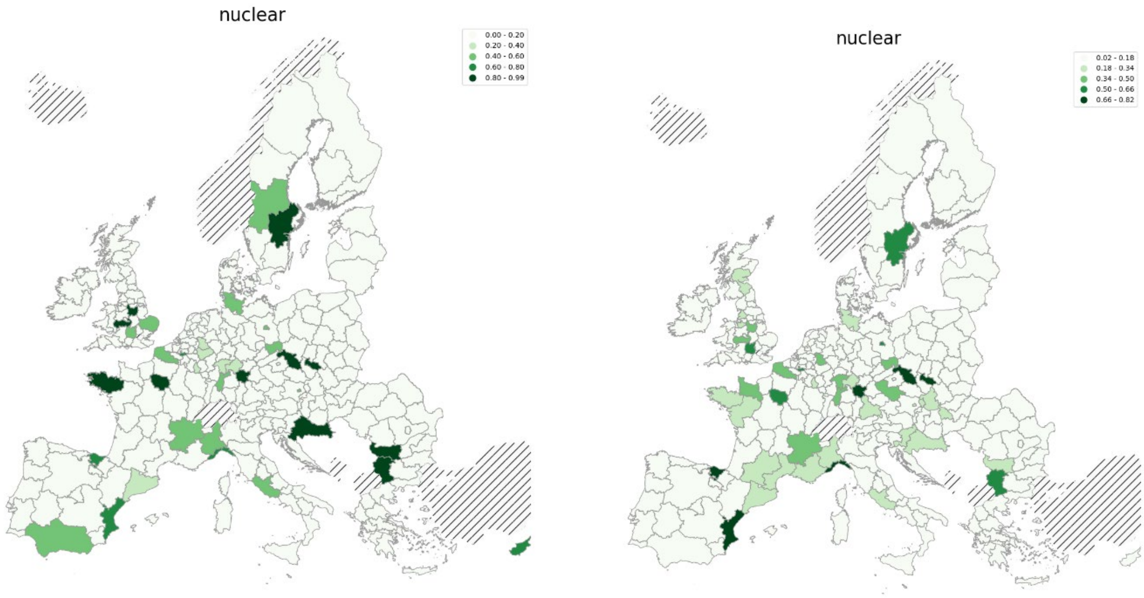
Figure 2: Revealed and potential technological advantage in wind technologies



(a) RTA 2013 (b) pRTA 2019

Source: Bruegel.

Figure 3: Revealed and potential technological advantage in nuclear technologies

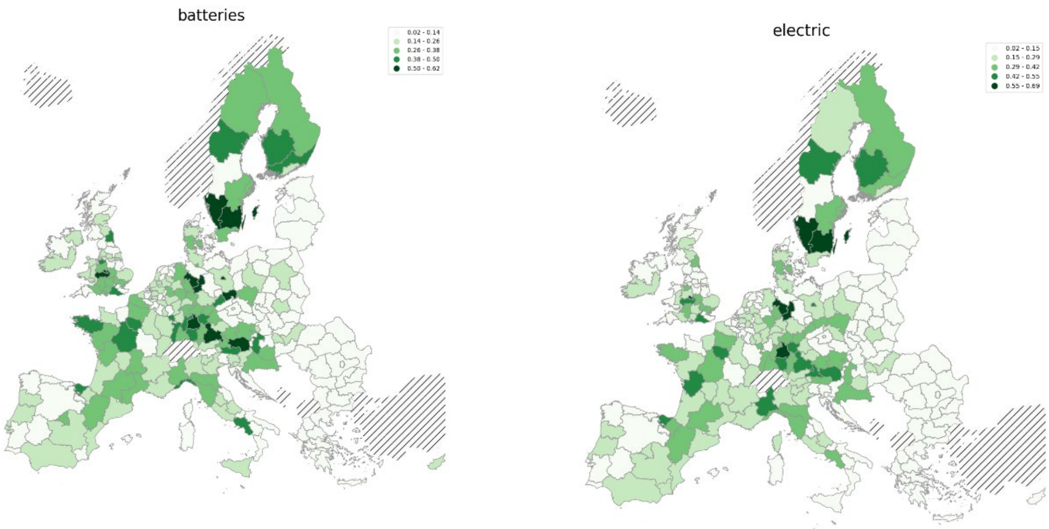


(a) RTA 2013 (b) pRTA 2019

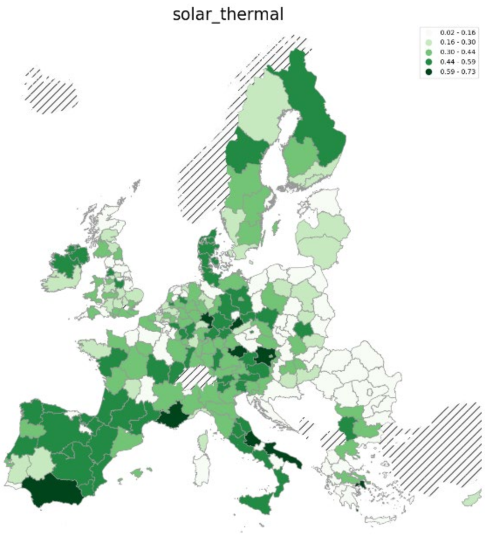
Source: Bruegel.

In the following maps, we present the estimates for pRTA values for different technologies, as of 2019.

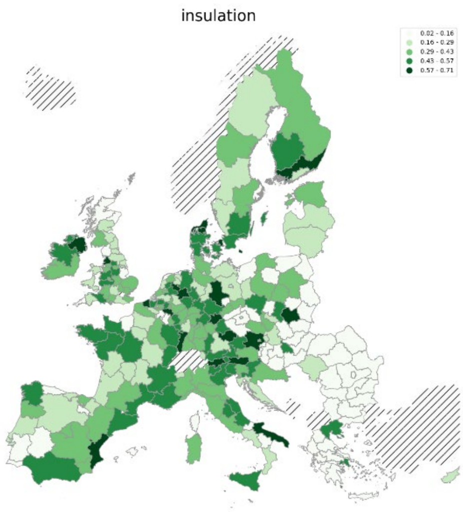
Figure 4: Potential technological advantage in 2019



(a) pRTA 2019: batteries (b) pRTA 2019: electric vehicles



(c) pRTA 2019: solar and thermal energy



(d) pRTA 2019: insulation technologies

Source: Bruegel.

In very complex and typically clustered technologies, such as electric vehicles, our measure of potential advantage seems to be more clustered in highly innovative regions. We correlate the measures for pRTA with the European Commission's innovation scoreboard for the same year. The scoreboard assesses how innovative regions are, based on a set of indicators, including patent activity, workforce education, data from the Community Innovation Survey and other information.

In the first column of Table 1, we present the correlation coefficients, for all the low-carbon technologies observed, between the Commission's indicator for 2019, and the pRTA for the same year. In the second column, we present the correlation between the same indicator and the RTA values, both in 2015. We find high correlation with pRTA for most of the technologies, particularly the most innovative (batteries, electric vehicles and energy management technologies). The volatility issues for the RTA calculations could explain the lower correlation with RTA. Interestingly, however, the discrepancy in correlations could also indicate that our pRTA measure is, indeed, picking up some of those latent factors, 'hidden' in patent data, and observed by the innovation scoreboard.

Table1: Correlations between revealed and potential advantage with the Regional Innovation Index (European Commission, 2019), across low-carbon technologies

	pRTA (2019)	RTA (2015)
Energy managament	0.67	0.28
Solar PV	0.62	0.37
Batteries	0.60	0.27
Electric vehicles	0.59	0.34
Efficient lighting	0.59	0.21
Biofuels	0.58	0.46
Efficient heating/cooling	0.54	0.32
Efficient combustion	0.53	0.38
Insulation	0.49	0.30
Rail	0.39	0.28
Solar thermal	0.34	0.27
Wind	0.29	0.24
Nuclear	0.24	0.25
Hydro	0.10	0.18

Source: Bruegel.

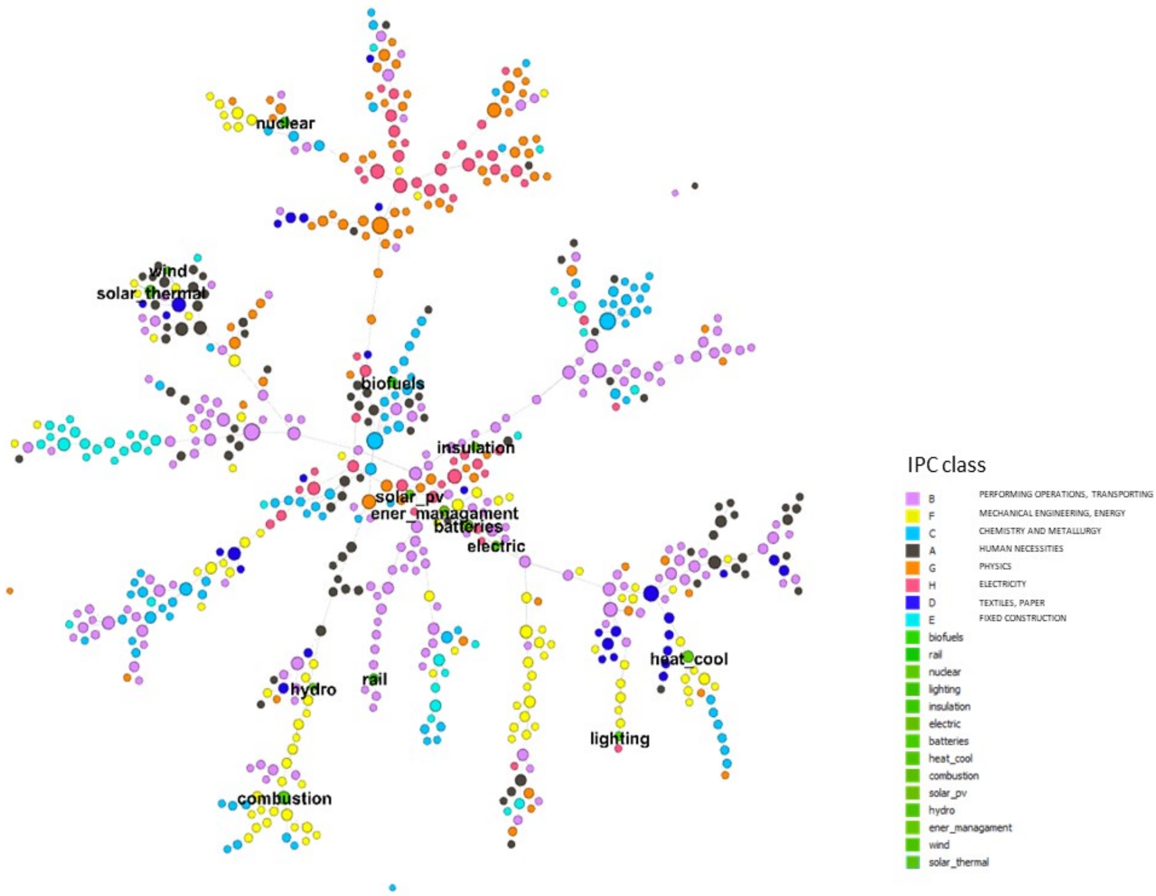
3.3 European networks of green innovation

The networks of proximity that we estimated, as described in section 3.1, can be very informative about the state of low-carbon innovation in Europe. A growing body of literature applies network theory to patent analysis. The technological space is a useful tool for visualisation of technological proximity and might allow forecasts of the direction of technology development in for countries/regions with strength in specific technology areas. Mariani *et al* (2019) focused on patent citations and used network centrality for technological forecasting. Wu and Yao (2012) created and tested on a specific technical field an artificial intelligence-based method for network analysis, combining text-mining techniques. Song *et al* (2016) applied overlay patent networks to analyse the design space evolution by looking at co-references of patents, in order to understand the possible directions of the most likely expansion paths.

In this section, we start by looking at the technological space, with a focus on low-carbon technologies. The technological space resembles the concept of the product space (Hidalgo *et al*, 2007), showing to what extent different technologies are related, based on how often they are patented together. For this, we keep the definition of proximity as simple as possible, using a simple co-patenting figure. Figure 5 plots the European technological space. This graph is built by constructing a technology-technology matrix between of IPC classes for all the technologies and Y-

CPC classes for low-carbon ones. Each node is a 4-digit IPC class or a green technology, and the weight of the nodes is given by the correlations between RTAs in different regions, for the 2013 time stack.

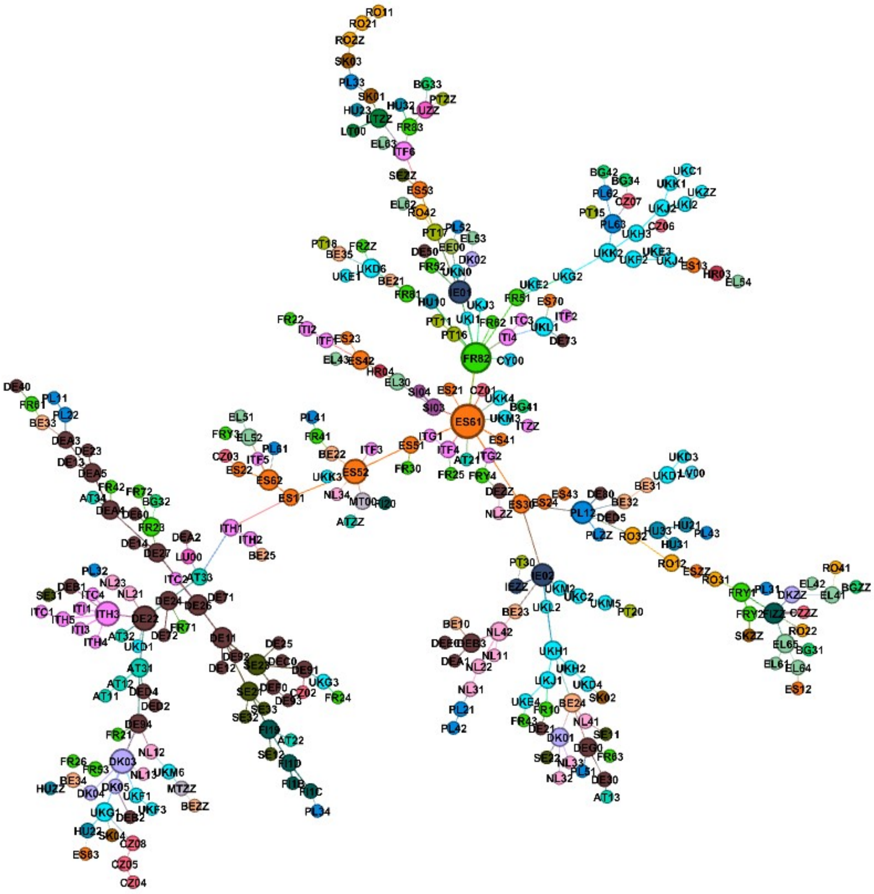
Figure 5: European product space based on patents registered in REGPAT 2018



Source: Bruegel. Note: the network is built on the RTA correlations between IPC 1-digit technological classes and low-carbon technologies based on the relevant CPC Y-codes.

Since the network is extremely dense, as all technologies are connected to some degree to one another, we present here a visualization of the Maximum Spanning Tree (the graph that maximises the total weights of the edges). We can observe how solar panels, energy management, batteries and electric cars seem to have good proximity in the network. Nuclear technologies are closer to classes G and F (physics and mechanical and energy engineering). As expected, rail technologies position close to the class B (performing operations, transportation).

Figure 6: Proximity of NUTS2 regions based on revealed comparative advantage in 2013

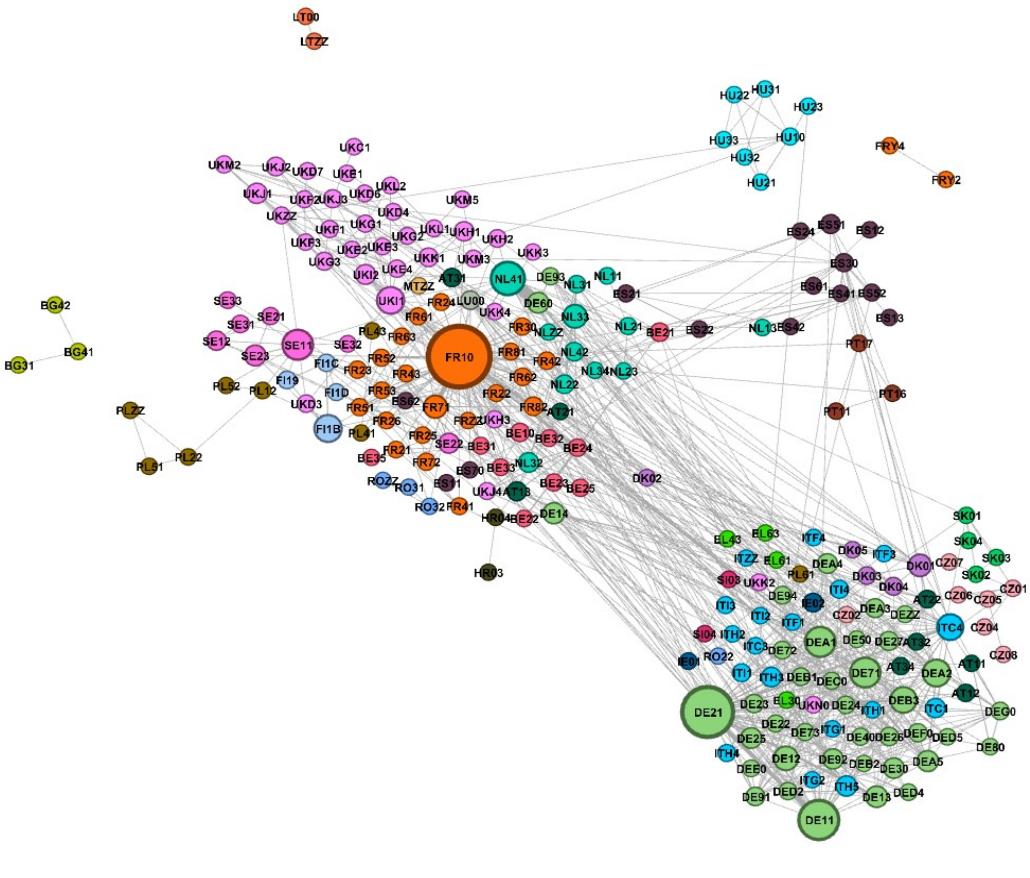


Source: Bruegel. Note: nodes are weighted by a measure of centrality in the graph (betweenness), edges represent correlations in RTA across 637 technological categories.

In the same way, this matrix can be inverted, to show how close regions are. The proximity of regions is based on their technological proximity. In Figure 6, we create a region-region network, in order to understand the proximity of European regions based on the correlations between the technological structures of their economies. The size of the nodes represents their degrees (number of connections with other nodes). A general first observation from this graph is that European regions have product mixes which go beyond country borders. Looking closer at the division, we notice the wedge between productive regions (bottom-left clusters) and less productive regions (mid-right branch). The regions that cluster on the left around Veneto (ITH3) include highly-productive regions, powertrains of European economy, such as Upper Franconia (DE24) or Rhône-Alpes (FR71). Here, the regions of northern Italy and most of the highly industrialized regions of the Germany branch out. Another interesting cluster to the bottom-right: the Dutch region of Noord-Brabant (NL41), a renowned high-tech region, is close to East Anglia (UKH1), the Danish capital region of Hovedstaden (DK01) and Ile de France (FR10). These visualisations are interesting, although more research is needed in order to properly assess the quality and robustness of the clusters, attempting to measure proximities with different metrics.

This network is based on the relative technological advantage of all the 637 4-digit IPC codes, and thus reflects the general innovation activity of regions. In order to observe how European regions collaborate in patenting low-carbon technologies, and observe green industrial clusters, we build networks based on simple co-patenting figures, rather than minimum pairwise correlation probability.

Figure 7: Co-patenting of European regions in low-carbon technologies



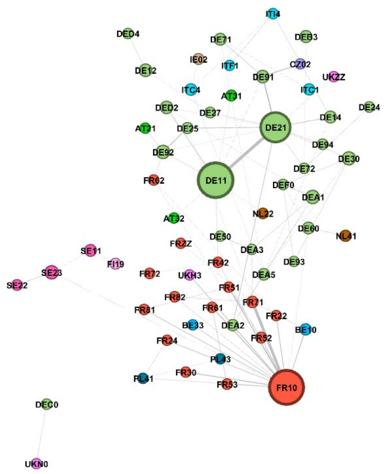
Source: Bruegel.

In Figure 7, we plot a graph based only on low-carbon technologies, in which the size of the nodes is relative to the number of patent applications, and the weight of the edges (i.e. the lines connecting the circles) represents the number of co-patents between two regions. The network is then clustered based on its modularity, a structural measure that tells us how well the graph can be divided into different modules. Specifically, the *OpenOrd* algorithm (Martin *et al*, 2011) is applied.

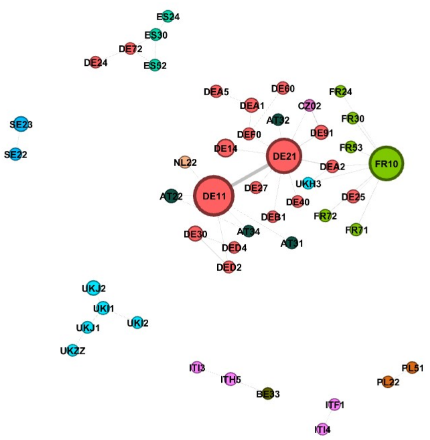
Two clusters emerge, although quite tightly connected. One is dominated by Ile de France (FR10), the region of the capital of France. The other is dominated by Germany, with Oberbayern (DE21) and Stuttgart (DE11). The United Kingdom, the Netherlands, Belgium and Sweden, among other, seem to be clustered more tightly with France, whereas on the other side we see Italy, Germany, Slovakia and Austria. The finding of highly concentrated centres of green innovation is in line with the literature. While the agglomeration phenomena of technological development are well-known, recent studies have focused on green technologies in particular. Barbieri *et al* (2020) found that the nature of green technologies is more complex than non-green technologies. This is illustrated by the agglomeration of low-carbon innovation in high-tech centres such as Paris and Oberbayern. In fact, this is in line with the path-dependency nature of complex technological products, and is logically related to the knowledge base of the region, and spill-over effects across different industries.

In addition, we compute the same co-patenting networks considering one low-carbon technology at a time. In the four panels of Figure 8, we show the examples of batteries, electric vehicles, wind and nuclear technologies. Which are the clusters of innovations? In the case of batteries and electric vehicles, we can see the clusters in France and Germany, whereas for wind, Denmark and Germany are the most central. In these three panels, we filter out the nodes that have no co-patenting and a small number of patents. For nuclear we cannot do this, as most nuclear-patenting regions do not co-patent nuclear patents with other nuclear-patenting regions. Moreover, observed nuclear co-patenting reflects national boundaries.

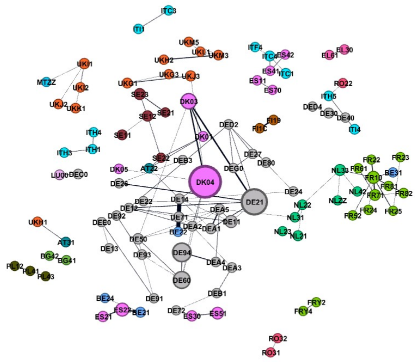
Figure 8: Co-patenting for European regions based on applicant's location, 1990-2016



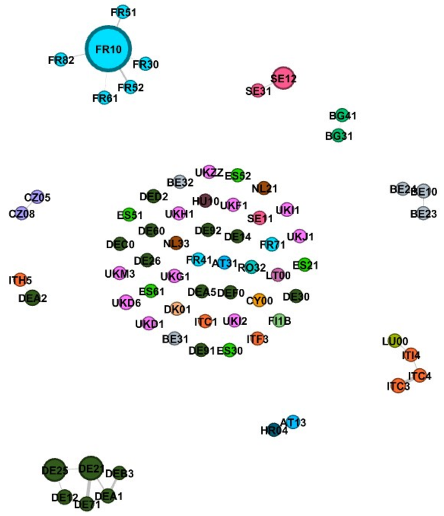
(a) Batteries



(b) Electric Vehicles



(c) Wind



(d) Nuclear

Source: Bruegel.

4 Dimensionality reduction: a data-driven selection of variables

After having calculated the RTA and estimated the pRTA we can now assess which labour market, economic and other regional characteristics are associated, together with potential specialisation, with a higher green RTA.

We investigate this by making use of a large, novel dataset of regional characteristics, and an exploratory approach. We begin with an agnostic view about what regional characteristics could be associated with RTA. First, we build a wide dataset at the NUTS2 level with all the variables present in the Eurostat database and in the Joint Research Centre's Urban Data Platform. The time dimension corresponds to the three-year non-overlapping period of the patents-based measures. Empirically, we rely on dimensionality reduction algorithms. We aim at understanding the association between our measures for revealed advantage and potential advantage in a region, and the large number of regressors on the right-hand side. We select relevant variables in the dataset by applying an elastic net regularisation to isolate the significantly associated coefficients.

In section 4.1, we explain the imputation methodology that we apply to the panel dataset at the NUTS2 level, in order to fill in the gaps across regions and time. In section 4.2, we present our elastic net regularisation for a selection of data-driven variables.

4.1 Data and imputation methodology

We start to build our database with an indiscriminate systematisation of all the NUTS2-level variables present in the Eurostat Database and the Cambridge Econometrics data, publicly available at the Joint Research Centre's Urban Data Platform. Data collection of regional statistics at the NUTS2 level, though improving, is highly inconsistent. As a result, many of the data sheets used to generate our dataset are incomplete in terms of time and location.

As a consequence of the indiscriminate scraping and querying techniques used to generate it, the dataset is incomplete and highly multicollinear, containing several repeated and aggregated indicators. Proper utilisation of NUTS2 regional statistics, especially in cases where full datasets are necessary, is therefore a challenge given such complications. Limitations will be discussed further in section 5.1.

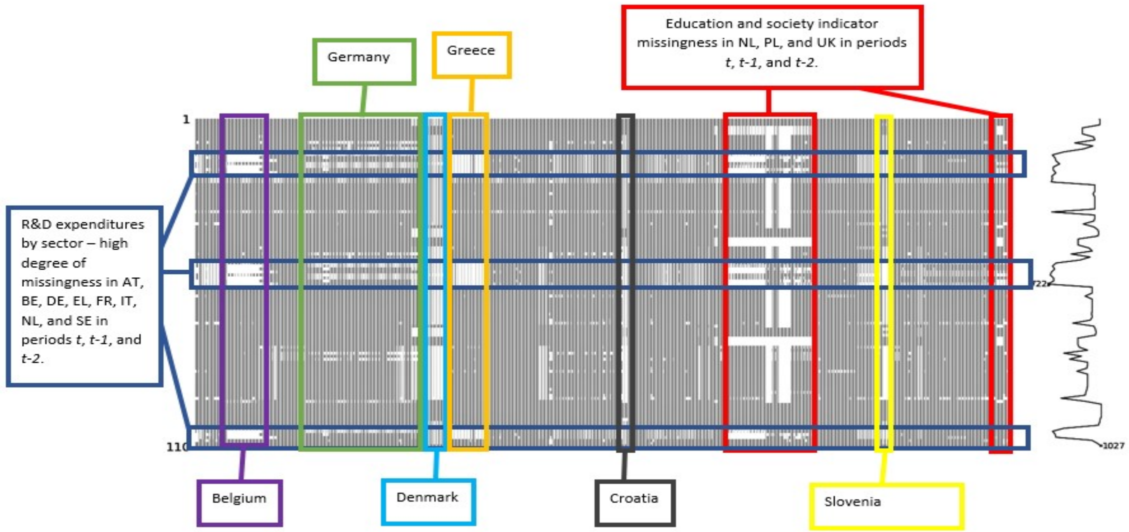
Regularised regression techniques, such as the widely-known LASSO regressions, are often used to reduce dimensionality through variable selection. However, these techniques fail in the presence of missing data, making complete datasets necessary. Therefore, we choose to impute missing values before using any data-driven variable selection technique.

Multiple Imputation by Chained Equations (MICE) is a data imputation algorithm which can attempt to capture the uncertainty associated with missing data values by *"randomly drawing multiple imputations from a distribution of imputations and also by introducing additional error variance to each imputation" (Lodder, 2014)*. MICE makes the assumption that data is missing at random (MAR), meaning that the presence of an underlying relationship between the propensity of a region to be missing data and the value of the missing data causes any results taken from imputation to be problematic (ie due to estimator bias). Additionally, MICE fails when imputing on non-invertible matrices. Therefore, it is important to reduce collinearity as much as possible prior to imputation.

Given the poor performance of advanced imputation methods on high-dimensional low-rank matrices, we first hand-select indicators based on the domain knowledge. We reduce the total number of collected indicators, with respect to the indiscriminate scraping, from 476 to 245, by removing, for example, similar measures of population density, broken down by different demographics. Additionally, we remove regions defined as extra NUTS2 regions (encoded with a ZZ). We also remove overseas territories (eg PT20 or FRY1) or regions with NUTS2 codes that have been replaced (eg UKI1 or IE01), but still persist in Eurostat or Urban Data Platform databases.

Further, these methods often utilise linear regressions to estimate missing values and lose a great deal of stability above certain thresholds of absence. Addo (2018) showed that MICE imputation using non-Bayesian linear regression exhibits stability for datasets missing up to 50 percent of observations. For our purposes, we place this threshold at 30 percent to get a dataset of 110 indicators for 258 NUTS2 regions.

Figure 9: Visualisation of dataset with missing data shown in white



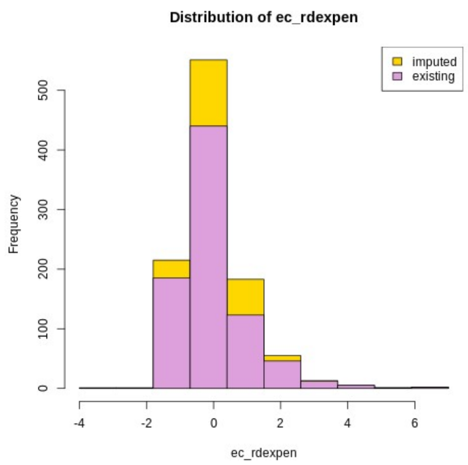
Source: Bruegel. Note: horizontal patterns of absence indicate data missing across regions while vertical patterns indicate absence within a region.

Using the Python missing data visualisation package *missingno*, we examine our dataset for patterns which might invalidate the MAR assumption. Investigation shows that several areas of absence can be attributed to lack of availability (ie Denmark in 2005). Given these patterns and the completeness of the remaining data, we proceed under MAR assumptions and impute with MICE.

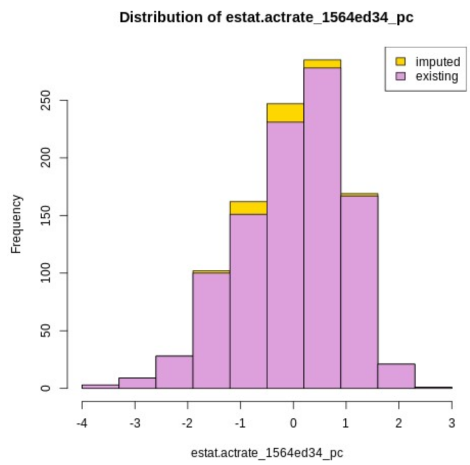
In line with MICE common practice, we first allow the algorithm to identify constant or collinear variables which could present problems during the imputation step. Three covariates are identified as being collinear and are removed. We then impute using MICE using non-Bayesian linear regression. See Figure 10 for histograms showing the results of imputation on selected covariates.

Figure 10: Distributions of selected variables before and after imputation

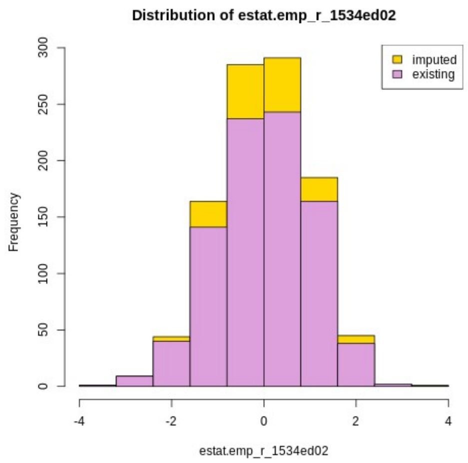
(c) 10c



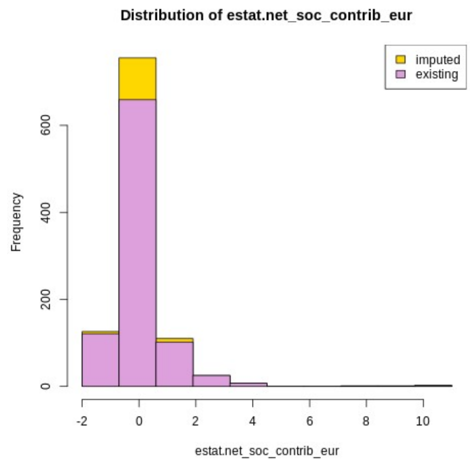
(d) 10d



(a) 10a



(b) 10b



Source: Bruegel.

4.2 Dimensionality reduction

After using MICE imputation to fill in the gaps without biasing the distributions of the variables, we apply elastic net regularisation to the dataset. To be able to assess which regional characteristics (and potentially regional policies) might be promising to engineer regional specialisation, we explore the relationship between current specialisation in green technologies on the one hand, and past potential specialisation and current and past regional characteristics on the other hand. As mentioned, we use an experimental approach, more in line with a machine-learning exercise rather than economic modelling. Therefore, we do not seek to make any causal inference at this stage, but only to look at associations. Future research should encompass causality.

We build a dataset based on four time stacks of RTAs and pRTAs estimated for non-overlapping periods of three years (2006, 2009, 2012, 2015). We aim to observe, row-wise (hence, for each region *r* at time *t),* the potential in that same technology at *t-1,* in order to account for the path dependency in green technologies. Including both the Eurostat and Urban data platforms, it contains 1032 observations and 110 variables, after the first selection from the Eurostat database to reduce multicollinearity, filtering explained in the previous paragraph.

We define RTA, an observed advantage for the low-carbon technology *i* in region *r* at time *t*, as a function of all the right-hand side variables. In turn, the right hand side is composed of all the indicators at time *t* (first term), *t-1* (second term) and *t-2* (third term), as well as potential advantage for that same technology in t-1:

(3)

$$RTA_{r,t,i} = \beta_0 + \sum_{k=1}^{K} \omega_k x_{r,t,i,k} + \sum_{k=1}^{K} \gamma_k x_{r,t-1,i,k} + \sum_{k=1}^{K} \theta_k x_{r,t-2,i,k} + \theta \mathrm{p} RTA_{r,t-1,i}$$

Where:

r = 258 *(NUTS2 regions)*

i = Low-carbon technology

t = 3 years non overlapping timestack

K = 110

Our empirical strategy aims at establishing which of the coefficients is significantly associated with RTA. We implement this specification using a regularisation technique. The two most common applications of regularisation regression are LASSO (Least Absolute Shrinkage and Selection Operator) and Ridge regressions. The intuition behind these techniques is to apply a penalty score to the magnitude of the coefficients of an OLS regression, maximising the relevant ones and shrinking the others to 0.

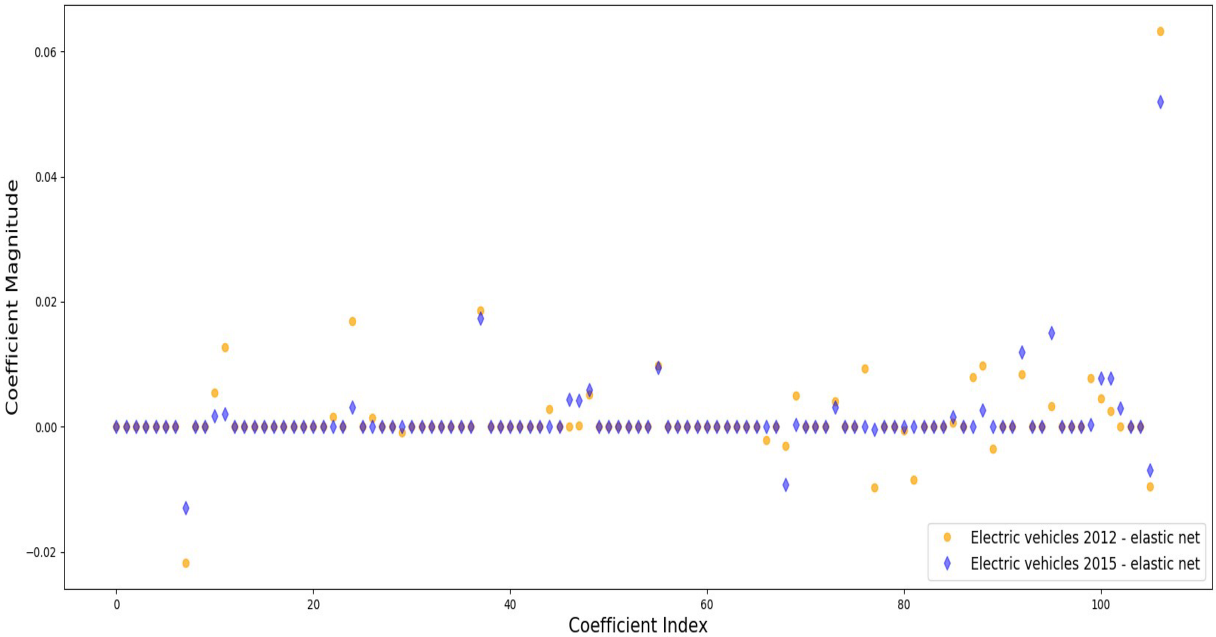
In the literature, there are two typical definitions of regularisation techniques: L1 and L2. The L1 type is known as Ridge regression, while L2 is known as LASSO regression. The main difference between the models is in the application of the ‘penalty factor’ to the cost function. LASSO regularisations can shrink coefficients all the way down to zero, performing a real variable selection. However, in presence of a large number of multicollinear variables, the selection of the feature is done randomly. On the other hand, Ridge can handle the issue of multicollinearity without selecting variables at random, but performs poorly with a high number of dimensions.

Elastic net regularisation combines L1 and L2 approaches, overcoming the respective limitations. This method is more flexible to our purpose, being able to better handle the high multi-dimensionality and collinearity of our dataset. In order to estimate the coefficients for the right-hand side, we first subset the panel to four cross-sectional datasets based on the 2006, 2009, 2012 and 2015 stacks.

We make use of the *ElasticNetCV* module of the python library *sklearn* to perform the regularisation. The module allows an automatic choice of the L1-L2 ratio, influencing the weight given to each penalty factor, by feeding in an array of possible values. The L1-L2 ratio ranges from 0 to 1, and tells how skewed the model should be towards LASSO or Ridge. The array chosen is skewed towards the LASSO-type regression, including more values above 0.5 than below. The *ElasticNetCV* library optimises the parameter selection based on a typical 10-fold cross-validation approach. In the same fashion, the alpha level (overall magnitude of the penalty score) is also selected.

Figure 11 shows the results of the regularisation for electric vehicles, for 2012 and 2015. On the horizontal axis we plot the 110 coefficients estimated, and on the vertical axis their magnitude. The coefficient distancing from the others on the right-hand side is the potential technological advantage (pRTA) at *t-1*. The representation in Figure 11, represents the results of the regularisation: on the y-axis we can see the magnitude of the coefficients for the 2012 (in yellow), and for the 2015 (in blue) time stacks.

Figure 11: Results of the Elastic net regularization where for electric vehicles in the 2012 (yellow) and 2015 stacks (blue)



Source: Bruegel.

After applying the regularisation to each low-carbon technology, we obtain a list of significant variables. These variables are associated with RTA for a low-carbon technology, with non-zero coefficients (both negative and positive). We combine together a count of all the 'surviving' variables. Tables 3 and 4 in the appendix present the results respectively for the 2012 and 2015 cross-sections. We provide a count of surviving variables at time *t*, *t-1* and *t-2*, and in total. Potential advantages at *t-1* are found to be always the highest-ranking coefficients, and are omitted from the tables. As expected, the variables surviving more often are related to the presence of a highly-educated labour force, higher wages and people employed in scientific fields. We discuss these results in the next section.

At this stage, the count of surviving variables is only based on the separate cross-sectional exercises. The panel structure is not exploited, as the cross-sections are considered separately.

In order to observe time-varying effects of these variables, we rescale the panel dataset to have a mean of zero, following a two-way fixed effects approach as in Imai and Kim (2020). However, unlike the two-way methodology, the panel is only rescaled for region-fixed effects.

We subtract from each variable the average value for that region over the year, on a region-time basis. Subsequently, we perform an elastic net regularisation, similar to the previous estimates, this time under the assumption of fixed effects estimation. We rely on the R library *glinternet* for this implementation. The results of the fixed effects regularisation are presented in Table 5 in the appendix. The R^2 values yielding from this model are all lower than 0.5 percent, suggesting that the variation in RTA comes more from cross-sectional differences rather than time-varying effects. It is not possible to perform regularisations that include interaction terms, in this fixed effects setup, as explained by Giesselmann and Schmidt-Catran (2018).

5 Results

5.1 Variables that associate with RTA

Tables 2, and 3 (set out in more length in the appendix), summarise the results of the regularization via elastic net, by showing the number of times a coefficient is found to be non-zero, at time *t*, *t-1*, *t-2* and in total. None of our results are causally identified, and should only be read as associations. The variable that survives the greatest number of times is the activity rate (labour participation) of people that have an ISCED level of education[3], higher than the upper secondary level.

Table 2: Cross-sectional estimation 2012

	t	t-1	t-2	all
Activity rate of population	15	16	21	52
Total duration of employment	13	21	14	48
Scientists and engineers	4	7	7	18
HH Paid current taxes on income wealth etc. mil EUR	5	6	5	16
HH Social benefits other than social transfers in kind recived mil EUR	5	8	3	16
Persons employed in science and technology	8	2	4	14
Long-term unemployment (12 months or longer) in thousands	6	3	5	14
Unemployment rate by age	4	5	5	14

Highest counts of coefficient survival for the 2012 stack, in an elastic net implementation.

Source: Bruegel.

[3] International Standard Classification of Education, classified by UNESCO.

Table 3: Cross-sectional estimation 2015

	t	t-1	t-2	all
Activity rates of population	12	14	13	39
Total duration of employment	11	12	10	33
Scientists and engineers	6	7	5	18
Average number of usual weekly hours in main job by age in hours	7	4	5	16
Persons employed in science and technology	8	4	4	16
HH Paid current taxes on income wealth etc. mil EUR	4	6	4	14
Long term unemployment (12 months or longer) in thousands	7	2	4	13

Highest counts of coefficient survival for the 2012 stack, in an elastic net implementation.

Source: Bruegel.

A longer duration of employment is also associated positively in many of the specifications with the advantage in low-carbon technologies. The number of engineers and scientists is also significant. In general, higher educational attainments of the population are associated with more innovation in green technologies. We also find some evidence of a positive association with tertiary educational attainment and (lower) unemployment rates of women.

In addition, different measures for R&D expenditure survive for all the green technologies. Both private and public sector spending result from our analysis, and seem to appear more often at time *t* rather than as a lag. The measure that scores higher in our framework, for the fixed effects model, is the percentage of GDP spent by the government on R&D. Intramural expenditure, defined by the OECD as the: *"amount of money spent on R&D that is performed within a reporting unit"* (Frascati Manual, 2015). In a causal model, future research shall try to establish the precise relationship between R&D expenditure in the private sector, in higher education institutions and companies.

Tables 4 and 5, meanwhile, present the counts of coefficients indicated by the regularisation procedure, one-way demeaned panel approach discussed in the previous section. Table 4 summarises the main effects, while table 5 summarises the interaction factors that we force with pRTA at *t-1*. Overall, our fixed-effects regularisation approach does not yield different results compared to the cross-sectional specifications. As mentioned, the lower R^2 could be indicative of a higher variation coming from the cross-regional rather than from time-varying effects.

5.2 Limitations and further research

The first stage of this paper used zero inflated beta regressions to predict pRTA values. Looking at the R^2 statistics, not higher than 0.35, suggests that the modelling could be improved. pRTA values correlate at 0.4, on average, with RTA. The definition of the stacks (three years versus a longer period), appears to be less problematic than the modelling, as it does not seem to have too much effect on the volatility of the RTAs, although it makes the RTA measures probably more precise. In further research, a model with less volatile RTAs could yield more consistent predictions for pRTAs.

The most significant limitation of this study is data availability at the NUTS2 level. Prior to dimensionality reduction, over 70 percent of indicators were missing more than 10 percent of observations. Of these, around 60 percent are time-lagged and 40 percent are non-lagged. As a consequence of this missing information, the 30 percent threshold used for keeping covariates is higher than preferred, and yet causes a large loss in availability of data eventually included. As mentioned, the performance of multiple imputation on high dimensional matrices with significant amounts of missing data is improving. However, the stability and quality of imputed values would obviously be much better given a more complete starting dataset. Moreover, because of the need to reduce dimensionality and missing data when using multiple imputation and regularised regression, we may be eliminating covariates or interactions among covariates and pRTA/RTA which have significant predictive power. For some covariates, though, the lack of data is so pervasive that the ability to achieve meaningful predictive power is precluded. Better and more consistent data collection at the regional level will help solve these dilemmas.

Related to the data availability problem, MICE imputation assumes that the absence mechanism of the underlying data is MAR. This would imply, conditional on our observed values, that the values of missing data have no relation to the missing data. After reviewing the patterns of missing data by indicator, region and time, it could be argued that absence is heterogeneous in its mechanism, with some being MAR and others being missing not at random (MNAR). Given a high proportion of missing data being consolidated across similar indicators and time periods, we felt comfortable making a blanket MAR assumption; however, a much closer evaluation of missing data should be performed to confirm this assumption.

Relatedly, adding different NUTS2 level datasets and types of variables could lead to different results. In particular, we believe that the use of datasets with diverse scopes and extensions could be particularly relevant, notably focusing on the infrastructural dimension, market structure, competitiveness and institutions.

Within our current framework, particularly exploratory and inductive, we have recognised the association between our right-hand side variables. Further research should try to establish with a causal approach the relationship between potential advantage in a region and its characteristics in terms of education, labour markets and public and private R&D spending. Furthermore, our empirical methodology should move forward to take into account time-varying effects and the interaction factors between potential and right-hand side variables.

Although RTAs represent the regional specialisation in a technology, the same approach could make use of different models for patent counts (or other types of data that captures innovation) to conduct similar exercises. In this sense and given that our results point towards what is generally recognised as good policy for innovation, further research should also try to model a better counterfactual in terms of what distinguishes green innovation from general technological innovation, leading to more precise policy recommendations.

6 Discussion and possible implications

In this paper we have explored, at European regional level, the extent of innovation in low-carbon technologies. The motivation for our study stems from the necessity that European states will face in the years to come to foster innovation in low-carbon technologies. This is not just because of the need to contain global warming below 2˚C above pre-industrial levels. The decarbonisation process will also lead to a strong change in our production systems, and consequently in labour markets and industrial sectors.

We started by exploring network-based methodologies to predict what advantages regions have in green technologies, based on observed specialisations, and latent, unobserved factors. The first contribution of our study is to provide an atlas of potential specialisation in green technologies for European NUTS2 regions, using regression-based forecasts. Furthermore, by clustering networks of co-patenting, we find that European regions have a good degree of in-country technological diversification, and that European policies could address smart specialisation in green technologies based on a clustering that is country-independent, and is linked to the local economic characteristics. Our representations of co-patenting figures across Europe seem to confirm the agglomeration effects present in innovation in the context of green technologies. In fact, as expected, a small number of leading regions (notably in France and Germany) are pushing the frontier of green patenting. However, we find a large number of others have potential to develop a technological specialisation in the low-carbon technologies in the future.

The European Union Green Deal aims to guide the decarbonisation process, while maintaining industrial competitiveness. Whereas these overarching objectives provide a clear direction, implementation should be informed at regional level, with strategic positioning of regions and low-carbon technologies. Our contribution aims to provide a tool based on innovative methodologies, within the broader concept guiding Smart Specialisation Strategies, which focuses on green technologies.

The second part of this study makes an exploratory contribution to the debate around a horizontal green industrial policy for regions, with a purely data-driven approach. Our results, despite being subject to strong methodological limitations, are generally in line with the common understanding of horizontal industrial policy. Clearly, endogeneity and reverse causality are the most difficult issues to overcome in empirical industrial policy exercises. Our contribution is a first attempt to investigate the role of horizontal green industrial policies for technological specialisation, making use of innovative methodologies at the regional level.

We find an association between a regional advantage in low-carbon products for which there is a higher activity rate for people who have a level of education above upper secondary, and a greater presence of science and technology knowledge-intensive workers. In addition, in terms of labour markets, we have evidence of this association where the total duration of employment is longer. In terms of R&D spending the correlations seem to be significant for both public general spending in R&D, as well as expenditure in the private sector and in higher education institutions. Although our results are in line with those of the literature, our empirical approach has several significant limitations, and should be further refined in order to bring the data-driven selection of variables to a more rigorous third-stage approach aiming to establish robust causal linkages, from which we could infer more informed policy recommendations.

Although other factors play a role in determining competitive advantage, technological specialisation can promote competitive industries, thereby shaping long-run growth dynamics. Policy can leverage strength in similar technologies by shaping innovation paths, strengthening learning capabilities, targeting sector-specific innovation regimes, and coordinating sectoral, national and regional policies.

References

Addo, E.D. (2018) 'Performance Comparison of Imputation Algorithms on Missing at Random Data', Electronic Theses and Dissertations, Paper 3422, available at https://dc.etsu.edu/etd/3422

Balland, P-A., R. Boschma, J. Crespo and D.L. Rigby (2019) 'Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification', *Regional Studies* 53(9):1252-1268

Barbieri, N., A. Marzucchi and U. Rizzo (2020) 'Knowledge sources and impacts on subsequent inventions: Do green technologies differ from non-green ones?' *Research Policy* 49(2): 103901

Boschma, R. and K. Frenken (2011) 'The emerging empirics of evolutionary economic geography', *Journal of Economic Geography*, 11(2): 295-307

Fiorini, A., A. Georgakaki, F. Pasimeni and E. Tzimas (2017) *Monitoring R&I in Low-Carbon Energy Technologies*, JRS Science for Policy Report, Publications Office of the European Union

Frascati Manual (2015) *Guidelines for collecting and reporting data on Research and Experimental Development*, Organisation for Economic Cooperation and Developement

Giesselmann, M. and A. Schmidt-Catran (2018) 'Interactions in Fixed Effects Regression Models', *DIW Berlin Discussion Paper* No. 1748

Hausmann, R., C.A. Hidalgo, D.P. Stock and M.A. Yildirim (2019) 'Implied Comparative Advantage', *HKS Faculty Research Working Paper Series* RWP14-003

Hidalgo, C.A., B. Klinger, A.L. Barabasi and R. Hausmann (2007) 'The product space conditions the development of nations', *Science* 317(5837): 482-487

Hidalgo, C.A. and R. Hausmann (2009) 'The building blocks of economic complexity', *Proceedings of the National Academy of Sciences*, 106(26): 10570-10575

Imai, K. and I. Kim (2020) 'On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data', *Political Analysis*, 1-11

Lim, M. and T. Hastie (2015) 'Learning interactions via hierarchical group-lasso regularization', *Journal of Computational and Graphical Statistics*, 24(3): 627-654

Lodder, P. (2014) 'To impute or not impute: That's the question', in G.J. Mellenbergh and H.J. Adèr (eds) *Advising on research methods: Selected topics 2013*, Johannes van Kessel Publishing

Mariani, M.S., M. Medo and F. Lafond (2019) 'Early identification of important patents: Design and validation of citation network metrics', *Technological Forecasting and Social Change*, 146: 644-654

Martin, S., W. Brown, R. Klavans and K. Boyack (2011) 'OpenOrd: An Open-Source Toolbox for Large Graph Layout', *Proceedings of SPIE*, The International Society for Optical Engineering 7868: 786806

Montresor, S. and F. Quatraro (2019) 'Green technologies and Smart Specialisation Strategies: a European patent-based analysis of the intertwining of technological relatedness and key enabling technologies', *Regional Studies*, pp.1-12

Oketch, T. (2017) 'Performance of Imputation Algorithms on Artificially Produced Missing at Random Data', Electronic Theses and Dissertations, Paper 3217, available at https://dc.etsu.edu/etd/3217/

Ospina, R. and S.L. Ferrari (2012) 'A general class of zero-or-one inflated beta regression models', *Computational Statistics Data Analysis* 56(6): 1609-1623

Stellner, F. (2014) 'Technological distance measures: theoretical foundation and empirics', paper for DRUID Society Conference 2014

Song, B., G. Triulzi, J. Alstott, B. Yan and J. Luo (2016) 'Overlay patent network to analyze the design space of a technology domain: the case of hybrid electrical vehicles', in *DS 84: Proceedings of the DESIGN 2016 14th International Design Conference*: 1145-1154

UNEP/EPO (2015) *Climate change mitigation technologies in Europe – evidence from patent and economic data*, United Nations Environment Programme and the European Patent Office

Wu, C.C. and C.B. Yao (2012) 'Constructing an intelligent patent network analysis method', *Data Science Journal*, 011-003

Yan, B. and J. Luo (2017) 'Measuring technological distance for patent mapping', *Journal of the Association for Information Science and Technology* 68.2: 423-437

Zachmann, G. (2016) 'An approach to identify the sources of low-carbon growth for Europe', *Policy Contribution* 2016/16, Bruegel

Zachmann, G. and A. Roth (2018) 'Learning for Decarbonisation: start early, concentrate on promising technologies, exploit regional strength and work with your national system', *Policy Brief* October 2018, COP21 RIPPLES project

Zachmann, G. and R. Kalcik (2018) 'Export and patent specialization in low-carbon technologies', *Global Innovation Index* 107

Zou, H. and T. Hastie (2005) 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society*, series B (statistical methodology), 67(2): 301-320

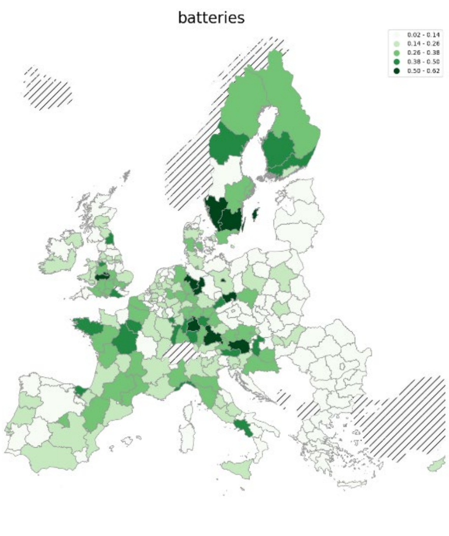
Appendix

A) List of CPC-Y codes for low-carbon technologies definitions

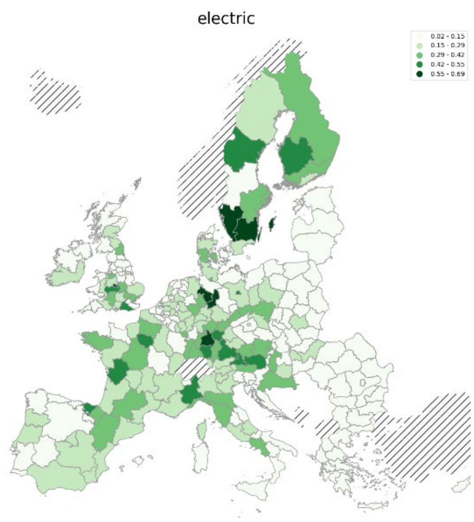
Technology	CPC-Y codes (patents)
Solar PV	Y02E1050, Y02E1052, Y02E1054, Y02E10541, Y02E10542, Y02E10543, Y02E10544, Y02E10545, Y02E10546, Y02E10547, Y02E10548, Y02E10549, Y02E1056, Y02E10563, Y02E10566, Y02E1058
Solar Thermal	Y02E1040, Y02E1041, Y02E1042, Y02E1043, Y02E1044, Y02E1045, Y02E1046, Y02E10465, Y02E1047
Wind	Y02E1070, Y02E1072, Y02E10721, Y02E10722, Y02E10723, Y02E10725, Y02E10726, Y02E10727, Y02E10728, Y02E1074, Y02E1076, Y02E10763, Y02E10766
Hydro	Y02E1020, Y02E1022, Y02E10223, Y02E10226, Y02E1028
Energy management	Y02B7030, Y02B7032, Y02B703208, Y02B703216, Y02B703225, Y02B703233, Y02B703241, Y02B70325, Y02B703258, Y02B703266, Y02B703275, Y02B703283, Y02B703291, Y02B7034, Y02B70343, Y02B70346
Lighting	Y02B2010, Y02B2012, Y02B20125, Y02B2014, Y02B20142, Y02B20144, Y02B20146, Y02B20148, Y02B2016, Y02B2018, Y02B20181, Y02B20183, Y02B20185, Y02B20186, Y02B20188, Y02B2019, Y02B2020, Y02B20202, Y02B20204, Y02B20206, Y02B20208, Y02B2022, Y02B2030, Y02B2032, Y02B20325, Y02B2034, Y02B20341, Y02B20342, Y02B20343, Y02B20345, Y02B20346, Y02B20347, Y02B20348, Y02B2036, Y02B2038, Y02B20383, Y02B20386, Y02B2040, Y02B2042, Y02B2044, Y02B20445, Y02B2046, Y02B2048, Y02B2070, Y02B2072
Heating and cooling	Y02B3008, Y02B3010, Y02B30102, Y02B30104, Y02B30106, Y02B30108, Y02B3012, Y02B30123, Y02B30126, Y02B3014, Y02B3016, Y02B3018, Y02B3020, Y02B3022, Y02B3024, Y02B3026, Y02B3028, Y02B3050, Y02B3052, Y02B3054, Y02B30542, Y02B30545, Y02B30547, Y02B3056, Y02B30563, Y02B30566, Y02B3060, Y02B3062, Y02B30625, Y02B3064, Y02B3066, Y02B3070, Y02B3072, Y02B3074, Y02B30741, Y02B30743, Y02B30745, Y02B30746, Y02B30748, Y02B3076, Y02B30762, Y02B30765, Y02B30767, Y02B3078, Y02B3080, Y02B3090, Y02B3092, Y02B3094
Combustion	Y02B8010, Y02B8012, Y02B8014, Y02B8020, Y02B8022, Y02B8024 Y02B8026, Y02B8028, Y02B8030, Y02B8032, Y02B8034, Y02B8040, Y02B8050
Residential insulation	Y02E2010, Y02E2012, Y02E2014, Y02E2016, Y02E2018, Y02E2030, Y02E2032, Y02E20322, Y02E20324, Y02E20326, Y02E20328, Y02E2034, Y02E20342, Y02E20346, Y02E20348, Y02E2036, Y02E20363, Y02E20366, Y02E20185, Y02E20344
Biofuels	Y02E5010, Y02E5011, Y02E5012, Y02E5013, Y02E5014, Y02E5015, Y02E5016, Y02E5017, Y02E5018, Y02E5030, Y02E5032, Y02E5034, Y02E50343, Y02E50346

Batteries	Y02E6012, Y02E60122, Y02E60124, Y02E60126, Y02E60128, Y02T1070, Y02T107005, Y02T107011, Y02T107016, Y02T107022, Y02T107027, Y02T107033, Y02T107038, Y02T107044, Y02T10705, Y02T107055, Y02T107061, Y02T107066, Y02T107072, Y02T107077, Y02T107083, Y02T107088, Y02T107094, Y02T1072, Y02T107208, Y02T107216, Y02T107225, Y02T107233, Y02T107241, Y02T10725, Y02T107258, Y02T107266, Y02T107275, Y02T107283, Y02T107291
Electric cars	Y02T1064, Y02T10641, Y02T10642, Y02T10643, Y02T10644, Y02T10645, Y02T10646, Y02T10647, Y02T10648, Y02T10649, Y02T1062, Y02T106204, Y02T106208, Y02T106213, Y02T106217, Y02T106221, Y02T106226, Y02T10623, Y02T106234, Y02T106239, Y02T106243, Y02T106247, Y02T106252, Y02T106256, Y02T10626, Y02T106265, Y02T106269, Y02T106273, Y02T106278, Y02T106282, Y02T106286, Y02T106291, Y02T106295
Rail transport	Y02T3000, Y02T3010, Y02T3012, Y02T3014, Y02T3016, Y02T3018, Y02T3030, Y02T3032, Y02T3034, Y02T3036, Y02T3038, Y02T3040, Y02T3042
Nuclear	Y02E3030, Y02E3031, Y02E3032, Y02E3033, Y02E3034, Y02E3035, Y02E3037, Y02E3038, Y02E3039, Y02E3040

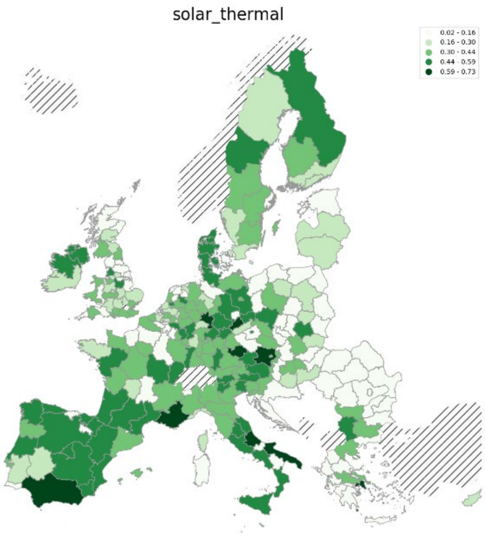
B) Potential in low-carbon technologies for EU regions (2018)



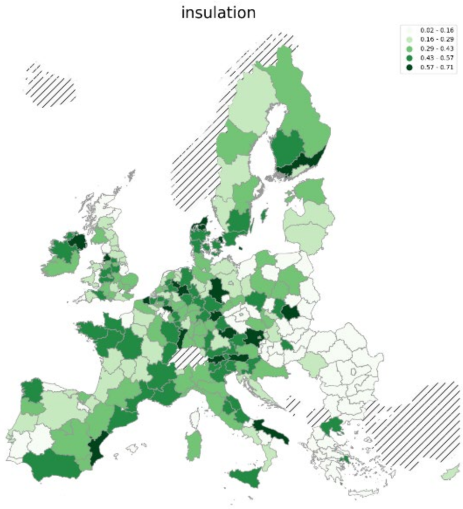
pRTA 2019 - Batteries



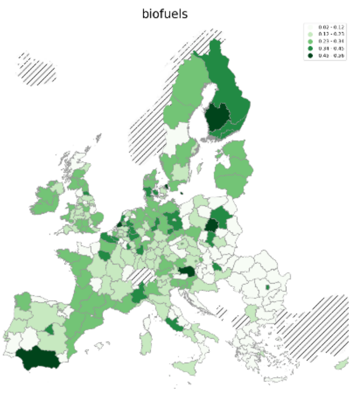
pRTA 2019 - Electric Vehicles



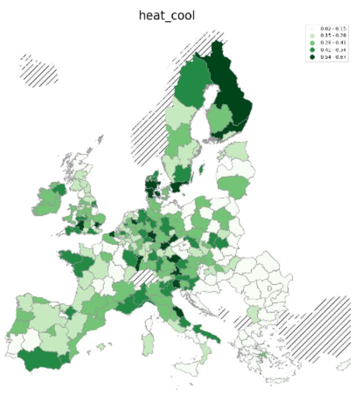
pRTA 2019 – Solar and thermal energy



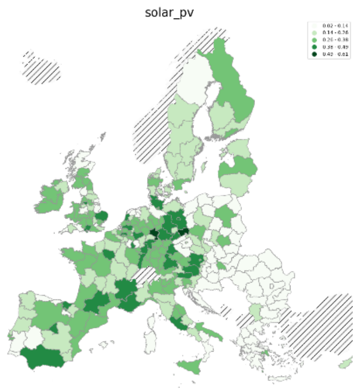
pRTA 2019 - Insulation technologies

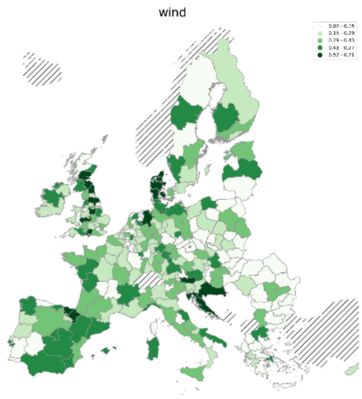


pRTA 2019 - Biofuels



pRTA 2019 - Heating and Cooling





pRTA 2019 – Solar panels

pRTA 2019 - Wind technologies

C) Variables selection summary tables

Table 2: Cross-sectional estimation 2012

	t	t-1	t-2	all
Activity rates by age education attainment level and citizenship	15	16	21	52
Total duration	13	21	14	48
Scientists and engineers	4	7	7	18
HH paid current taxes on income wealth etc. € milliions	5	6	5	16
HH social benefits other than social transfers in kind received € millions	5	8	3	16
Persons employed in science and technology	8	2	4	14
Long-term unemployment (12 months or longer) in thousands	6	3	5	14
Unemployment rate by age	4	5	5	14
Average number of usual weekly hours in main job by age in hours	6	1	5	12
Participation rate in education and training (last 4 weeks) total age 25-64	0	6	5	11
Gross domestic expenditure on R&D € millions government	5	6	0	11
Self-employed persons	2	4	5	11
HH net social contributions € millions	3	4	4	11
Gross domestic expenditure on R&D, € millions business enterprise sector	6	4	0	10

Table 3: Cross-sectional estimation 2015

	t	t-1	t-2	all
Activity rates by age education attainment level and citizenship	12	14	13	39
Total duration	11	12	10	33
Scientists and engineers	6	7	5	18
Average number of usual weekly hours in main job by age in hours	7	4	5	16
Persons employed in science and technology	8	4	4	16
HH paid current taxes on income wealth etc. € millions	4	6	4	14
Long term unemployment (12 months or longer) in thousands	7	2	4	13
Self-employed persons	3	2	6	11
Proportion of population aged 20-39	4	2	5	11
Participation rate in education and training (last 4 weeks) total age 25-64	0	6	5	11
HH social benefits other than social transfers in kind received € millions	4	7	0	11
Gross domestic expenditure on R&D € millions government	8	3	0	11
Persons with tertiary education (ISCED) and/or employed in science and technology % of active population	2	4	3	9
Unemployment rate by age	2	5	2	9

Table 4: Fixed effects estimator, main effects

	t	t-1	t-2	all
Total duration	12	12	17	41
Activity rates ISCED>3	11	12	13	36
Age dependency ratio (0-19 and over 60 to pop. aged 20-59)	4	5	5	14
Scientists and engineers	6	4	4	14
Long term unemployment (12 months or longer) in thousands	6	2	6	14
Students (ISCED 5-6) at regional level - as % of total country level students (ISCED 5-6)	0	6	7	13
Persons employed in science and technology	8	3	2	13
Gross domestic expenditure on R&D € millions business enterprise sector	6	4	3	13

Table 5: Fixed effects estimator, interactions

	t	t-1	t-1	all
Total duration of employment	0	2	3	5
Activity rates ISCED >3	2	1	1	4
HH Net social contributions € millions	0	1	1	2
Self-employed persons	0	2	0	2
Total R&D personnel business enterprise sector full-time equivalent (FTE)	1	1	0	2
Unemployment Rate (Females)	2	0	0	2

bruegel

Bruegel, Rue de la Charité 33, B-1210 Brussels
(+32) 2 227 4210
info@bruegel.org
www.bruegel.org