Knowledge complexity of European metropolitan regions

ADAR project report, WS 17/18

Nico Pintar, h1052260

February 2024

Abstract

Knowledge creation is widely considered as the central driver for innovation, and accordingly, for creating competitive advantage. However, most measurement approaches have so far mainly focused on the quantitative dimension of knowledge creation, neglecting that not all knowledge has the same value (Balland and Rigby, 2017). The notion of knowledge complexity has come into use in this context just recently as an attempt to measure the quality of knowledge in terms of its uniqueness and its replicability. The central underlying assumption is that more complex knowledge is more difficult to be replicated, and therefore provides a higher competitive advantage for firms, or at an aggregated level, regions and countries. The objective of this study is to advance and apply measures for regional knowledge complexity to a set of European regions, and to highlight its potential in a regional policy context.

1 Background and motivation

The ability to create and adapt new knowledge as a pre-requisite for successful innovation is widely considered as an important driver for the competitiveness and productivity of firms, and accordingly the regions where these firms are located (see e.g. Malecki, 2014). Therefore, the development and application of indicators on knowledge production has become one of the central concerns not only in a scientific context per se, but increasingly as a major instrument to orientate and evaluate research and innovation policies (see e.g. Godin, 2003; Katz, 2006), often referred to under the notion of policy learning and evidence-based policy making (Borras and Laatsit, 2019; Sanderson, 2002). In this context, many scholars point to the growth of the *indicators landscape* in terms of quantity and data availability, which is at the same time characterised by more limitations in times of increased complexity of innovation processes and systems (see e.g. Freeman and Soete, 2009).

One weakness that we can specifically observe of indicators on knowledge production – with knowledge production considered as the main underlying driver for innovation – is its focus on *quantity*. In other words, common indicators measuring knowledge production activities, often derived from patenting information, focus on pure quantitative counts of some knowledge outputs, relating such outputs to innovative outcomes. This is also common practice for most policy-oriented indicator systems, such as the European innovation scoreboard (Hollanders et al., 2009). Accordingly, in such works it is implicitly assumed that all knowledge has the same value, i.e. the quality of knowledge is often neglected (Balland and Rigby, 2017).

Against this background, the notion of *knowledge complexity* has come into play more prominently in the recent past (see e.g. Balland and Rigby, 2017). Within this conception, the complex nature of knowledge is associated with its value and quality in terms of accessibility and mobility in (geographical) space, with a higher complexity reflecting increasing quality but decreasing accessibility for others due to its higher degree of *tacitness* and, accordingly, spatial *stickiness*. In essence, the approach is derived from the work presented by Hidalgo and Hausmann (2009) who introduced the complexity notion to grasp the ability of countries to export non-ubiquitous product groups. This was transferred to the concept of knowledge complexity, and to the regional level (Balland and Rigby, 2017; Ivanova et al., 2017; Pintar and Scherngell, 2018) as an equivalent in terms of the knowledge domains regions are capable of. From the perspective of complexity and systems theory (Fleming and Sorenson, 2001; Kauffman, 1993; Simon, 1962), the complexity of knowledge can be well related to the variety of differing knowledge components it contains, and the interdependencies of those components.

With its focus on knowledge complexity, this study lies in the vein of the research stream exploring knowledge production processes and their dynamics, but takes a spatial economics and regional policy perspective in its conceptualisation and application. The objective is to advance measures for regional knowledge complexity, and to apply them to a set of European regions, highlighting its potential in a regional policy context. Furthermore, we apply the regional knowledge complexity measures to a new and more meaningful set of European regions. Most works investigating regional knowledge production rely on standard administrative classifications (mostly NUTS-2) that might artificially intersect agglomerations of knowledge creation (cities) leading to problematic interpretations in a spatial context (Lepori et al., 2019). If a considerable share of residents in areas close to or around large cities commute to work, economic activity in these commuter belts should be counted towards the main city, essentially creating functional areas. The so-called *metropolitan regions* defined by EUROSTAT (2019) - which we use in this study as a unit of observation - aim to do that. The final contribution of this study is to relate regional knowledge complexity - compared to non-complex knowledge production - to regional productivity in a spatial regression framework. As regions are increasingly embedded into their surroundings in their knowledge production activities - be it via formal or informal research and development collaborations - it is important to relax the independent observation assumption of basic linear regression models and explicitly take space into account. Specifically, we employ a spatial durbin model (SDM) that introduces a spatially lagged dependent variable and multiple spatially lagged independent variables to explain regional knowledge production. In order to also explore whether different subsets of the population of regions are differently equipped to properly extract economic (or productivity) gain from complex knowledge production, an effort is made to estimate a spatially eigenvector filtered unconditional quantile regression.

2 Methods and empirical approach

2.1 Operationalisation of knowledge complexity

The concept of economic complexity grasps the ability of countries to export non-ubiquitous product groups, which can only be traded by relatively few countries. The fact that a country is able to export such sophisticated products competitively should signal the existence of a large set of necessary latent (technological) capabilities which in turn should constitute a competitive advantage. Indeed, economic complexity has been shown to be superior in predicting future economic growth of a country to other indicators such as education and institutional quality (Hausmann et al., 2011). This fundamental concept of economic complexity can be very well translated to (regional) knowledge production and has recently been applied in the context of knowledge complexity. In this study, we follow this recent research direction of employing the approach of Hidalgo and Hausmann (2009) to capture the complexity of knowledge of spatial entities (countries or regions), using their technological patent portfolio (patents of a specific technological domain applied for in a specific region). Then we combine – in the same way as done with exports for the economic complexity index – the diversity and ubiquity of the citation-adjusted patent portfolio of a country/region in a knowledge complexity index for countries or regions (see e.g. Antonelli et al., 2017; Balland et al., 2019; Balland and Rigby, 2017; Ivanova et al., 2017; Whittle, 2017). In terms of interpretation, the knowledge complexity index proposed by Balland and Rigby (2017) - and applied in this study - of countries or regions is understood as their ability to create and sustain knowledge bases that are non-ubiquitous in the system.

More specifically, a region's knowledge complexity is understood as a function of its diversity in terms of different technologies produced, and the ubiquity of these technologies, i.e. how many other regions are capable of producing and 'exporting' knowledge related to a specific technological field. Accordingly, the knowledge complexity of regions is based on the region-by-technology network matrix, representing the technological portfolio of all regions as it connects each spatial entity $i = (1 \dots N)$ with technological fields $i = (1 \dots K)$ in which it is specialised in. Similar to previous literature we use the concept of Revealed Comparative Advantage (RCA) by Balassa (1965) to find apparent specialisations of regions in technologies for the time period given by t (subscript t omitted for clarity purposes).

$$RCA_{ik} = \frac{X_{ik}}{\sum_k X_{ik}} / \frac{\sum_i X_{ik}}{\sum_i \sum_k X_{ik}} \tag{1}$$

The RCA of a region in a specific technological field is the ratio between the share of the regions' knowledge production in this field and the share of the same technological field in the whole sample. X_{ik} can be any proxy of knowledge production of region i in technology k. Similar to the majority of related literature, We use patent applications to account for knowledge production. A value larger than one signals a relative regional specialisation in the specific field. Consequently, we define the matrix M as

$$M_{ik} = \begin{cases} 1 \;\; if \; RCA_{ik} > 1 \\ 0 \;\; if \; RCA_{ik} \leq 1, \end{cases} \tag{2}$$

i.e. elements are set to 1 if a region is specialised in a certain technology, and to zero otherwise. M can − from a graph theoretic perspective − also be described as a bipartite graph with two distinct sets of nodes (the N regions and K technological fields) where only nodes of different types can be connected. Region i is connected to field k in the European knowledge production network if, and only if, $M_{ik} = 1$. The diversity in knowledge production of region i is then simply given by its degree centrality, $d_i = \sum_k M_{ik}$. Analogously, the ubiquity of k is equal to its degree centrality, $u_k = \sum_i M_{ik}$. Hidalgo and Hausmann (2009) introduced the so-called *Method of Reflections* in order to infer the complexity of countries (and products) from the network of global exports of products. Translated to our notation and applied to knowledge production, this iterative, self-referential algorithm (see eq. 3) takes regional diversification and the ubiquity of technological fields and then recursively refines these variables with n iterations to yield estimates of regional and technological complexity $(d_i^n; u_k^n)$.

$$\begin{cases} d_i^n = \frac{1}{d_i^0} \sum_k M_{ik} \; u_k^{n-1} \\ u_k^n = \frac{1}{u_k^0} \sum_i M_{ik} \; d_i^{n-1} \end{cases} \tag{3}$$

In other words, this algorithm produces generalised measures of diversification and ubiquity where each iteration uses information from previous iterations to yield a finer estimate of regional and technological complexity, respectively. Each even iteration of d_i^n is a finer estimate of regional knowledge complexity, calculated as the average ubiquity of technological fields (at iteration (n − 1)) in which this region is specialised in. Analogously, each uneven iteration of u_k^n produces a better estimate of technological complexity as the average diversification of regions (at iteration (n − 1)) that are able to produce knowledge in that particular field.

2.2 Theoretical framework and modelling approach

2.3 Data

As elaborated in the previous section, the main focus of this study is to analyse the relationship between regional complex knowledge capital and regional productivity while at the same time explicitly accounting for spatial dependence and spatial heterogeneity in regional innovation activities. We proxy regional innovation activities or knowledge production with patent applications to the European patent office (EPO) by inventors within *metropolitan regions* defined by EUROSTAT (2019). These regions aim to capture urban agglomerations in Europe and defined as aggregates of NUTS-3 regions where at least 50% of the population lives inside a functional urban area that is composed of 250.000 or more inhabitants.

We retrieve patent applications to the European Patent Office (EPO) by EU and EFTA inventors between 1996 and 2016 from the OECD REGPAT database, which offers regionalised patent data. Patents are allocated to NUTS-3 regions in the REGPAT database, where patents are attributed to regions by inventor residence. We map patents located in these NUTS-3 regions to *metropolitan regions* as defined by EUROSTAT and remove (fractional) patents that are located in peripheral regions according to this classification. Similar to related literature, we define knowledge capital as aggregates of patent applications of regional inventors using a five-year moving window. Patent applications associated with the year 2000 (the first period in the sample), for example, are then the sum of patent applications from years 1996 to 2000.

Using patents mapped to *metropolitan regions* regions, it is then possible to calculate regional knowledge complexity scores for each period 2000-2016, following elaborations in section 2.1 and specifically equation 3. Regional complex knowledge capital is then defined as the complexity weighted regional knowledge capital.

Regional productivity is calculated using regional output and regional production factor inputs of labour and capital.
We define the regional total factor productivity index (p) adapted from Caves et al. (1982).

$$p_{it} = q_{it} - s\, l_{it} - (1 - s)\, c_{it} \tag{4}$$

Lower case letters refer to variables in logged form. Here, lower case s is the assumed share of labour costs in the production process. Similar to related studies (e.g. Beugelsdijk, Klasing and Milionis, 2018), we set s equal to 2/3. Regional output (q) is measured via real regional gross value added. Labour input (l) is the number of employees, adjusted by differences in the average working hours per country. The capital stock of a region is defined as the five year sum of past real gross fixed capital formation (investment). All non-patent variables described above are sourced for the period 2000 to 2021 from ARDECO. As the growth of productivity is arguably even more of relevance for policy decisions than the level of productivity per se, we focus in this analysis on the five-year growth rate of productivity as our main dependent variable. We define the five-year growth rate of productivity as the difference of $p_{i(t+5)} - p_{it}$. Consequently, the last year under investigation is $t = 2016$ where the dependent variable refers to TFP growth from 2016 to 2021.

Keeping only regions that produce at 50 patent applications per period (which is necessary to assure sensible knowledge complexity scores) and regions where output, labour and capital data is available until 2021, yields a balanced panel with 192 *metropolitan regions* and 17 time periods (2000-2016).

3 Analysing European regional knowledge complexity in *R*

3.1 Descriptive analysis

3.2 Inferential analysis

4 Conclusion