

The co-evolution of knowledge and economic structure: Evidence from European Regions.

Francesco Quatraro^{a,b*}

a) University of Nice Sophia Antipolis
CNRS-GREDEG
250, rue Albert Einstein
06560 Valbonne
France

b) Bureau of Research on Innovation, Complexity and Knowledge (BRICK)
Collegio Carlo Alberto, Moncalieri (Torino)

ABSTRACT. This paper articulates an analysis of the co-evolutionary patterns of structural change in knowledge and economics. The former is made operationally through the analysis of co-occurrences of technological classes within patent documents, so as to derive the indicators of coherence, variety and cognitive distance. The latter is instead made operational in a synthetic way by implementing a shift share analysis, which allows to decomposing the growth in labour productivity in the effects due to the changing allocation of employment, those ascribed to intra-sector productivity growth and those due to the interaction of these two. The results of the analysis conducted on a sample of 227 European regions, show the existence of interesting lead-led dynamics between knowledge and economic structure that call for further investigation.

* Corresponding address: GREDEG, CNRS and University of Nice Sophia Antipolis, 250 rue Albert Einstein, 06560 Valbonne, France. Tel. +33 4 93954373; Fax: +33 4 93653798. Email address: francesco.quatraro@unice.fr.

1 Introduction

The relationships between technological change and structural change have long been neglected in economic analysis. Indeed, the two subjects have been mostly analyzed independently or in relation to the mechanisms of economic growth, but not much attention has been given to their strict connection. On the one hand, the former scholars interested in the dynamics of structural change explicitly acknowledged the importance of technological change, but did not provide an adequate account of its role (Kuznets, 1930; Burns, 1934). On the other hand, the most prominent scholar in the study of technological change, Joseph Schumpeter, elaborating upon the concept of creative destruction, provided a theoretical ground showing many interesting implications for structural change that have not been properly developed (Schumpeter, 1942).

Within the domain of regional economics, an early effort to integrate the analysis of technological and structural change is due to François Perroux (1955), who integrated the role of technological change in his “growth pole” theory following Schumpeter’s legacy. Regional economic systems are characterized by rounds of growth, i.e. periods in which firms within the propulsive industry grow at faster rates, propagating the positive effects across firms directly and indirectly related to the propulsive industry. The main driving factor of such expansion is technical efficiency gained through innovation efforts.

More recently, the evolutionary approach to economic geography has articulated a framework for the analysis of the linkages between technology and structural change at the local level able to account for the inherent complexity of the dynamics at stake (Boschma and Frenken, 2006 and 2011). In this strand of analysis, the competences accumulated at the local level matters in shaping the process of industrial diversification, so that the change in the allocation of workforce across different sectors is influenced by the degree of technological relatedness amongst the involved activities.

This paper aims at contributing to the ongoing debate by analyzing the co-evolutionary patterns between structural change and technological change at the regional level. In particular, by adopting a collective approach to the analysis of technological knowledge, we propose the concept of knowledge structure (Quatraro, 2012) and propose a framework to analyze its relationship with the changes in economic structure. To this purpose we will couple three different methodological approaches. The former allows to elaborating a set of indicators able to provide a synthetic account of the architecture of knowledge structure. In

particular we will draw upon co-occurrences matrixes to calculate coherence, cognitive distance and variety indicators. Secondly, we will provide a synthetic account of the change of economic structure by implementing a “shift-share analysis” in order to disentangle the contribution to (labour) productivity growth of within-sector productivity dynamics and of reallocation of labour force across the different sectors. Finally, the relationships between these two sets of indicators have been investigated by using a vector autoregression (VAR) model, which we estimated via ‘reduced form’ applying the least absolute deviation (LAD) estimator due to the distributional properties of the variables.

The analysis is carried out on European NUTS II regions and provide an interesting insight into the dynamic feedbacks between economic and knowledge structure. While some relationships goes in the expected direction, in some cases we are confronted with somewhat more articulated patterns that call for a finer grained representation of search behaviours of innovating agents. The rest of the chapter is organized as follows. The next section articulates the theoretical framework proposes the working hypotheses. Section 3 elaborates a model and introduces ‘shift-share’ analysis. Section 4 provides a description of the data used and outlines the econometric strategy. Section 5 present the results of the estimations and a general discussion. Section 6 finally provides some temporary conclusions.

2 Economic Structure and Knowledge Structure: The missing link

The appreciation of the role of structural change in the process of economic growth can be traced back to the seminal contributions by Marshall (1919) and Kuznets (1930). According to Marshall, the advances in industries in which the country already possess a competitive advantage is likely to strengthen international trade. The pattern of industrial specialization is therefore a key aspect influencing the trade between nations. However, competitive advantages are not supposed to characterize the same industries forever, and accordingly industrial leadership of countries is likely to follow the evolution of the main industries they are specialized in. This does not imply necessarily the switch to different activities, which can be eventually attained only as a result of a very slow process. An alternative to cope with the challenges coming from emergent countries which are likely to follow a delayed development path similar to those of the advanced ones, is the introduction of improvements, not only technical, to increase the efficiency of production processes.

Some decades later, in the 1930s, the issue of changes in economic structure came to the fore in the so called three-sectors approach. Key authors in this framework are Arthur Burns (1934), Allan Fisher (1939) and Simon Kuznets (1930). This latter has been clearly the one having proposed a detailed analysis of such dynamics, and he can be surely identified as the founder of the strand of empirical analysis of structural change. The building block of Kuznets' approach is the growth retardation hypothesis, which states that industry growth rates are declining over time, and then that industries whose period of development comes later are likely to overtake the mature ones. This implies that one would observe an alternation of leading industries, and of leading countries as well. Such diversity across industries generates a process of change in the economic structure of production, in terms of relative composition of activities. Differential growth rates across branches of an industry are hence likely to create structural change.

Kuznets also stressed the bearing of Schumpeter's analysis of innovation dynamics upon the analysis of structural change. He noted that the process of creative destruction entails two parts, the creation of new combinations on the one hand, and the destruction of the old ones on the other hand. The introduction of radical innovations alters the structure of the economy, creating new jobs and making the existing ones obsolete. This in turn engenders a dislocating effect upon employment, which tends to shift from the old sector to the new one, with major difficulties in terms of switching costs (Kuznets, 1972). Schumpeter indeed argues that innovation represents the main engine of economic progress within the capitalistic system (Schumpeter, 1928 and 1939), in which "the opening up of new markets, foreign or domestic, and the organizational development [...] illustrate the same process of industrial mutation [...] that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one" (Schumpeter, 1942: p.83).

Although the relevance of technological change in the rejuvenation of mature industries or in the birth of new ones is explicitly acknowledged, there is no systematic effort to bring it into the analysis of structural change. A former contribution in this direction is provided by François Perroux (1955), who develops the implications of Schumpeter's theoretical speculation about innovation on the analysis of the process of regional development. Perroux proposed indeed a view according to which the development of local economies is shaped by centripetal and centrifugal forces. Some sectors are likely to be stronger in some areas and weaker in other areas, so that the economic development of a specific area is influenced by the structural ties of the propulsive sector with the rest of the local economic activities. Vertical and horizontal linkages can therefore enhance the positive

effects of outperforming sectors. Out of the main sources of competitive advantage in this framework, innovation plays a key role in the development of technical efficiency. A few decades later, Thomas (1975) articulated the implications of Perroux' framework on regional economic growth using a product life-cycle perspective, wherein the saturation of product markets are the main responsible for the slowdown of growth rates and the quest for innovations aims at opening new markets.

The interplay between Schumpeterian dynamics and retardation theory seems therefore useful to enhance the understanding of regional differences in the transition dynamics typical of structural change processes (Quatraro, 2009). Boschma and Frenken (2006 and 2011) have recently proposed a far reaching integration of these issues into an evolutionary approach combining explicitly industrial dynamics with economic geography. In this framework, regional growth emerges out of a process of industrial diversification, in which the introduction of new varieties is constrained by the competencies accumulated at the local level. The emergence of new industries at the local level, i.e. the shift away of employment from one sector to another, is influenced by the technological relatedness amongst the sectors. Proximity matters not only in the geographical but also in the technical and technological space (Boschma, 2005; Antonelli and Quatraro, 2012).

The evolutionary economic geography allows to appreciating the role of the competences accumulated over time in local contexts, so as to predict the more likely direction of structural change dynamics. Moreover, it provides the basis to link structural change with search patterns in the technology landscape at the regional level. In this perspective, the grafting of the so called recombinant knowledge approach onto the investigation of the relationship between technological and structural change can be particularly fertile to take into account the heterogeneous nature of regional knowledge dynamics (Quatraro, 2010). This strand of analysis has moved from key concepts brought forward by Schumpeter (1912 and 1942) and Usher (1954), and then elaborated upon the models proposed within evolutionary economics (Nelson and Winter, 1982). The creation of new knowledge is indeed represented as the outcome of a search process across a set of alternative components that can be combined one another. A crucial role is played within this framework by the cognitive mechanisms underlying the search process aimed at exploring the knowledge space so as to identify the pieces that might be combined together. The set of potentially combinable pieces turns out to be a subset of the whole knowledge space. Search is supposed to be local rather than global, while the degree of localness appears to be the outcome of cognitive, social and technological influences. The ability to engage in a search

process within spaces that are distant from the original starting point is likely to generate breakthroughs stemming from the combination of brand new components (Nightingale, 1998; Fleming, 2001; Fleming and Sorenson, 2001; Sorenson et al., 2006). In this direction regional innovation capabilities may be defined as the ability of regional actors to engage in the combinatorial process that gives rise to the structure of the regional knowledge base (Lawson and Lorenz, 1999; Romijn and Albu, 2002; Antonelli, 2008).

The recombinant knowledge approach provides a useful framework to represent the internal structure of regional knowledge bases as well as to enquire into the effects of its evolution (Quatraro, 2012). If knowledge stems from the combination of different technologies, knowledge structure can be conceptually represented as a web of connected elements. The nodes of this network stands for the elements of the knowledge space that may be combined with one another, while the links represent their actual combinations. The frequency with which two technologies are combined together provides useful information on the basis of which one can characterize the internal structure of the knowledge base according to the average degree of complementarity and similarity of the technologies which knowledge bases are made of, as well as to the variety of the observed pairs of technologies (Saviotti, 2004 and 2007; Krafft, Quatraro and Saviotti, 2009).

The combination of the recombinant knowledge approach with the tradition of analysis of structural change provides the bases to appreciate the relationship between the dynamics of knowledge and economic structure. In view of the arguments elaborated so far we can therefore spell out the working hypotheses underlying the analysis conducted in this paper. Technological change is the main driver of economic growth. This has been proven at different levels of aggregation, from the firm-level to the region-level to the country level. However both of the terms of this fairly general relationship can be better qualified in order to appreciate the underlying heterogeneous mechanisms. Not only technological knowledge matters for economic growth, but also the structure of knowledge may have sensible effects on the dynamics of economic growth and even more on the changes in the structure of economic activities. In the following section we will articulate an analytical model relating changes in economic and knowledge structure, which will serve as a guide to the implementation of the empirical strategy that will be presented in the remainder of the paper.

3 A model for knowledge and economic structure: The shift-share analysis.

A formal model linking the change of economic structure to that of knowledge structure can be easily derived by using a traditional Cobb-Douglas production function like the following;

$$Y_{i,t} = AC_{i,t}^{\alpha} L_{i,t}^{\beta} K_{i,t}^{\delta} \quad (1)$$

One can therefore that the production in a region i at time t can be represented by such kind of function, in which C stands for fixed capital, L stands for labour services and K stands for knowledge inputs. As usual, α , β and δ are the output elasticities of capital, labour and knowledge respectively. Following Nesta (2008), let us apply the decomposition of knowledge input as follows:

$$K \cong EDR \quad (2)$$

Where E is the traditional measure of regional knowledge capital stock, D measures technological variety while R represents the coherence of the regional knowledge base. Let us now substitute Equation (2) into (1) as follows:

$$Y_{i,t} = AC_{i,t}^{\alpha} L_{i,t}^{\beta} [E^{\omega_E} D^{\omega_D} R^{\omega_R}]_{i,t}^{\delta} \quad (3)$$

Where ω_E , ω_D and ω_R are the weights attributed to each of the three properties. By multiplying the exponent δ by such weights we obtain the following:

$$Y_{i,t} = AC_{i,t}^{\alpha} L_{i,t}^{\beta} E_{i,t}^{\theta_E} D_{i,t}^{\theta_D} R_{i,t}^{\theta_R} \quad (4)$$

Assume now that such production function is characterized by constant returns to scale in the traditional inputs capital and labour, such that:

$$\alpha + \beta = 1$$

By multiplying both sides of the equation by L^{-1} we obtain

$$(Y/L)_{i,t} = A(C/L)_{i,t}^{\alpha} E_{i,t}^{\theta_E} D_{i,t}^{\theta_D} R_{i,t}^{\theta_R} \quad (5)$$

The left hand side of this equation clearly is a labour productivity index. In order to investigate the relationship between the change in knowledge structure and change in economic performances we need to total differentiate equation (5) as follows:

$$\Delta \left(\frac{Y}{L} \right) = \Delta A \frac{\partial(Y/L)}{\partial A} + \Delta(C/L) \frac{\partial(Y/L)}{\partial(C/L)} + \Delta E \frac{\partial(Y/L)}{\partial E} + \Delta D \frac{\partial(Y/L)}{\partial D} + \Delta R \frac{\partial(Y/L)}{\partial R} \quad (6)$$

Now, after calculating all the derivatives on the right hand side of equation (6), and dividing both sides by (Y/L) we yield the following:

$$\frac{\Delta(Y/L)}{(Y/L)} = \frac{\Delta A}{A} + \alpha \frac{\Delta(C/L)}{(C/L)} + \theta_E \frac{\Delta E}{E} + \theta_D \frac{\Delta D}{D} + \theta_R \frac{\Delta R}{R} \quad (7)$$

Equation (7) relates the change in knowledge characteristics to the change in labour productivity. While this has proven to be a useful result, we still need to decompose ‘generic’ labour productivity growth into the differential contribution provided by changing

reallocation of employment across sectors, i.e. the most traditional utilization of the concept of structural change in economics.

To this purpose, the so-called shift-share analysis provides an interesting methodology that can be integrated in this framework with a few more passages. As noted by Houston (1967), shift the origins of shift-share analysis can be dated back to the seminal work by Daniel Creamer (1942), although it did not reach great success at least until 1960, when Perloff, Dunn, Lampard and Mutt employed it as an analytical tool in their work *Regions, Resources and Economic Growth*. It has been mostly used to investigate disentangle the compositional mix and the competitive position of regions in the face of observed changes in some relevant variables (Esteban, 1972 and 2000). In this chapter we will follow the approach developed by Fagerberg (2000), who decomposed labour productivity in three major components, i.e. the allocative, the productivity differential and the interaction between the two. We start by rearranging labour productivity as follows (region subscripts are omitted for the sake of clarity):

$$\frac{Y}{L} = \frac{\sum_j Y_j}{\sum_j L_j} = \sum_j \left[\frac{Q_j}{N_j} \frac{N_j}{\sum_j N_j} \right] \quad (8)$$

Labour productivity at the system level can therefore be decomposed in the contribution provided by labour productivity of each sector j as well as by share of sector j in total employment.

If we set:

$$P_j = \frac{Q_j}{N_j} \quad (9)$$

$$S_j = \frac{N_j}{\sum_j N_j} \quad (10)$$

Then:

$$\frac{Y}{L} = \sum_j [P_j S_j] \quad (11)$$

The variation in labour productivity can be therefore expressed as follows:

$$\Delta \frac{Y}{L} = \sum_j [P_{j,t-1} \Delta S_j + \Delta P_j S_j + S_{j,t-1} \Delta P_j] \quad (12)$$

Equation 9.12 can be therefore expressed in growth rates by dividing it by (Y/L) :

$$\frac{\Delta(Y/L)}{(Y/L)} = \sum_j \left[\frac{P_{j,t-1} \Delta S_j}{(Y/L)} + \frac{\Delta P_j S_j}{(Y/L)} + \frac{S_{j,t-1} \Delta P_j}{(Y/L)} \right] \quad (13)$$

The first term between parentheses is the contribution to productivity growth from **changes in the allocation of labour** between industries. It will be positive if the share of high productivity industries in total employment increases at the expenses of industries with low productivity. The second term measures the **interaction between changes in productivity** in

individual industries **and changes in the allocation of labour** across industries. It will be positive if fast growing sectors in terms of productivity will also increase their share in total employment. The third term is the **contribution from productivity growth within industries**.

We can now substitute Equation (13) into equation (7) to articulated in an explicit form the relationship between change in economic and knowledge structure:

$$\sum_j \left[\frac{P_{j,t-1} \Delta S_j}{(Y/L)} + \frac{\Delta P_j \Delta S_j}{(Y/L)} + \frac{S_{j,t-1} \Delta P_j}{(Y/L)} \right] = \frac{\Delta A}{A} + \alpha \frac{\Delta(C/L)}{(C/L)} + \theta_E \frac{\Delta E}{E} + \theta_D \frac{\Delta D}{D} + \theta_R \frac{\Delta R}{R} \quad (14)$$

Equation (14) provides a useful starting point to the elaboration of an empirical strategy for the assessment of the dynamic interactions between structural change in knowledge and the economy. However, as we moved from a Cobb-Douglas production function, one would think that the l.h.s. of the equation is a function of the r.h.s., which would be clearly inconsistent with the main hypothesis of this book, according to which structures are endogenous and mutually interdependent. For this reason, we will use Equation (14) mostly as a hint rather than an indication of the functional form characterizing the relationship between knowledge and economic structure. The empirical analysis will be indeed carried out by adopting a somewhat less restricted approach which will be based on the application of vector autoregression (VAR) models, which we will describe in the next section.

4 The Variables, Methodology and Data

4.1 The Variables

The three properties of the KB which we will use in our analysis are the knowledge capital stock, its variety, related or unrelated, its coherence (and the complementary cognitive distance: see the Appendix for the details).

The traditional **regional knowledge stock** is computed by applying the permanent inventory method to patent applications. We calculated it as the cumulated stock of past patent applications using a rate of obsolescence of 15% per annum: $E_{i,t} = \dot{h}_{i,t} + (1 - \delta)E_{i,t-1}$, where $\dot{h}_{i,t}$ is the flow of regional patent applications and δ is the rate of obsolescence.

The **variety** of a KB measures the extent of its diversification, with related variety measuring it at a lower level of aggregation and unrelated variety at a higher level of aggregation (Frenken et al, 2007). Technological variety can be measured by using the information entropy index. It was introduced by Shannon (1948) to measure the information

content of messages, and can be used as a distribution function in a number of circumstances (Theil, 1967, Frenken 2004). The use of information entropy to measure variety is based on the rise in the information content of systems as the number of their distinguishable components increases: a system with a large number of distinguishable components requires more information to be described than a system with a smaller number of distinguishable components.

The information entropy index has interesting features, like its decomposability into a between-group and within-group component, and the extension to multidimensional cases. According to the latter, one may calculate the variety of the actual combinations of technological classes in a given context (say a firm or a sector). The former property allows for the operationalization of the distinction between related and unrelated variety. One could say that related variety (within-group entropy) measures diversification at a local level, or within a technological class, while unrelated variety (between-group entropy) measures diversification at a more global level in a knowledge space. The important implication of this distinction is that while a growth in unrelated variety implies a rise in cognitive distance, a growth in related variety is compatible with a more incremental development and even a fall in cognitive distance.

The **coherence** of a KB measures the extent to which different types of knowledge can be combined. This is of a fundamental importance since the types of knowledge required by firms to create new products or services are not necessarily found within a discipline, but need to be combined to produce the desired output. The ability of firms to combine these different types of knowledge is not constant but can be expected to vary systematically during particular phases of the evolution of knowledge. For example, we can expect the ability of firms to combine different types of knowledge to fall as a completely new type of knowledge emerges at a discontinuity and to rise again as the new type of knowledge starts maturing. The coherence of the knowledge base can be calculated by adapting a procedure developed by Teece et al (1994) to measure the coherence in the output of a firm. The basic principle underlying the calculations is that the higher the frequency with which different technologies are used together by a firm the more coherent is its knowledge base. The calculation proceeds by first calculating the frequency of co-occurrence of each pair of technologies in the KB and then by averaging them over the whole firm, or sector in the present case (see Nesta Saviotti, 2005, 2006 and Krafft, Quatraro, Saviotti, 2009). In addition to coherence, we also investigate the relationship between the terms of shift-share and **cognitive distance**, which measures the extent of discontinuity involved in the emergence of a new type of knowledge. It is the

inverse of an index of similarity. This measure is of fundamental importance to be able to distinguish the effect of the emergence of a discontinuity from that of the subsequent period of normal or incremental development. There are many ways to calculate cognitive distances but we used the complement of the index of similarity proposed by Jaffe (1989).

4.2 Methodology

The main focus of this chapter is on the observation of the co-evolutionary dynamics between knowledge and economic structure. We have proposed in the previous section a synthetic representation of change in economic structure by introducing shift share analysis. For the sake of clarity, let us assign a symbol to each of the identified components:

$$\mu = \sum_j \frac{P_{j,t-1} \Delta S_j}{(Y/L)} \quad (15)$$

$$\pi = \sum_j \frac{\Delta P_j \Delta S_j}{\left(\frac{Y}{L}\right)} \quad (16)$$

$$\alpha = \sum_j \frac{S_{j,t-1} \Delta P_j}{(Y/L)} \quad (17)$$

In view of the complex and endogeneous nature of the relationships between the properties of knowledge and those of economic structure, we apply a VAR model.

The regression of interest is the following:

$$w_{i,t} = c + \beta_z w_{i,t-z} + \varepsilon_{i,t} \quad (18)$$

Where w_{it} is an $m \times 1$ vector of random variables for region i at time t , β is an $m \times [m \times z]$ matrix of slope coefficients that are to be estimated. In our particular case $m=9$ and corresponds to the vector $[\mu(i,t), \pi(i,t), \alpha(i,t), \text{growth of knowledge capital (i,t), coherence growth (i,t), growth of cognitive distance (i,t), variety growth (i,t), related variety growth (i,t), unrelated variety growth (i,t)}]$. ε is an $m \times 1$ vector of disturbances. Knowledge capital is obtained by applying a permanent inventory method approach, the same way as the previous chapter. The properties of knowledge structure are instead calculated following the procedure described in Section 5.2.

In line with previous studies, the measure of growth rates is based on the difference of the logarithms of the respective variables. Let $X_i(t)$ represent the absolute value of the variable in region i at time t . Define the normalized (log) value of the variable as:

$$x_i(t) = \log(X_i(t)) - \frac{1}{N} \sum_{i=1}^N \log(X_i(t)) \quad (19)$$

Where N is the number of regions. In what follows, growth rates are defined as the first difference of normalized (log) values according to:

$$g_i(t) = x_i(t) - x_i(t - 1) \quad (20)$$

In such a way, common macroeconomic shocks are already controlled because the growth rate distribution was normalized to zero for each variable in each region in each year.

Following a growing body of literature (Coad, 2010; Buerger, Broekel and Coad, 2012; Colombelli, Krafft and Quatraro, 2011), Equation (18) is estimated via ‘reduced form’ VARs, which do not impose any a priori causal structure on the relationships between the variables, and are therefore suitable for the purposes of this analysis. These reduced-form VARs effectively correspond to a series of m individual ordinary least squares (OLS).

However, previous studies have emphasized how the empirical distribution of the growth rates is closer to a Laplacian than to a Gaussian distribution (Bottazzi et al. 2007; Bottazzi and Secchi 2003; Castaldi and Dosi 2009). Such evidence suggests that standard regression estimators, like ordinary least squares (OLS), assuming Gaussian residuals may perform poorly if applied to these empirical frameworks. To cope with this, a viable and increasingly used alternative consists of implementing the least absolute deviation (LAD) techniques, which are based on the minimization of the absolute deviation from the median rather than the squares of the deviation from the mean.

It must be noted that we do not include any individual dummies in the analysis. Even though unobserved heterogeneity can have important effects on the estimation results, the inclusion of individual dummies along with lagged variables may engender some biases for fixed-effect estimation of dynamic panel-data models, a problem known as Nickell-bias. Some alternative approaches relate to the use of instrumental variable (IV) or GMM estimators (Blundell and Bond, 1998). The main problem with this lies in the difficulty to find out good instruments, which is particularly hard when dealing with growth rates. When instruments are weak, IV estimation of panel VAR thus leads to imprecise estimates. Binder et al. (2005) propose instead a panel VAR model including firm-specific effects, which is however based on the assumption of normally distributed errors, which is not the case for what concerns the growth rates of the variables used in our regressions.

Since we are dealing with growth rates, instead of levels, we can maintain that any region-specific component has been largely removed. Moreover, we follow the wide body of literature on the analysis of firms’ growth rates stating that the non-Gaussian nature of growth rate residuals are a far more important econometric problem deserving careful attention even in regional level analyses (Buerger, Broekel and Coad, 2012).

4.3 The Data

In order to implement the analysis outlined in the previous section we gather together two datasets. The shift-share analysis has been conducted by using the branch accounts of NUTS II European regions¹ provided by the Eurostat within the European System of Integrated Economic Accounts. As is well known, these data are available only since 1995, the year in which the Eurostat has implemented a standardized procedure to collect data from European countries, so as to build a coherent and homogeneous dataset. As a result we were able to calculate the μ , the π , and the α components for a subset of European regions on a time span ranging from 1995 to 2007. The properties of knowledge structure, i.e. coherence, cognitive distance and variety (based on the information entropy index) have instead been calculated by using patent information contained in the OECD REGPAT database which covers patent data that have been linked to regions utilizing the addresses of the applicants and inventors. The analysis has been conducted by adopting the inventor-based regionalization², and by using 4-digits technology codes.

We obviously merged the two sets of indicators on the basis of the NUTS II regional code and the year. We end up with an unbalanced panel of 227 regions observed on average on 8 years. The descriptive statistics for the whole sample are reported in Table 1, while Figure 1 shows instead the distributional properties of the variables under scrutiny, providing empirical support to their non-Gaussian distribution. In particular all the variables appears to follow a Laplace-like distribution, which makes or empirical strategy outlined in the previous section the best approach to the analysis.

>>> INSERT Figure 1 AND Table 1 ABOUT HERE <<<

¹ We acknowledge that the use of administrative regions to investigate the effects of knowledge creation represents only an approximation of the local dynamics underpinning such process. Indeed administrative borders are arbitrary, and therefore might not be representative of the spontaneous emergence of local interactions. It would be much better to investigate these dynamics by focusing on local systems of innovation. However, it is impossible to find out data at such a level of aggregation. Moreover, the identification of local systems involve the choice of indicators and threshold values according to which one can decide whether to unbundle or not local institutions. This choice is in turn arbitrary, and therefore it would not solve the problem, but it would only reproduce the issue at a different level. Thus we think that despite the unavoidable approximation, our analysis may provide useful information on the dynamics under scrutiny.

² The assignment of patent to regions on the basis of inventors' addresses is the most widespread practice in the literature (see for example Maurseth and Verspagen, 2002; Henderson et al., 2005; Breschi and Lissoni, 2009, Paci and Usai, 2009, to quote a few). A viable alternative may rest on the use of applicants' addresses, above all when the assessment of knowledge impact on growth is at stake (see Antonelli, Krafft and Quatraro, 2010). However, when the analysis is conducted at local level of aggregation, and the geography of collective processes of knowledge creation is emphasized, the choice of inventors' addresses remains the best one.

The elaboration of a regional breakdown of descriptive statistics turns out to be very much complicated when dealing with a sample of 227 regions. For this reason we decided to show the cross-regional distribution of average values by implementing a map for each of the variables under consideration. In Figure 2 we report the cross-regional distribution of the three components contributing to labour productivity growth. Let us recall that μ is the contribution of the changing mix of regional industries, and is positive if regions tend to specialize in high-productivity activities, π is the interaction between productivity growth and the change in the industry mix, and is positive to the extent that regions specialize in fast growing sectors, while α is the contribution of within-sector productivity growth weighted by the sector share on total employment.

>>> INSERT Figure 2 ABOUT HERE <<<

It is interesting to note that for most of sampled regions the effect of change in the industry mix is positive, suggesting that structural change plays an important role in the process of economic growth. Most of European regions tend therefore to specialize in high-productivity sectors, with the only exception of some Greek regions and in the British midlands. The process is more pronounced in Italy and in central-eastern Europe than in Spain and France. The second diagram shows that the interaction term is positive again in most of Italian regions, Spain, France and Germany, while the evidence is more mixed in the other regions. Italy, France, Spain and Germany in the observed period are subject to changes favoring the increasing share of fast-growing sectors. Finally, the within-sector productivity growth seems to matter the most for Northern regions, like Finland, Sweden and Denmark, and at a somewhat lesser extent for some Eastern and Greek regions.

In Figure3 we report instead the cross regional distribution of knowledge capital, coherence and cognitive distance (log values). The top diagram reports the figures concerning the knowledge capital. We can notice how knowledge capital is higher in central European regions and in northern regions, while it is lower in the periphery of the continent. A look at the coherence index reveals that on average search behaviors are more like organized search than random screening, while cognitive distance is on average very low in most of the European regions, suggesting that exploration is conducted across the safe boundaries of established knowledge competences. Only for a few scattered regions in France, Spain and Finland we observe both low values of coherence and of cognitive distance, suggesting a search strategies characterized by exploration behaviors conducted within well defined boundaries of the knowledge space.

>>> INSERT Figure 3 ABOUT HERE <<<

In Figure4 we show the cross distribution of the variety index, articulated in unrelated and related knowledge variety. The top diagram indicates that on average European regions are characterized by a high degree of variety, with the only exception of some peripheral regions in Portugal and in Greece. When we look at the distinction between related and unrelated variety we notice that the distribution looks very similar to that of total variety. By observing also the ranges assigned to each classes, we can also emphasize that on average related knowledge variety is higher than unrelated variety.

>>> INSERT Figure 4 ABOUT HERE <<<

The maps reported in Figure 3 and in Figure 4 are based on absolute (log) values of the properties of the knowledge structure. In the following section we will implement the estimation of equation (18), which is based instead on the normalized growth rates of such variables.

5 Econometric results

The results of the ‘reduced-form’ VAR are reported in Table 3, which should be read as follows. Each column corresponds to each of the dependent variables in the model. Thus in column (1) the dependent variable is the normalized growth of μ , in colomun (2) that of π , and so on and so forth. The rows indicate instead the explanatory variables, which are grouped by lag (three lags are included). At the end of table we report also the number of observations and the R-squared for each regression.

With respect to the observed autocorrelation, it is impressive to note that none of the variables under scrutiny shows any degree of persistence. On the contrary, coefficients are negative and significant across all the three lags considered, suggesting erratic growth dynamics for all the variables. Such results on knowledge-related variables are consistent with the findings of Buerger et al (2011), who ascribe this kind of evidence to the intrinsic uncertainty and volatility characterizing innovation. Evidently, even though we attempted to counterbalance such volatility by letting each paper last 5 years, this has been not enough.

>>> INSERT Table 3 ABOUT HERE <<<

We now move to analyze in more detail the lead-lag relationship between the change in knowledge and in economic structure. As far as the first lag is concerned, knowledge coherence and knowledge capital show a positive and significant coefficient on α , which is

consistent with the largest part of the literature linking knowledge and productivity growth. The α component stands indeed for the contribution stemming from within-sector productivity growth, which is positively affected by the growth of knowledge coherence and that of knowledge capital. Cognitive distance is instead negatively linked to π . The search across dispersed area of the knowledge landscape is therefore likely to jeopardize the increase in the share of fast growing sectors. The knowledge variety indexes do not seem to affect significantly the economic structure. Unrelated variety appears instead to be positively affected by π , suggesting the increasing share of fast growing sectors is likely to favor the introduction of further variety in the innovation system. It is also interesting to note that α affects negatively the growth of coherence and cognitive distance. This evidence is in line with previous work (Colombelli, Krafft and Quatraro, 2011), according to which higher performances are likely to create the economic conditions to stimulate exploration activities, although in domains that are not too far from the established technological competences.

When we move to the second lag, we see that knowledge coherence affects positively and significantly π , i.e. faster growth of coherence is associated with the faster increase of faster growing sectors. Cognitive distance is again negatively related to π , while the related and unrelated variety indexes are instead both related positively to π . This evidence is quite puzzling, as by definition when related variety rises, unrelated variety decreases. However the twin positive coefficients can be interpreted in the light of the mixed nature of the π component, whereby related variety positively affects productivity growth, while unrelated variety positively affects the change in the industry mix. For what concerns the effects of the economic structure on knowledge structure, the μ component does not yield any significant effect on the knowledge characteristics. The π component instead affects positively coherence and negatively cognitive distance: the increasing share of faster growing sectors stimulates the establishment of exploitation activities dominated by organized search strategies within the comfortable fences of established competences. Once again, α negatively affects knowledge coherence and cognitive distance, like in the one-lag coefficient.

Finally, the third lag presents an interesting negative and significant coefficient on the effect of knowledge coherence on μ , which suggests that the decrease of knowledge coherence, which signals the undertaking of exploration activities, is likely to engender a reallocation effect of labor force across sectors, i.e. to foster the change in economic structure. The effect on π is again positive, signaling the prevalence of the positive effects on productivity dynamics. The coefficient of cognitive distance on π is instead negative and significant, which coupled with the positive one of coherence, suggests that exploitation strategies based on

organized search are likely to engender the movement towards faster growing activities. For what concerns the effects of economic on knowledge structure, both π and μ yield negative and significant effects on knowledge variety, and in particular on related variety. Thus it would seem that increasing variety foster the changing allocation of labor across sectors, but that this in turn is likely to be followed by a reduction in variety. The convergence towards faster growing sectors is also followed by a sharp decrease of cognitive distance. The within-sector productivity dynamics do not seem to hold significant effects on knowledge structure.

6 Conclusions

In this paper we have conducted an exploratory analysis of the co-evolutionary patterns of knowledge and economic structure. Drawing upon a theoretical framework which stresses the dynamic nature of the interactions between these two components as well as the endogenous character of their change process, we decided to implement an empirical framework based on the indicators proposed in Section 5.2 in order to characterize the architecture of knowledge structure. We have coupled such methodological approach with the shift-share technique, which allow grasping in a synthetic way the effects of the change in economic structure, and in particular we focused on the changing allocation of labor force across sectors.

The empirical analysis, given the dynamic effects feeding back from economic and knowledge structure and vice versa, has been conducted by implementing a set of ‘reduced-form’ VARs, which allowed us to investigate the lead-lag relationships between the two systems, without imposing any aprioristic causal structure.

The results of the analysis are encouraging and call for further research in this direction, showing a clear interactive pattern between the two structures. Changes in knowledge structure that signal the undertaking of exploitation strategies based on organized screening are likely to engender increasing within-sector productivity growth, while exploration strategies are likely to be followed by the changing allocation of labor force across sectors. We also noted how the increasing share of faster growing sectors stimulates the establishment of exploitation activities dominated by organized search strategies within the comfortable fences of established competences. Moreover, the implementation of VAR(3) allowed us to appreciate also some interesting dynamics, like the one relating variety and μ ,

which can be according to which increasing variety foster the changing allocation of labor across sectors, but that this in turn is likely to be followed by a reduction in variety.

These results are obviously somewhat preliminary and do not pretend to have a final word on the relationship between knowledge and economic structure. We think however that they are interesting both with respect to the mechanisms on which they shed new light and with respect to the identification of new methodological approaches to address these issues.

7 Appendix – The Implementation of Knowledge Indicators

The adoption of these variables marks an important step forward in the operational translation of knowledge creation processes. In particular, they allow for a better appreciation of the collective dimension of knowledge dynamics. Knowledge is indeed viewed as the outcome of a combinatorial activity in which intentional and unintentional exchange among innovating agents provides the access to external knowledge inputs (Fleming and et al., 2007). The network dynamics of innovating agents provide the basis for the emergence of new technological knowledge, which is in turn represented as an organic structure, characterized by elementary units and by the connections amongst them. The use of such variables implies therefore a mapping between technology as an act and technology as an artefact (Arthur, 2009; Lane et al., 2009; Krafft and Quatraro, 2011). Co-occurrences matrixes are very similar to design structure matrixes (DSM) (Baldwin and Clark, 2000; Murmann and Frenken, 2006; Baldwin, 2007), in that they can be thought as adjacency matrixes in which we are interested not only in the link between the elements, but also by the frequency with which such links are observed.

In other words these measures capture the design complexity of knowledge structure, and allow for featuring the innovation behaviour of firms, as well as its evolution, in relation with the changing architecture of such structure (Henderson and Clark, 1990; Murmann and Frenken, 2006). In this perspective, an increase in knowledge coherence is likely to signal the adoption of an exploitation strategy, while a decrease is linked to exploration strategies. Increasing values of cognitive distance are instead related to random screening across the technology landscape, while decreasing cognitive distance is more likely to be linked to organized search behaviour. Knowledge variety is likely to increase in any case when new combinations are introduced in the system. However the balance between related and unrelated variety should be such that the related one is likely to dominate during exploitation phases, while the unrelated one gains more weight in the exploration strategies (Krafft, Quatraro, Saviotti, 2009).

7.1 Variety

Let us start by the variety indicator, which we decided to measure by using the information entropy index. Entropy measures the degree of disorder or randomness of the

system, so that systems characterized by high entropy will also be characterized by a high degree of uncertainty (Saviotti, 1988). Differently from common measures of variety and concentration, the information entropy has some interesting properties (Frenken and Nuvolari, 2004). An important feature of the entropy measure is its multidimensional extension. Consider a pair of events (X_l, Y_j) , and the probability of co-occurrence of both of them p_{lj} . A two dimensional total variety (TV) measure can be expressed as follows:

$$TV \equiv H(X, Y) = \sum_l \sum_j p_{lj} \log_2 \left(\frac{1}{p_{lj}} \right) \quad (A.1)$$

If one considers p_{lj} to be the probability that two technological classes l and j co-occur within the same patent, then the measure of multidimensional entropy focuses on the variety of co-occurrences of technological classes within regional patents applications.

Moreover, the total index can be decomposed in a “within” and a “between” part anytime the events to be investigated can be aggregated into a smaller numbers of subsets. Within-entropy measures the average degree of disorder or variety within the subsets, while between-entropy focuses on the subsets measuring the variety across them. Frenken et al. (2007) refer to between- and within- group entropy respectively as unrelated and related variety.

It can be easily shown that the decomposition theorem holds also for the multidimensional case. Hence if one allows $l \in S_g$ and $j \in S_z$ ($g = 1, \dots, G$; $z = 1, \dots, Z$), we can rewrite $H(X, Y)$ as follows:

$$TV = H_Q + \sum_{g=1}^G \sum_{z=1}^Z P_{gz} H_{gz} \quad (A.2)$$

Where the first term of the right-hand-side is the between-entropy and the second term is the (weighted) within-entropy. In particular:

$$H_Q = - \sum_{g=1}^G \sum_{z=1}^Z P_{gz} \log_2 P_{gz} \quad (A.3)$$

$$H_{gz} = - \sum_{l \in S_g} \sum_{j \in S_z} p_{lj} \log_2 p_{lj} \quad (A.4)$$

$$P_{gz} = \sum_{l \in S_g} \sum_{j \in S_z} P_{lj}$$

$$H_g = - \sum_{l \in S_g} \sum_{j \in S_z} P_{lj} \log \frac{1}{P_{lj}}$$

We can therefore refer to between- and within-entropy respectively as *unrelated technological variety (UTV)* and *related technological variety (RTV)*, while total information entropy is referred to as *general technological variety*.

7.2 Knowledge similarity and dissimilarity (cognitive distance)

We need a measure of cognitive distance (Nooteboom, 2000) able to express the dissimilarities amongst different types of knowledge. A useful index of distance can be derived from the measure of *technological proximity*. Originally proposed by Jaffe (1986 and 1989), who investigated the proximity of firms' technological portfolios. Subsequently Breschi et al. (2003) adapted the index in order to measure the proximity, or relatedness, between two technologies. The idea is that each firm is characterized by a vector V of the k technologies that occur in its patents. Knowledge similarity can first be calculated for a pair of technologies l and j as the angular separation or un-centred correlation of the vectors V_{lk} and V_{jk} . The similarity of technologies l and j can then be defined as follows:

$$S_{lj} = \frac{\sum_{k=1}^n V_{lk} V_{jk}}{\sqrt{\sum_{k=1}^n V_{lk}^2} \sqrt{\sum_{k=1}^n V_{jk}^2}} \quad (\text{A.5})$$

The idea underlying the calculation of this index is that two technologies j and l are similar to the extent that they co-occur with a third technology k . The cognitive distance between j and l is the complement of their index of the similarity:

$$d_{lj} = 1 - S_{lj} \quad (\text{A.6})$$

Once the index is calculated for all possible pairs, it needs to be aggregated at the industry level to obtain a synthetic index of technological distance. This can be done in two steps. First of all one can compute the weighted average distance of technology l , i.e. the average distance of l from all other technologies.

$$WAR_{it} = \frac{\sum_{j \neq i} d_{ijt} P_{jt}}{\sum_{j \neq i} P_{jt}} \quad (A.7)$$

Where P_j is the number of patents in which the technology j is observed. Now the average cognitive distance at time t is obtained as follows:

$$CD_{it} = \frac{1}{P_{it}} \sum_{j \neq i} WAR_{ijt} P_{jt} \quad (A.8)$$

7.3 Knowledge coherence

Cognitive distance measures the degree of dissimilarity among technologies. We expect it to provide us with an indication of the difficulty, or cost, a firm has to face to learn a new type of knowledge. Typically a firm needs to combine, or integrate, many different pieces of knowledge to produce a marketable output. Thus, in order to be competitive a firm not only needs to learn new 'external' knowledge but it needs to learn to combine it with other, new and old, pieces of knowledge. We can say that a knowledge base in which different pieces of knowledge are well combined, or integrated, is a coherent knowledge base. The technologies contained in the knowledge base are by definition complementary in that they are jointly required to obtain a given outcome. For this reason, we turned to calculate the coherence of the knowledge base, defined as the average relatedness of any technology randomly chosen within the sector with respect to any other technology (Nesta and Saviotti, 2005 and 2006; Nesta, 2008).

To yield the knowledge coherence index, a number of steps are required. In what follows we will describe how to obtain the index at whatever level of analysis i . First of all, one should calculate the weighted average relatedness WAR_j of technology j with respect to all other technologies present within the sector. Such a measure builds upon the measure of technological relatedness τ_{jm} (see below). Following Teece et al. (1994), WAR_j is defined as the degree to which technology j is related to all other technologies $j \neq m$ in the aggregate, weighted by patent count P_{mt} :

$$WAR_{jit} = \frac{\sum_{m \neq j} \tau_{jm} P_{mit}}{\sum_{m \neq j} P_{mit}} \quad (\text{A.9})$$

Finally the coherence of knowledge base within the aggregate i (be it a firm, a sector or a region) is defined as weighted average of the WAR_{it} measure:

$$R_{it} = \sum_{j \neq m} WAR_{jit} \times \frac{P_{jit}}{\sum_j P_{jit}} \quad (\text{A.10})$$

It is worth stressing that such index implemented by analysing co-occurrences of technological classes within patent applications, measures the degree to which the services rendered by the co-occurring technologies are complementary to one another. The relatedness measure τ_{jm} indicates indeed that the utilization of technology j implies that of technology m in order to perform specific functions that are not reducible to their independent use. This makes the coherence index appropriate for the purposes of this study.

In order to calculate the parameter τ , i.e. *technological relatedness*, we start by calculating the relatedness matrix (Nesta, 2008). The technological universe consists of k patent applications. Let $P_{jk} = 1$ if the patent k is assigned the technology j [$j = 1, \dots, n$], and 0 otherwise. The total number of patents assigned to technology j is $O_j = \sum_k P_{jk}$. Similarly, the total number of patents assigned to technology m is $O_m = \sum_k P_{mk}$. Since two technologies may occur within the same patent, $O_j \cap O_m \neq \emptyset$, and thus the observed the number of observed co-occurrences of technologies j and m is $J_{jm} = \sum_k P_{jk} P_{mk}$. Applying this relationship to all possible pairs, we yield a square matrix Ω ($n \times n$) whose generic cell is the observed number of co-occurrences:

$$\Omega = \begin{bmatrix} J_{11} & & J_{j1} & & J_{n1} \\ \vdots & \ddots & & & \vdots \\ J_{1m} & & J_{jm} & & J_{nm} \\ \vdots & & & \ddots & \vdots \\ J_{1n} & \cdots & J_{jn} & \cdots & J_{nn} \end{bmatrix} \quad (\text{A.11})$$

We assume that the number x_{jm} of patents assigned to both technologies j and m is a hypergeometric random variable of mean and variance:

$$\mu_{jm} = E(X_{jm} = x) = \frac{O_j O_m}{K} \quad (\text{A.11})$$

$$\sigma_{jm}^2 = \mu_{jm} \left(\frac{K - O_j}{K} \right) \left(\frac{K - O_m}{K - 1} \right) \quad (\text{A.12})$$

If the observed number of co-occurrences J_{jm} is larger than the expected number of random co-occurrences μ_{jm} , then the two technologies are closely related: the fact the two technologies occur together in the number of patents x_{jm} is not casual. The measure of relatedness hence is given by the difference between the observed number and the expected number of co-occurrences, weighted by their standard deviation:

$$\tau_{jm} = \frac{J_{jm} - \mu_{jm}}{\sigma_{jm}} \quad (\text{A.13})$$

It is worth noting that such relatedness measure has lower and upper bounds: $\tau_{jm} \in]-\infty; +\infty[$. Moreover, the index shows a distribution similar to a t-student, so that if $\tau_{jm} \in]-1.96; +1.96[$, one can safely accept the null hypothesis of non-relatedness of the two technologies j and m . The technological relatedness matrix Ω' may hence be thought about as a weighting scheme to evaluate the technological portfolio of regions.

8 References

- Antonelli, C., 2008. *Localized Technological Change: Towards the Economics of Complexity*. London: Routledge.
- Antonelli, C., Krafft, J. and Quatraro, F., 2010. Recombinant knowledge and growth: The case of ICTs. *Structural Change and Economic Dynamics* 21, 50-69.
- Antonelli, C. and Quatraro, F., 2012. Localized Technological Change and Efficiency Wages: the Evidence across European Regions. *Regional Studies*, forthcoming.
- Arthur, W.B., 2009. *The Nature of Technology*. Free Press, New York.
- Baldwin, C.Y., 2007. Where do transactions come from? Modularity, transactions and the boundaries of firms. *Industrial and Corporate Change* 17, 155-195.
- Baldwin, C. Y. and Clark, K.B., 2000. *Design Rules, Volume I, The power of Modularity*. Cambridge MA, MIT Press.
- Binder, M., C. Hsiao and C. H. Pesaran, 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795–837.
- Blundell R.W and Bond S.R., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.
- Boschma, R. 2005. Proximity and innovation: A critical assessment. *Regional Studies* 39: 61-74.
- Boschma, R. and Frenken, K., 2006. Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography* 6, 273-302.
- Boschma, R. and Frenken, K., 2011. Technological relatedness and regional branching. In: Bathelt, H., Feldman M.P. and Kogler, D.F. (eds) *Dynamic geographies of knowledge creation and innovation*. London and New York, Routledge.
- Boschma, R. and Iammarino, S., 2009. Related variety, trade linkages, and regional growth in Italy. *Economic Geography* 85, 289-311.
- Bottazzi, G., Cefis, E., Dosi, G. and Secchi, A., 2007. Invariances and Diversities in the Patterns of Industrial Evolution: Some Evidence from Italian Manufacturing Industries. *Small Business Economics* 29, 137-159.
- Bottazzi, G., Secchi, A., 2003. Common properties and sectoral specificities in the dynamics of U.S. manufacturing companies. *Review of Industrial Organization* 23, 217–232.
- Breschi, S. and Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 94, 439-468.
- Breschi, S., Lissoni, F., and Malerba, F., 2003. Knowledge relatedness in firm technological diversification, *Research Policy* 32: 69-97.
- Buerger, M., Broekel, T. and Coad, A., 2012. Regional dynamics of innovation: Investigating the co-evolution of patents, research and development (R&D), and employment. *Regional Studies* 46, 566-582.
- Burns A. F., 1934. *Production trends in the United States since 1870*. NBER, Boston.
- Castaldi, C. and Dosi, G., 2009. The patterns of output growth of firms and countries: Scale invariances and scale specificities. *Empirical Economics* 37, 475-495.

- Coad, A., 2010. Exploring the processes of firm growth: evidence from a vector auto-regression. *Industrial and Corporate Change* 19, 1677-1703.
- Colombelli, A., Krafft, J. and Quatraro, F., 2011. High Growth Firms and Technological Knowledge: Do gazelles follow exploration or exploitation strategies? LEI-BRICK Working Paper no 14/2011.
- Creamer, D., 1942. *Shifts of Manufacturing Industries, in Industrial Location and National Resources*. Washington, D. C.: U. S. National Resources Planning Board.
- Esteban, J.M., 1972. Shift-share analysis revisited. *Regional and Urban Economics* 2, 249-61.
- Esteban, J.M., 2000. Regional convergence in Europe and the industry mix: a shift-share analysis. *Regional Science and Urban Economics* 30, 353-364.
- Fagerberg, J., 2000. Technological Progress, Structural Change and Productivity Growth: A Comparative Study. *Structural Change and Economic Dynamics* 11, 393-411.
- Fisher, A.G.B., 1939. Production, primary, secondary and tertiary. *The Economic Record* 15, 24-38.
- Fleming, L., 2001. Recombinant uncertainty in technological search. *Management Science* 47, 117-132.
- Fleming, L., Sorenson, O. 2001. Technology as a complex adaptive system: Evidence from patent data. *Research Policy* 30: 1019-1039.
- Frenken, K. 2004. Entropy and information theory. In: Hanusch, H, and Pyka, A. (eds.), *The Elgar Companion to Neo-Schumpeterian Economics*. Cheltenham, Edward Elgar, pp. 544-555.
- Frenken, K., Nuvolari, A., 2004. Entropy Statistics as a Framework to Analyse Technological Evolution. In: J. Foster and W. Hözl (eds) *Applied evolutionary economics and complex systems*. Cheltenham, U.K. and Northampton, Mass.: Edward Elgar.
- Frenken, K., von Oort, F., Verburg, T., 2007. Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies* 41: 685-97.
- Henderson, R.M and Clark, K.B, 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* 35, 9-30.
- Henderson, R. M., Jaffe, A. and Trajtenberg. M., 2005. patent citations and the geography of knowledge spillovers: A reassessment: Comment. *American Economic Review* 95, 416-464.
- Houston, G.B., 1967. The Shift and Share Analysis of Regional Growth: A Critique. *Southern Economic Journal* 33, 577-581.
- Krafft, J. and Quatraro, F., 2011. The Dynamics of Technological Knowledge: From Linearity to Recombination. In: Antonelli, C. (ed) *Handbook on the Economic Complexity of Technological Change*. Cheltenham: Edward Elgar, pp. 181-200.
- Krafft, J., Quatraro, F. and Saviotti, P.P., 2009. Evolution of the knowledge base in knowledge intensive sectors. LEI-BRICK Working Paper no 06/2009.
- Jaffe, A.B. 1986. Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. *American Economic Review* 76, 984-1001.
- Jaffe, A., 1989. Real Effects of Academic Research. *American Economic Review* 79: 957-70.

- Kuznets S., 1930. *Secular Movements in Production and Prices*. Houghton Mifflin, Boston.
- Kuznets, S., 1972. Innovation and Adjustments in Economic Growth, *Swedish Journal of Economics*, 74, 431-451.
- Lane, D.A., van Der Leeuw, S.E., Pumain, D., West, G. (eds.), 2009, *Complexity perspectives in innovation and social change*. Springer, Berlin.
- Lawson C. and Lorenz, E., 1999. Collective learning, tacit knowledge and regional innovative capacity. *Regional Studies* 33, 305-317.
- Marshall, A., 1919. *Industry and Trade. A Study of industrial technique and business organization; and of their influences on the condition of various classes and nations*. London: Macmillan and Co., Ltd.
- Maurseth, P. B. and Verspagen, B., 2002. Knowledge spillovers in Europe: A patent citations analysis. *Scandinavian Journal of Economics* 104, 531-45.
- Murmann, J.P. and Frenken, K., 2006. Towards a systematic framework for research on dominant designs, technological innovations, and industrial change. *Research Policy* 35, 925-952.
- Nelson, R.R. and Winter, S., 1982, *An Evolutionary Theory of Economic Change*, Cambridge, Harvard University Press.
- Nesta, L., 2008. Knowledge and productivity in the world's largest manufacturing corporations. *Journal of Economic Behavior and Organization* 67: 886-902.
- Nesta, L., and Saviotti, P.P., 2005. Coherence of the Knowledge Base and the Firm's Innovative Performance: Evidence from the U.S. Pharmaceutical Industry, *Journal of Industrial Economics* 53: 123-42.
- Nesta, L., and Saviotti, P.P., 2006. Firm Knowledge and Market Value in Biotechnology. *Industrial and Corporate Change* 15: 625-52.
- Nooteboom, B., 2000. *Learning and innovation in organizations and economies*. Oxford: Oxford University Press.
- Paci, R. and Usai, S., 2009. Knowledge flows across European regions. *The Annals of Regional Science* 43, 669-690.
- Perloff, H. S., Dunn, E. S., Lampard, E. E. And Muth, R. F., 1960. *Regions, Resources and Economic Growth*. Baltimore, Maryland: Johns Hopkins Press.
- Perroux F., 1955. Note sur la notion de 'pole de croissance'. *Économie Appliquée* 7, 307-320.
- Quatraro, F., 2009. Innovation, structural change and productivity growth. Evidence from Italian regions, 1980-2003. *Cambridge Journal of Economics* 33, 1001-1022.
- Quatraro, F., 2010. Knowledge Coherence, Variety and Productivity Growth: Manufacturing Evidence from Italian Regions, *Research Policy*, 39, 1289-1302.
- Quatraro, F., 2012. *The economics of structural change in knowledge*. London and New York, Routledge.
- Romijn H. and Albu, M., 2002. Innovation, networking and proximity: Lessons from small high technology firms in the UK. *Regional Studies* 36, 81-86.
- Saviotti, P. P., 1988. Information, variety and entropy in technoeconomic development. *Research Policy* 17: 89-103.

- Saviotti, P.P., 2004. Considerations about the production and utilization of knowledge, *Journal of Institutional and Theoretical Economics* 160: 100-121.
- Saviotti, P.P., 2007. On the dynamics of generation and utilisation of knowledge: The local character of knowledge, *Structural Change and Economic Dynamics* 18: 387-408.
- Schumpeter, J., 1912, *The Theory of Economic Development*. Harvard University Press, Cambridge.
- Schumpeter, J. A., 1928. The Instability of Capitalism. *Economic Journal*, 38, 361-386.
- Schumpeter, J. A., 1939. *Business Cycles. A Theoretical, Historical and Statistical Analysis of the Capitalist Process*. New York and London, McGraw Hill.
- Schumpeter. 1942. *Capitalism, Socialism and Democracy*, London, Unwin.
- Shannon, C.E., 1948. A mathematical theory of communication, *Bell System Technical Journal* 27: 379-423.
- Sorenson, O., Rivkin, J.W. and Fleeming, L., 2006. Complexity, network and knowledge flows. *Research Policy* 35, 994-1017.
- Teece, D.J., Rumelt, R., Dosi G. and Winter, S., 1994. Understanding Corporate Coherence: Theory and Evidence. *Journal of Economic Behavior and Organisation* 23: 1-30.
- Theil, H., 1967. *Economics and Information Theory*. Amsterdam: North Holland.
- Thomas M. J., 1975. Growth pole theory, technological change and regional economic growth. *Papers in Regional Science* 34, 3-25.
- Usher, A., 1954. *A History of Mechanical Invention*. Dover, Cambridge, MA.

Figure 1 - Distribution of the 9 relevant variables describing knowledge and economic structure.

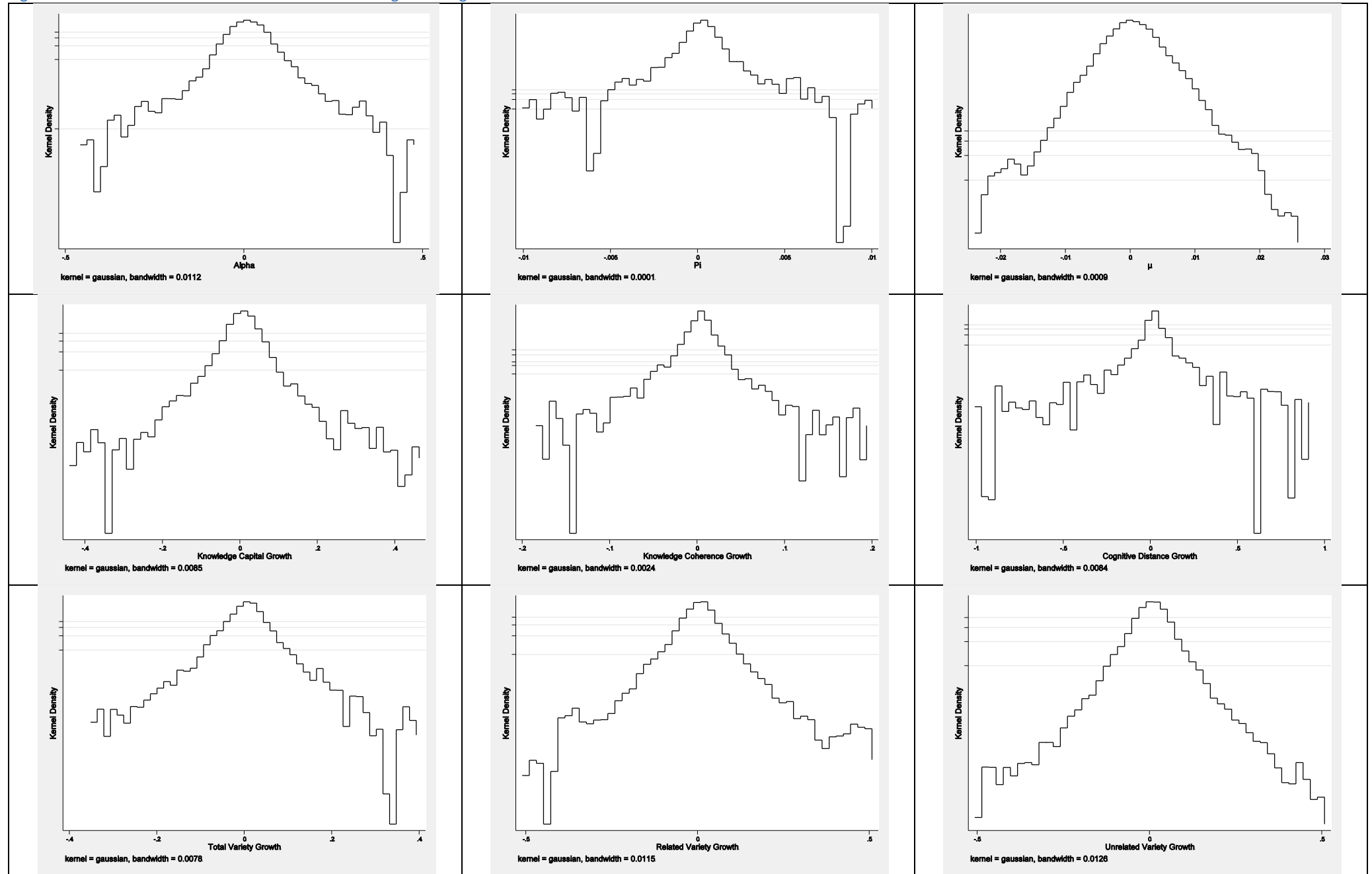


Figure 2 – Distribution of the three components of shift-share decomposition

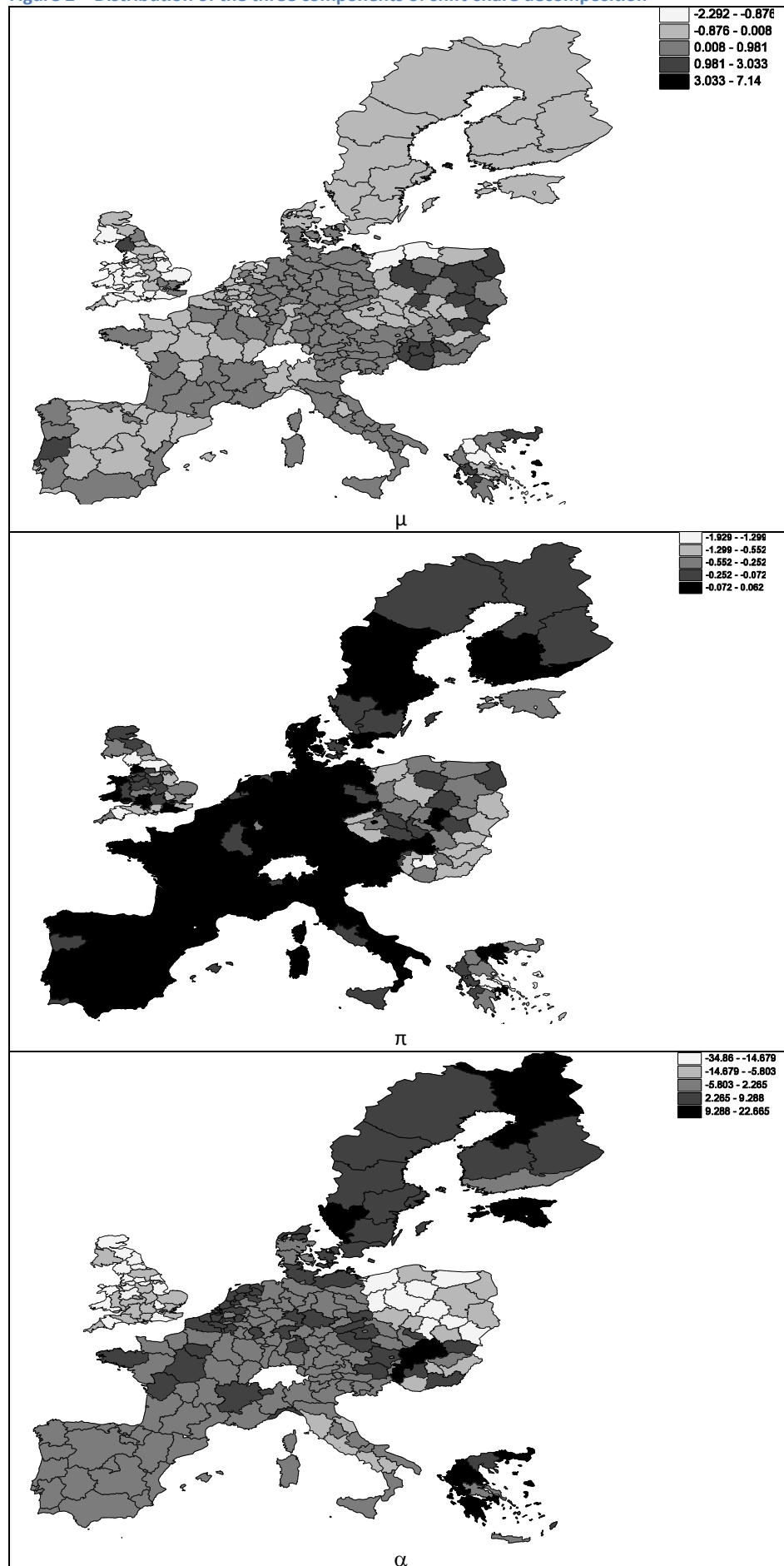


Figure 3 – Distribution of the properties of knowledge structure (I)

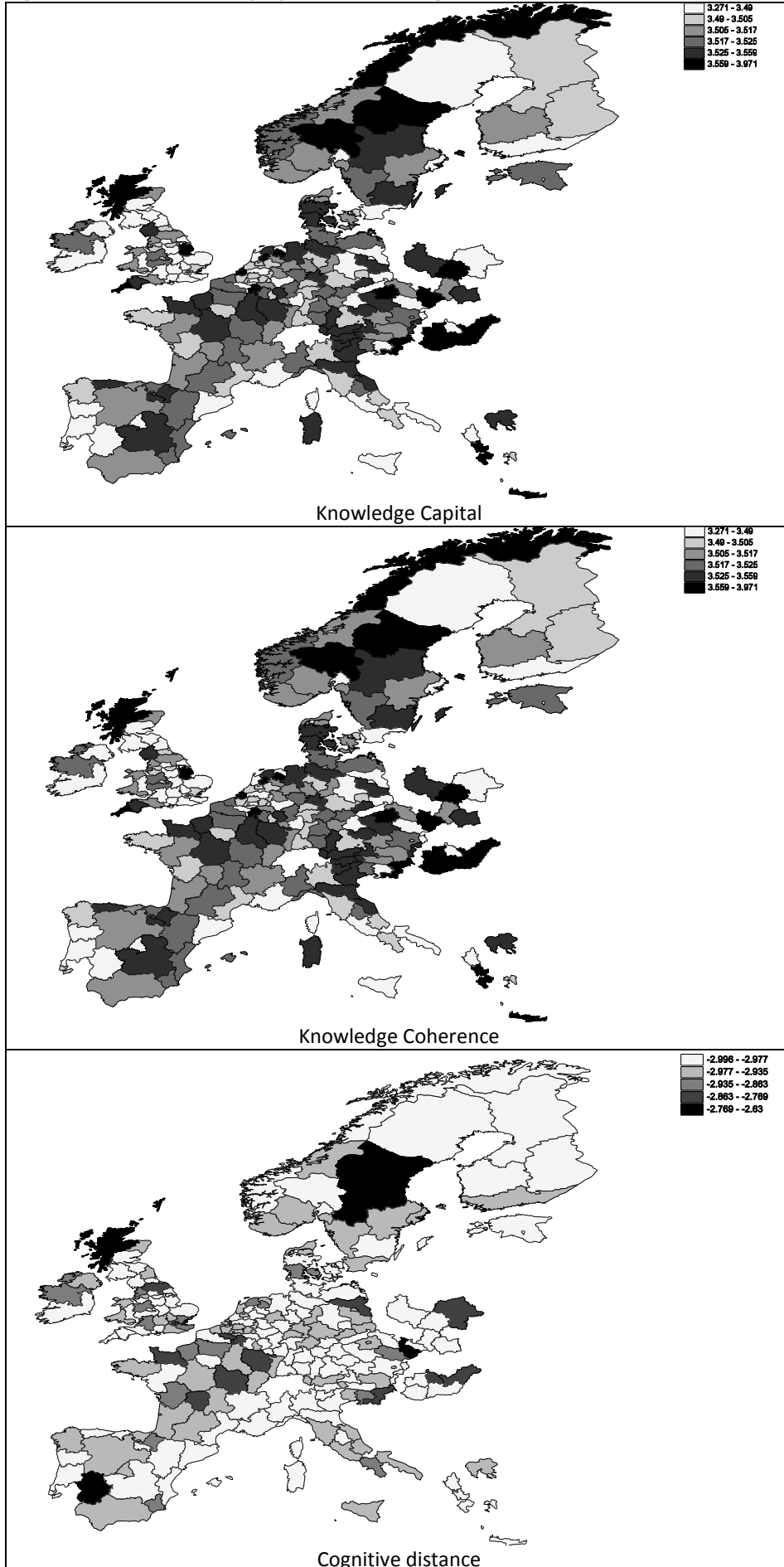


Figure 4 – Distribution of the properties of knowledge structure (II)

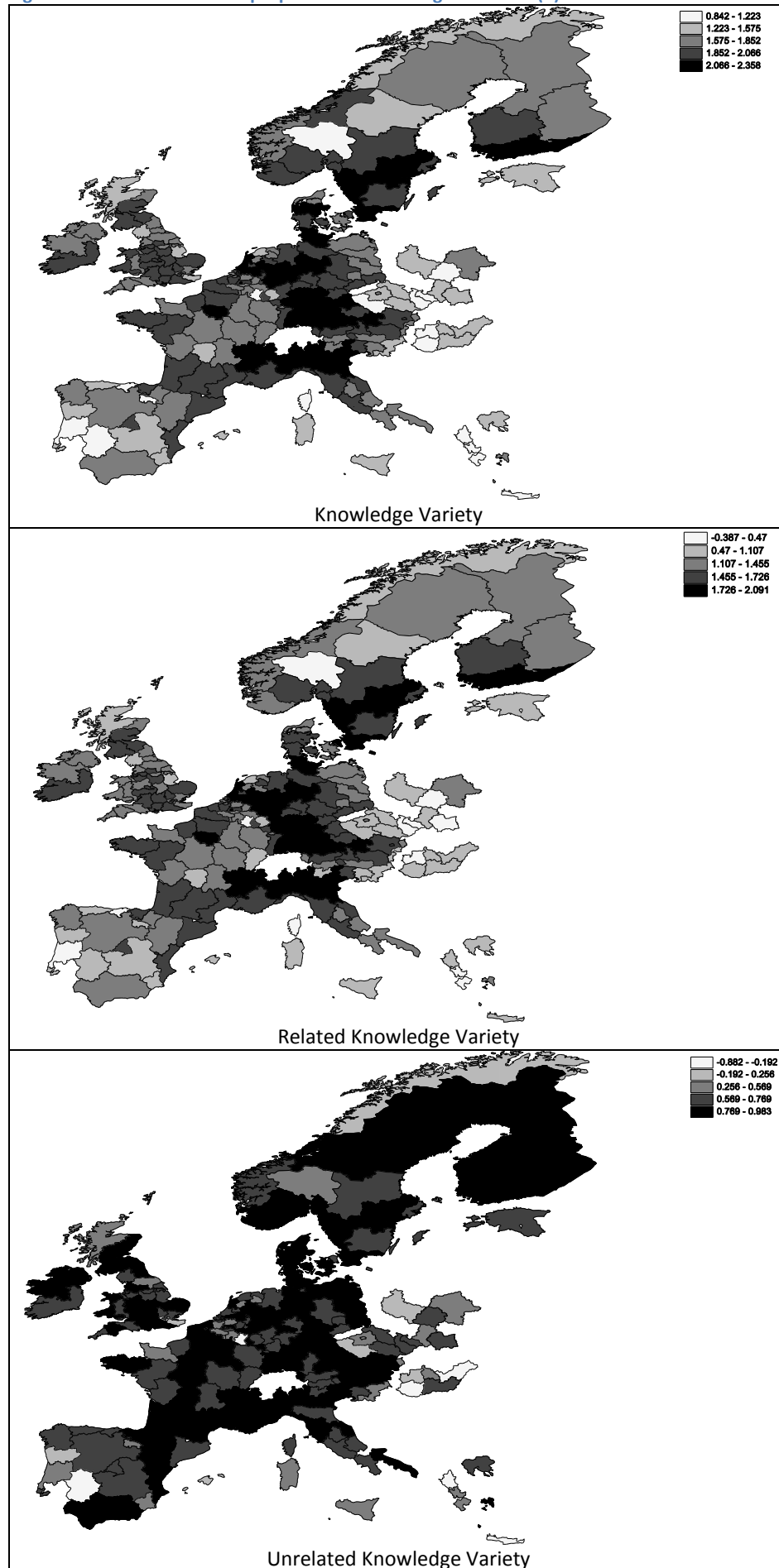


Table 1 – Descriptive statistics of the 9 variables before normalization

Variable		Mean	Std. Dev.	Min	Max	Obs.
μ	overall	0.003	0.006	-0.030	0.060	N = 1876
	between		0.005	-0.030	0.033	n = 227
	within		0.005	-0.028	0.042	
π	overall	-0.001	0.002	-0.032	0.002	N = 1873
	between		0.002	-0.022	0.001	n = 227
	within		0.001	-0.023	0.008	
α	overall	0.024	0.068	-0.349	0.630	N = 1873
	between		0.044	-0.217	0.173	n = 227
	within		0.061	-0.361	0.507	
Knowledge Capital	overall	0.048	0.109	-0.163	1.253	N = 1711
	between		0.102	-0.101	0.644	n = 202
	within		0.090	-0.659	0.756	
Knowledge Coherence	overall	-0.004	0.032	-0.644	0.578	N = 1744
	between		0.046	-0.159	0.578	n = 205
	within		0.026	-0.489	0.239	
Cognitive Distance	overall	0.001	0.115	-1.099	1.099	N = 1744
	between		0.056	-0.409	0.549	n = 205
	within		0.111	-1.098	1.107	
Knowledge Variety	overall	0.003	0.064	-0.333	0.568	N = 1744
	between		0.069	-0.145	0.494	n = 205
	within		0.052	-0.294	0.393	
Related Knowledge Variety	overall	0.004	0.113	-0.601	1.615	N = 1744
	between		0.115	-0.601	0.726	n = 205
	within		0.095	-0.664	0.892	
Unrelated Knowledge Variety	overall	0.005	0.103	-1.163	1.125	N = 1742
	between		0.094	-0.519	0.518	n = 205
	within		0.090	-0.728	0.998	

Table 2 - Correlation matrix

	Coherence	Knowledge Capital	Cognitive distance	TV	RTV	UTV	μ	π	α
Coherence	1.000								
Kn. Capital	-0.027 0.244	1.000							
Cog. Distance	-0.018 0.434	0.075* 0.001	1.000						
TV	-0.034 0.133	0.043 0.062	-0.115* 0.000	1.000					
RTV	0.009 0.693	-0.007 0.749	-0.128* 0.000	0.727 0.000	1.000				
UTV	-0.110* 0.000	0.010 0.661	-0.004 0.857	0.234 0.000	-0.331* 0.000	1.000			
μ	-0.026 0.312	0.042 0.104	0.026 0.316	0.019 0.471	-0.049 0.056	0.010 0.709	1.000		
π	-0.024 0.348	-0.046 0.075	-0.048 0.062	-0.088 0.001	0.091* 0.000	-0.044 0.085	-0.201* 0.000	1.000	
α	0.027 0.291	-0.029 0.274	-0.016 0.529	0.006 0.810	-0.058* 0.025	0.024 0.363	-0.019 0.432	-0.129 0.000	1.000

Note: Variables are expressed in normalized growth rates. Pairwise correlation coefficients. The stars signal a significance of at most 5%.

Table 3 – Results of ‘reduced-form’ VAR estimation of Equation (18)

VARIABLES		(1) mu	(2) pi	(3) alfa	(4) Koh	(5) CD	(6) Kap	(7) TV	(8) RTV	(9) UTV
β_{t-1}	Koh	-0.012 -0.008	0.000 -0.002	0.357*** -0.082	-0.887*** -0.017	0.163*** -0.050	-0.116 -0.115	-0.052 -0.100	-0.327*** -0.117	0.048 -0.146
	Kap	-0.002 -0.002	0.000 0.000	0.0343** -0.015	0.00798*** -0.003	-0.014 -0.009	-0.549*** -0.022	0.009 -0.018	0.0493** -0.022	-0.006 -0.028
	CD	0.001 -0.001	-0.000462* 0.000	0.004 -0.014	-0.00577** -0.003	-1.198*** -0.008	0.005 -0.017	0.0379** -0.016	0.0442** -0.019	-0.0487** -0.024
	TV	0.001 -0.012	-0.002 -0.002	0.064 -0.121	0.019 -0.024	-0.570*** -0.073	0.275 -0.170	-0.369** -0.146	0.277 -0.169	0.408** -0.207
	RTV	-0.001 -0.007	0.001 -0.001	-0.051 -0.073	-0.005 -0.015	0.329*** -0.044	-0.167 -0.103	0.007 -0.089	-0.603*** -0.102	-0.256** -0.124
	UTV	0.000 -0.004	0.001 -0.001	0.000 -0.037	0.000 -0.007	0.175*** -0.023	-0.141*** -0.054	0.009 -0.046	-0.007 -0.051	-0.656*** -0.066
	Mu	-0.618*** -0.031	-0.001 -0.006	-0.494 -0.324	-0.091 -0.062	-0.062 -0.187	-0.052 -0.430	0.013 -0.366	-0.114 -0.447	0.134 -0.552
	Pi	-0.548*** -0.147	-0.691*** -0.029	-1.222 -1.523	0.321 -0.289	-0.248 -0.877	-3.765* -2.071	1.115 -1.736	2.074 -2.023	5.892** -2.558
	alfa	0.00492** -0.002	0.00156*** 0.000	-0.724*** -0.026	-0.00930* -0.005	-0.0435*** -0.015	0.043 -0.033	0.038 -0.029	0.010 -0.035	0.059 -0.044
	Koh	-0.001 -0.008	0.00328** -0.001	0.101 -0.078	-0.526*** -0.019	0.106* -0.055	-0.091 -0.120	0.226** -0.110	-0.123 -0.132	0.629*** -0.156
	Kap	-0.001 -0.002	0.000 0.000	0.020 -0.016	0.00858*** -0.003	0.000 -0.010	-0.382*** -0.023	-0.011 -0.019	0.033 -0.024	-0.010 -0.029
	CD	0.001 -0.002	-0.000781** 0.000	-0.001 -0.017	-0.00593* -0.003	-0.823*** -0.010	0.002 -0.023	0.026 -0.019	0.025 -0.024	-0.0740** -0.030
β_{t-2}	TV	0.003 -0.013	-0.00606** -0.003	-0.003 -0.132	-0.0950*** -0.027	0.137 -0.084	0.817*** -0.187	-0.028 -0.165	0.110 -0.189	0.188 -0.230
	RTV	-0.001 -0.008	0.00324** -0.002	-0.035 -0.079	0.0707*** -0.016	-0.062 -0.051	-0.521*** -0.113	-0.076 -0.100	-0.321*** -0.114	-0.078 -0.136
	UTV	0.001 -0.004	0.00164** -0.001	0.010 -0.038	0.0258*** -0.008	-0.016 -0.026	-0.315*** -0.056	-0.035 -0.051	0.024 -0.056	-0.247*** -0.072
	Mu	-0.327*** -0.034	-0.0210*** -0.007	0.403 -0.354	0.019 -0.068	-0.305 -0.205	0.364 -0.476	-0.561 -0.400	-0.554 -0.492	0.752 -0.601
	Pi	0.476*** -0.172	-0.469*** -0.035	-1.884 -1.805	0.979*** -0.355	-3.110*** -1.062	-5.204** -2.463	-3.913* -2.090	-1.289 -2.529	0.568 -3.138
	alfa	0.003 -0.003	0.00171*** -0.001	-0.543*** -0.028	-0.0110** -0.005	-0.0321** -0.016	0.027 -0.037	0.050 -0.032	0.003 -0.038	0.0781* -0.047
	Koh	-0.00952* -0.005	0.00217*** -0.001	0.048 -0.052	-0.213*** -0.014	0.011 -0.040	0.015 -0.089	0.105 -0.082	-0.200** -0.098	0.339*** -0.116
	Kap	-0.002 -0.001	0.000 0.000	0.019 -0.014	-0.001 -0.003	0.003 -0.008	-0.0918*** -0.019	-0.021 -0.016	-0.018 -0.019	-0.024 -0.025
	CD	-0.001 -0.001	-0.000581** 0.000	0.0346** -0.014	-0.003 -0.003	-0.399*** -0.008	-0.004 -0.018	0.0523*** -0.015	0.0823*** -0.019	-0.0504** -0.023
	TV	0.004 -0.011	-0.002 -0.002	0.334*** -0.119	-0.0770*** -0.024	0.183** -0.073	0.663*** -0.164	-0.865*** -0.144	-1.090*** -0.175	-0.015 -0.213
	RTV	0.003 -0.006	0.000 -0.001	-0.221*** -0.072	0.0671*** -0.014	-0.116*** -0.044	-0.439*** -0.099	0.347*** -0.087	0.316*** -0.105	0.031 -0.127
	UTV	0.002 -0.003	0.000 -0.001	-0.107*** -0.036	0.0272*** -0.007	-0.023 -0.023	-0.207*** -0.048	0.144*** -0.044	0.320*** -0.054	-0.309*** -0.065
	mu	-0.0970*** -0.0970***	-0.0161*** -0.0161***	-0.099 -0.099	0.023 0.023	-0.222 -0.222	0.145 0.145	-0.617* -0.617*	-0.788* -0.788*	0.253 0.253

VARIABLES	(1) mu	(2) pi	(3) alfa	(4) Koh	(5) CD	(6) Kap	(7) TV	(8) RTV	(9) UTV
	-0.028	-0.006	-0.291	-0.056	-0.169	-0.387	-0.335	-0.404	-0.497
pi	0.382**	-0.292***	2.362	0.092	-4.403***	-2.540	-3.320*	-4.620**	2.027
	-0.151	-0.031	-1.570	-0.305	-0.931	-2.126	-1.829	-2.183	-2.738
alfa	0.0113***	0.00120***	-0.245***	-0.001	-0.014	-0.003	-0.001	-0.008	0.015
	-0.002	0.000	-0.021	-0.004	-0.012	-0.028	-0.024	-0.028	-0.035
Constant	0.000	0.000218***	-0.00323**	0.00184***	0.001	0.001	0.000	0.002	-0.002
	0.000	0.000	-0.001	0.000	-0.001	-0.002	-0.002	-0.002	-0.002
R ²	0.178	0.209	0.265	0.301	0.378	0.170	0.195	0.204	0.225
Observations	935	935	935	926	926	922	926	926	926