Persistence of innovation and knowledge structure: Evidence from a sample of Italian firms[1]

Alessandra Colombelli[a,b,c] and Francesco Quatraro[b,c]

a) DISPEA, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino
b) University of Nice Sophia Antipolis, GREDEG-CNRS, 250 rue Albert Einstein, 06560 Valbonne
c) BRICK, Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri (Torino)

Preliminary Draft

(this version 1/23/2012)

ABSTRACT.

The paper investigates the patterns of persistence of knowledge and innovation across a sample of Italian firms in the period 1998-2006. While most of the empirical analysis of persistence focus on innovation as proxied by capital stock or the introduction of process/product innovation, this paper proposes to investigate the persistence of knowledge structure. The analysis draws upon a theoretical representation of knowledge as a collective good, stemming from the recombination of knowledge bits that are fragmented and dispersed across economic agents. In this perspective, knowledge is a co-relational structure, whereby the constituting elements are connected in a dense network of interlinks. On this basis, we derived some important properties of knowledge structure, i.e. the coherence, the cognitive distance and the variety, and investigated their pattern of persistence over time. The empirical analysis is implemented by exploring the autocorrelation structure of such properties within a quantile regression framework, and compare them with the more familiar evidence about knowledge stock. The results suggest that the properties of knowledge are featured by somewhat different patterns with respect to knowledge stock, and that such evidence is also heterogeneous across firms in different quantiles.

JEL Classification Codes: L20, L10, 032

Keywords: Persistence, Innovation, Knowledge Structure, Quantile regression, autocorrelation

[1] The authors acknowledge the financial support of the European Union D.G. Research with the Grant number 266959 to the research project 'Policy Incentives for the Creation of Knowledge: Methods and Evidence' (PICK-ME), within the context Cooperation Program / Theme 8 / Socio-economic Sciences and Humanities (SSH) and of the Collegio Carlo Alberto with the framework of the research project IPER. We wish to thank Second Rolfo for making possible the access to the Bureau Van Dijk AIDA dataset at the CERIS-CNR.

1 Introduction

The issue of innovation persistence has attracted increasing attention in the last decades, following the seminal contribution by Geroski et al. (1997). The idea that innovation activities are featured to some extent by dynamic increasing returns made possible by learning and creative accumulation has shaped this stream of literature. Success breeds success, in line with a kind of Matthew effect that is more likely to reward firms that have better performed in the past. In this sense the phenomenon of persistence keeps retaining a particular appeal even when the Schumpeterian underpinnings are grafted onto the more recent complexity-based economic thinking. Persistence can indeed be though as a particular kind of preferential attachment mechanism, according to which only a few agents show relentless introduction of innovation over time.

The traditional approaches to persistence of innovation instead do not take in great consideration the systemic dynamics of innovation. The explicit acknowledgement of the collective and systemic nature of technological knowledge allows to making some step forwards in the appreciation of the intrinsic heterogeneity of technological knowledge. Knowledge is not an unbundled stock. It is indeed a composite asset, which stems from interactions among a wide variety of agents who socialize their own knowledge and combine it with other knowledge inputs present in the environment they operate.

In this paper we propose to extend the analysis of persistence of innovation by introducing the concept of knowledge structure (Quatraro, 2012). The combinatorial activity at the basis of knowledge creation processes allows indeed for a conceptual representation of knowledge structure as a web of interconnected elements. This opens up different methodological avenues to the implementation of synthetic indicators to describe the changing structure of knowledge bases at different levels of aggregation.

The analysis conducted in this paper is based on the analysis of the co-occurrence of technological classes within patent documents, which allow to provide operational translation to concepts like knowledge variety, coherence and cognitive distance. These can be thought as peculiar properties of the structure of knowledge. The investigation of persistence patterns (or of their absence) is carried out on a sample of Italian firms observed between 1998 and 2006. We analyze the serial autocorrelation of growth rates of such properties, looking at the first three lags. We also implement quantile regression analyses to see if persistence patterns changes across the distribution of sampled firms. The results suggest that, while innovation shows a great deal of persistence, the properties of knowledge structure are more likely to be characterized by negative autocorrelation, which is to say alternation of high-growth and slowdown periods. This evidence is even more marked for those firms characterized by dramatically low or exceptionally high growth rates.

The rest of the paper is organized as follows. Section 2 provides a short review of the literature on persistence and proposes its grafting onto a complexity-based to technological knowledge which leads to the introduction of the concept of knowledge structure. Section 3 presents the data and discusses the methodology. In Section 4 we show and discuss the empirical results of econometric estimations. Finally Section 5 provides some preliminary conclusions and avenue for further research.

2 Persistence of Innovation and Knowledge Structure

The persistence of innovation activities has been the object of the analysis of a large body of literature in the last decade, both from a theoretical and empirical viewpoint. The main theoretical underpinnings lie in the concepts of cumulativeness and technological learning. According to neo-Schumpeterians, knowledge accumulation and technological learning account for the main forces leading to innovation persistence. Schumpeter himself distinguished between two different patterns of innovations (Schumpeter, 1912 and 1942). On the one hand, in the 'creative destruction' dynamics knowledge is conceived as a free good and, thus, all the firms can fish in the same pool of accessible technologies. As a consequence, new innovators introduce new technology while old innovators rest stuck in old innovation. On the other hand, the pattern of 'creative accumulation' emphasizes the cumulative nature of technological change. Knowledge is created and accumulated within firms. This builds high barriers to entry and, as a consequence, established large firms become key actors in the process of technological change. Within this framework success breeds success, current innovation is explained by past innovation and, thus, innovation is persistent (Alfranca, Rama and von Tunzelmann, 2002).

In evolutionary theory, the persistence of innovation activities stems from competition and selection mechanisms. In this view, the accumulation of knowledge and learning dynamics lead to the formation of firm-specific routines that may generate a stable pattern of economic activities. Yet, the inertia stemming from routines can be counteracted by dynamic forces like technological competition and innovation that push the economic system towards evolution (Nelson and Winter's, 1982). As a consequence, firms that survive to the market competition are those that persistently implement new techniques and introduce new ideas, which, in turn, increase their profitability and market share. Thus, the selection mechanism that pushes firms to persistently rely on innovation is a function of their internal competencies, technological capability and profitability.

A recent strand of literature has tried to empirically analyze the persistence of innovation. It is possible to distinguish two main lines of research in this area. A first set of studies aims at analyzing the persistence in the introduction of innovation trying to understand whether innovators have a stronger probability than non-innovators to keep innovating. In particular, these empirical works focus on the determinants and the features of the persistency by observing firms' patenting activity over time (Geroski, Van Reenen and Walters 1997, Malerba et al. 1997, Cefis and Orsenigo 2001, Cefis 2003, Alfranca, Rama and von Tunzelmann 2002) or the introduction of product and process innovation as revealed by innovation surveys repeated over time (Peters 2009, Raymond et al. 2006, Roper and Hewitt-Dundas 2008). These works, explicitly or implicitly, are based on the dynamic capabilities theory (Teece and Pisano 1994) and refer to the idea that technical change builds upon accumulated competencies and that new knowledge are generated by what has been learned in the past. A second set of studies examines persistency in the effects of innovation rather than the persistence of innovation per se (Cefis and Ciccarelli 2005, Latham and Le Bas 2006). These works build upon the idea that the stream of profits generated by past innovation gives firms the opportunity to keep on innovating and confirm that the impact of innovation on performance is cumulative and long lasting (Antonelli, Crespi, Scellato, 2012).

While the theoretical and empirical literature on the subject has much focused on the importance of internal technological capabilities and financial resources for the persistence of innovation, less attention has been paid to the collective and systemic nature of knowledge creation (Colombelli and von Tunzelmann, 2011). In particular, the focus on innovation as a simple count of the patents a firm has applied for, or as the count of product or process innovations introduced by firms, limits the scope of the analysis to the intensity of the innovation effort, but it does not say anything about the properties of the technological knowledge generated and the underpinning search strategies. The emphasis on patent counts, or on any kind of knowledge stock measure, has the undesirable disadvantage of implying a representation of technological knowledge as a homogeneous stock, as if it were the outcome of a quite uniform and fluid process of accumulation made possible by R&D investments, the same way as capital stock (Griliches, 1979; Mansfield, 1980). Such kind of representation is however hardly useful to investigate the nature of firms' search strategies, as it only allows for evaluating it from a quantitative rather than a qualitative viewpoint.

On the contrary there is a large consensus in the literature on the fact that knowledge technological knowledge can be depicted as an outcome of a collective undertaking strongly influenced by the availability of local sources of knowledge and by the quality of interactions (Allen, 1983; von Hippel 1988; Antonelli, 1999). The collective knowledge approach implies therefore the existence of agents characterized by bounded rationality, which cannot have the full command of the whole knowledge space, and therefore need to access knowledge dispersed and fragmented in the economic system in order to feed the combinatorial dynamics leading to the production of new knowledge (von Hayek, 1945).

In this perspective, more recently an increasingly share of scholars in the economics of innovation has elaborated theoretical approaches wherein the process of knowledge production is viewed as the outcome of a recombination process (Weitzmann, 1998; Kauffman, 1993). The creation of new knowledge is represented as a search process across a set of alternative components that can be combined one another. A crucial role is played here by the cognitive mechanisms underlying the search process aimed at exploring the knowledge space so as to identify the pieces that might possibly 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).

Based on these achievements, we can conveniently introduce the concept of knowledge structure (Quatraro, 2012). If new knowledge is the outcome of a combinatorial activity, one can imagine the space of human knowledge as a structure of relations that can be represented as a network the nodes of which are either variables or concepts and the links of which are the connections between different variables or concepts. Both the number of nodes and the number of links of such a knowledge network can be expected to change in the course of time as new concepts and variables are discovered and as new links are created between previously unconnected variables or concepts.

While the collective approach to the process of knowledge creation recalls the attention on the network of innovating agents, the adoption of a structure-based approach to knowledge turns out to be even more appropriate from a terminological viewpoint, as the idea of collectivism is much more related to a collection of agents rather than on the links amongst them. Structural holism is different from collective holism in that the whole is not just the juxtaposition of the individual elements, but is the outcome of the relationships among the components. The structuralism so conceived is consistent with the adoption of systemic thinking, according to which knowledge is an emergent property of a complex set of interactions at the agent level, which is in turn characterized by a complex set of interactions among knowledge inputs combined together, according to the principle of recursivity (Arthur, 2009; Lane, 2011)., By combining holism and individualism, the structuralist heuristic combines the interest in the relationships with the attention to the properties of the single components of the system (Bloch and Metcalfe, 2011).

Firms' knowledge bases can be accordingly understood as complex sets of interacting elements, say technological competences. Both the dynamics of the connections and the properties of the nodes provide useful information to qualify their combinatorial strategies. A synthetic, although partial, representation of the internal structure of the knowledge base can be built by drawing upon the frequency with which two technologies are combined together in the firms' knowledge base. Basically, this characterization takes into account 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 that lead us to derive three main properties of knowledge structure at a general level:

- Variety is related to the technological differentiation within the knowledge base, in particular with respect to the diverse possible combinations of pieces of knowledge in the sector, from the creation of a radically new type of knowledge to the more incremental recombination of already existing types of knowledge.
- Coherence can be defined as the extent to which the pieces of knowledge that agents within the sector combine to create new knowledge are complementary one another.
- Similarity (or dissimilarity) refers to the extent to which the pieces of knowledge used in the sector are close one another in the technology space.

The dynamics of technological knowledge can therefore be understood as the patterns of change in its own internal structure, i.e. in the patterns of recombination across the elements in the knowledge space. In other words, the investigation of such properties of the knowledge base provides an interesting implementation of the idea of architectural change (Henderson and Clark, 1990) applied to the analysis of knowledge dynamics. Moreover, this approach captures both the cumulative character of knowledge creation and the possible link to the relative stage of development of a technological trajectory (Dosi, 1982; Saviotti, 2004 and 2007; Krafft, Quatraro and Saviotti, 2009).

The persistence of knowledge structure takes a sharply different meaning from the persistence of innovation. It is indeed much more related with the ability by which firms can, or intend to, perseverate in pursuing a given search strategy, be it of 'exploration' or 'exploitation' (March, 1991). Our basic research question in this direction is to what extent firms knowledge bases are characterized by persistent integration or variety. By exploring the autocorrelation structure of the growth rates of knowledge properties, we investigated the degree to which these appear to be characterized by self-enforcing rather than cyclical dynamics. We also wonder whether some differential behaviors can be detected featuring high-growing (in terms of knowledge properties) from low-growing firms. For example, a firm showing a faster growth rate of internal coherence is one that explores the knowledge space in the domain of complementary technologies, profiting from the exploitation of the cumulated technological competences. The investigation of the persistence of coherence may therefore help understanding whether increasing coherence is a self-enforcing dynamics or it is preceded by former exploration activities in which many different alternatives are tried and eventually discarded by selecting only those with higher fitness values.

Similarly, we can gain better understanding of the dynamics by which firms develop their knowledge bases by increasing the average degree of similarity or amongst the technologies in their portfolios or their variety. In this perspective, we turn now to describe the data and the methodology we will use to provide an operational definition of the properties of knowledge structure and to investigate their persistence over time.

3 Data and Methodology

3.1 The dataset

The dataset used in this paper is an unbalanced panel of Italian firms that applied for a patent on the period 1998-2006. The dataset has been obtained by merging three sources of information. First of all, the PATSTAT database (April 2011) contains detailed information on worldwide patent applications to the European Patent Office[2]. This information is crucial to implement the properties of knowledge structure that will be described in what follows. Secondly we obtained information on Italian firms by the Bureau Van Dijk AIDA dataset. Finally, we used the harmonized matching tables described by Thoma et al. (2010) to combine the EPO and the AIDA datasets on the basis of the Bureau Van Dijk firm identification code.

Our final sample consists of 3,499 active firms having applied for more than a patent at the EPO in the period under scrutiny. Table 1 shows the distribution of firms across ISIC 4 macro sectors. Quite in line with expectations, the bulk of the sample operates within the manufacturing sector (about 77%). Besides manufacturing, ‘Wholesale and retail trade’ and ‘Professional, scientific and technical activities’ sectors also show relatively high shares, i.e. 8.66% and 4.58% respectively. Table 2 shows instead the size distribution of sampled firms[3]. Also in this case the evidence is hardly surprising, as most of the firms can be classified as small firms (about 39%). If one sums up the figures about the first three size classes, we obtain that about the 76% of the sampled firms operates at a scale that characterizes them as small-medium sized.

>>> INSERT Table 1 AND Table 2 ABOUT HERE <<<

[2] The limits of patent statistics as indicators of technological activities are well known. The main drawbacks can be summarized in their sector-specificity, the existence of non patentable innovations and the fact that they are not the only protecting tool. Moreover the propensity to patent tends to vary over time as a function of the cost of patenting, and it is more likely to feature large firms (Pavitt, 1985; Griliches, 1990). Nevertheless, previous studies highlighted the usefulness of patents as measures of production of new knowledge. Such studies show that patents represent very reliable proxies for knowledge and innovation, as compared to analyses drawing upon surveys directly investigating the dynamics of process and product innovation (Acs et al., 2002). Besides the debate about patents as an output rather than an input of innovation activities, empirical analyses showed that patents and R&D are dominated by a contemporaneous relationship, providing further support to the use of patents as a good proxy of technological activities (Hall et al., 1986).

[3] The total number of firms in Table 2 is slightly higher than that in Table 1 as the industrial classification field contains some missing values.

In Figure 1 we finally show the geographical distribution of firms. It is quite evident that the large majority of both firms and applicants is in Northern regions. The Lombardy region shows the highest number of innovating firms, followed by Piedmont, Emilia-Romagna and Veneto. Friuli Venezia-Giulia and two central regions, i.e. Lazio and Tuscany fall instead in the median class.

>>> INSERT Figure 1 ABOUT HERE <<<

The general picture that emerges from this preliminary exploration of the dataset is that the sample is mostly composed of small or medium-sized firms, active in the manufacturing sectors and located in the Northern regions of Italy. The following section will introduce the indicators proxying for the properties of the knowledge structure of the firms that are the object of our analysis.

3.2 The Implementation of Knowledge Indicators

For what concerns the definition of the knowledge related variables, let us start by the traditional firm's knowledge stock. This 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 patent applications and δ is the rate of obsolescence[4].

The implementation of knowledge characteristics proxying for variety, coherence and similarity, rests on the recombinant knowledge approach. In order to provide an operational translation of such variables one needs to identify both a proxy for the bits of knowledge and a proxy for the elements that make their structure. For example one could take scientific publications as a proxy for knowledge, and look either at keywords or at scientific classification (like the JEL code for economists) as a proxy for the constituting elements of the knowledge structure. Alternatively, one may consider patents as a proxy for knowledge, and then look at technological classes to which patents are assigned as the constituting elements of its structure, i.e. the nodes of the network representation of recombinant knowledge. In this paper we will follow this latter avenue[5]. Each technological class *j* is linked to another class *m* when the same patent is assigned to both of them. The higher is the number of patents jointly assigned to classes *j* and *m*, the stronger is this link. Since technological classes attributed to patents are reported in the patent document, we will refer to the link between *j* and *m* as the co-occurrence of both of them within the same patent document[6]. On this basis we calculated the following three key characteristics of firms' knowledge bases (see the appendix A for the methodological details):

a) Knowledge variety (KV) measures the degree of technological diversification of the knowledge base. It is based on the information entropy index, and it can be decomposed in related knowledge variety (RKV) and unrelated knowledge variety (UKV).

[4] Different depreciation rates have been implemented, which provided basically similar results.
[5] See Strumsky et al. (2012) for a compressive discussion on the use of patents technological classes to study technological change.
[6] It must be stressed that to compensate for intrinsic volatility of patenting behaviour, each patent application is made last five years.

b) Knowledge coherence (COH) measures the degree of complementarity among technologies. In the specific case coherence is meant to capture the average degree of complementarity among the technologies featuring firms' patents portfolios.
c) Cognitive distance (CD) expresses the dissimilarities amongst different types of knowledge. Similarly to the previous indicator, cognitive distance is meant to capture the average degree of similarity among the technologies featuring firms' patents portfolios.
d) We finally calculated knowledge diversification (KDIV) as the total number of technologies observables in firms' patents portfolios.

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).

3.3 Methodology

The empirical analysis of persistence has traditionally focused on innovation proxied by the application for a patent or the introduction of process/product innovations. The former explicit analysis by Geroski et al. (1997) adopts a proportional hazards model, while in subsequent works the most widespread methodology is the analysis of transition probability matrixes, which basically consists in assessing the probability that a firms innovate at time *t*, given its innovation performances at time *t-1* (Cefis and Orsenigo, 2001; Cefis, 2003; Antonelli et al. 2012).

A somewhat less covered methodology to tackle the issue of persistence consists in the analysis of serial autocorrelation of growth rates. Cefis and Ciccarelli (2005) provide in this respect a former implementation of such empirical strategy to analyze the differential in the persistence of profits between innovators and non-innovators. In the recent years, however, the analysis of the serial correlation of growth rates has gained momentum, for what concerns mainly firms performances in terms of sales or employment growth. These analyses are clearly cast in a persistence perspective, and emphasize the potentials of such methodology both in terms of predictive power and of policy design(Coad, 2007 and 2009; Coad and Hoelzl, 2009).

In this vein the application of serial autocorrelation analysis to the growth rates of knowledge structure properties may yield important results concerning the existence (or the absence) of persistence. In order to proceed with the analysis, we define growth rates of the relevant variables as follows:

$$Growth_{i,t} = \ln(X_{i,t}) - \ln(X_{i,t-1}) \tag{1}$$

Where X is measured in terms of knowledge capital stock; knowledge coherence; cognitive distance; knowledge variety; related knowledge variety; unrelated knowledge variety; knowledge diversification. All these variables, which have been introduced in the previous section and better explained in appendix A, are calculated for firm i at time t. Following previous empirical works (Bottazzi et al, 2011; Coad, 2010), the growth rates distributions have been normalized around zero in each year by removing means as follows:

$$\mathrm{s_{i,t}} = \mathrm{Growth_{i,t}} - \frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{Growth_{i,t}} \tag{2}$$

Where N stands for the total number of firms in the sample. This procedure effectively removes average time trends common to all the firms caused by factors such as inflation and business cycles.

The empirical model we will run in the analysis will take therefore the following form:

$$s_{i,t} = \alpha_0 + \sum_{k=1}^{K} \beta_k s_{i,t-k} + \epsilon_{i,t} \tag{3}$$

Figure 2 shows the distribution of firms' growth rates for all the relevant variables. As evidenced by the figure, the empirical distribution of the growth rates for our sample seems closer to a Laplacian than to a Gaussian distribution (with the only exception of knowledge capital). This is in line with previous studies analysing the distribution of firm growth rates (Bottazzi et al. 2007; Bottazzi and Secchi 2006; Castaldi and Dosi 2009).

>>>INSERT Figure 2 ABOUT HERE<<<

Such evidence suggests that standard regression estimators, like ordinary least squares (OLS), assuming Gaussian residuals may perform poorly if applied to these data. 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.

Descriptive statistics for the properties of knowledge structure are shown in Table 3. The reported variables are growth rates before normalization. On average we can observe that the sampled firms seem to be characterized by decreasing coherence, increasing cognitive distance and knowledge capital. The values across the percentiles also suggest that growth rates are characterized by highly skewed distributions.

>>>INSERT Table 3 AND Table 4 ABOUT HERE<<<

In Table 4 we show instead the matrix of correlations among the variables we use in the empirical exercise, marking a significance level of 5%. Although some significant pattern of correlation can be identified, these do not raise any sever concern, as the variables are not used together in the regression estimates.

4 Econometric Results and Discussion

The main focus of this paper is the analysis of the persistence (or the absence of it) of the properties of knowledge structure. The investigation of serial autocorrelation of growth rates allows us to appreciate the dynamics of such indicators, by revealing whether they are characterized by substantial self-enforcing mechanisms according to which firms building their technology portfolio around routinized exploitation activities (or random exploration) at some point in time are likely to perseverate in that direction, or they are rather likely to change direction in the course of time.

In Table 5 we report the results of the estimation of equation (3) carried out by implementing the LAD estimator. Due to the nature of the data, we limit ourselves to the analysis of the first three lags. The first column reports the evidence concerning the traditional measure of knowledge capital stock. We can notice how from this data knowledge production appears to be clearly persistent. The coefficient on the first lag is indeed positive and significant. The same applies also to the coefficient on the second lag, whose magnitude is anyway lower than the previous one. The third lag shows instead a negative and significant coefficient. These results would suggest that persistence in knowledge production is gradually achieved by firms.

>>> INSERT Table 5 ABOUT HERE <<<

Column (2) reports the results concerning knowledge coherence. Let us recall that knowledge coherence is a proxy of the degree to which the technologies that make the technological portfolio of firms are complementary to one another. In other words it is an indicator of the degree of integration of firms' knowledge base. The coefficient on the first lag is negative and significant, and the same applies to the coefficient on the second lag, although the magnitude is lower. The coefficient on the third lag is again negative, although not significant. This would suggest that at the overall level, increasing growth rates of coherence are preceded by (at least) two years of decreasing growth rate. This evidence is consistent with the idea that search behaviors characterized by the organized strategies of exploitation emerge out of an evolutionary process in which the preliminary phases involve the exploration, somewhat random, in many different directions in the knowledge space, unless a set of complementary technologies coherent with the established technological competences are selected.

Columns (3) to (5) report the results of estimations for what concern knowledge variety, and its components related and unrelated variety. The variety indexes we used, which are described in detail in the appendix, are based on the information entropy index, and in particular on its multi-dimensional extension. This means that variety here refers to the observed combination of technologies in firms' knowledge bases. This index provides therefore an idea of the extent to which firms try and experiment new combination. The results indicate that there is some degree of persistence in knowledge variety (KV), as revealed by the positive and significant coefficient on the second lag. The same applies also to related variety (RKV), which show a positive and significant coefficient also on the third lag. The strongest persistent dynamics characterize instead unrelated variety (UKV), which show positive and significant coefficients on all of the three lags. Thus it seems that pursuing variety in the past brings about new variety in the future, in terms of both related and unrelated components. The procedure by which the index is derived reveals that the concepts of 'related' and 'unrelated' variety refer basically to the belonging of technologies to the same technological domain, as defined by the classification system under utilization (in our case the International Patent Classification). This means that an increase in unrelated variety may signal an increase in combinations between technology from different technological domains, but that can have a high degree of complementarity or similarity.

Column (6) provides the results concerning the cognitive distance index. In this case only the negative coefficient on the first lag is significant. This would suggest the existence of a somewhat erratic dynamics of cognitive distance, which can be characterized by 'interruption to growth'. Column(7) finally provides the evidence about knowledge diversification (KDIV), which is characterized by negative and significant coefficients on all of the three lags. Increasing diversification at one moment in time appears therefore to follow decreasing diversification in at least three preceding years (or viceversa). Once again, this is consistent with the idea about the cyclical behavior of firms search strategies, according to which firms tend to smooth their diversification efforts once some profitable research avenues are identified, and eventually invigorate them when their research activities enter some decreasing returns phase.

4.1 Quantile regression analysis

In the preceding section we have investigated the serial autocorrelation of the growth rates of knowledge structure properties at the overall level. However, we can expect that firms in the dataset do not behave in the same way. To the purposes of this paper we are in particular interested in detecting possible differential behaviors for firms characterized by different growth rates of the variables under scrutiny. For example, are firms characterized by higher growth rates of coherence featured by peculiar autocorrelation patterns as compared to firms characterized by decreasing coherence?

The use of quantile regression techniques, first introduced by Koenker and Bassett (1978) can be of great help in this direction[7]. Besides of being robust to outliers and heavy-tailed distributions, the quantile regression methodologies are able to describe the entire conditional distribution of the dependent variable. Firms showing significanty higher or lower growth rates of the relevant variables are of particular relevance for the present study, and thus we can have a special focus on them by calculating coefficient estimates at various quantiles of the conditional distribution. Finally, it also worth recalling that this empirical approach avoids the restrictive assumption that the error terms are identically distributed at all points of the conditional distribution. This allows for accounting for

[7] More on quantile regressions can be found in Koenker and Hallok (2001).

firm heterogeneity and for considering the possibility that estimated slope parameters vary at different quantiles of the conditional growth rate distribution.

In Table 6 we report the results of the quantile estimation, while Figure 3 provides a summary representation (taking into account only the first lag). The coefficients can be interpreted as the partial derivative of the conditional quantile of the dependent variable with respect to particular regressors.

The results evaluated at the median are by definition the same as the those of the LAD estimation. If we look at first column, we can notice that the persistence of knowledge capital is robust across all of the percentiles identified. This means that both firms characterized by slow growth rates of knowledge capital and those characterized by high growth rates experience self-enforcing dynamics. The magnitude of the first lag coefficient increases as growth rates increase. Therefore, some ‘learning effect’ can be devised, according to which those firms more proactive in the generation of technological knowledge will experience a stronger boost on the future innovation performances.

>>> INSERT Table 6 and Figure 3 ABOUT HERE <<<

Column (2) shows the results concerning knowledge coherence. We can notice that the negative coefficient on the autocorrelation coefficients is common to all the quantiles. The uppermost quantile is however the only one characterized by a negative and significant coefficient also on the third lag. The diagram (b) of Figure 3 helps us to interpret these findings, supporting the idea that firms experiencing a particular dramatic fall in knowledge coherence have probably experienced above-average growth in the previous period, while firms showing particularly high growth rates of coherence have probably experienced poor performances in the past.

Columns (3) to (5) report the results concerning KV, RKV and UKV respectively. As far as KV is concerned, the first lag is positive and significant only for the 75% quantile, suggesting that firms in this class experience some persistence dynamics. Moreover, regression results for this area of the distribution also show positive and significant coefficients on the second and third lag. The results are somehow mixed in the other quantiles, whereby in the lowermost quantile the only significant (and negative) coefficient is on the third lag, while in the 25% quantile the coefficient on the second lag is positive and significant while that on the third lag is negative and significant. Also the evidence about RKV and UKV is rather mixed across quantiles, suggesting however that some degree of persistence mainly characterized the 10% and the 90% quantiles, i.e. those firms characterized either by particularly high of dramatically low growth rates of such indicartors.

Column (6) provides the results about cognitive distance. The negative and significant coefficient on the first lag is robust to all quantiles. However, by looking also at Figure 3, we can notice that firms at the 10% and those at the 90% quantiles are featured by relatively higher (in absolute values) coefficients. This is again consistent with the evidence we have found about knowledge coherence, and with the idea that firms are likely to introduce a discontinuity in their search behavior when they have achieved either a too much or a too little average degree of similarity in the technologies they can command.

5 Conclusions

This paper has proposed to extend the analysis of innovation persistence to the investigation of the dynamics of the properties of knowledge structure. This concept stems from a complexity-based approach to the analysis of technological knowledge, according to which this is an emergent property of systemic interactions among agents and is in turn characterized by a set of dynamic interactions among its constituting elements (Quatraro, 2012). To this purpose, we have proposed a set of indicators to synthetically describe the structure of knowledge, and analyzed the serial autocorrelation patterns of their growth rates.

The empirical results has provided support to the existence of persistent dynamics in innovation as measured by traditional proxies like knowledge capital stock. When the properties of knowledge structure are at stake, negative autocorrelation is mostly observed, suggesting that periods of growth are more likely to be followed by sharp decrease (or viceversa). The implementation of quantile regression techniques shows that such evidence is even more marked for firms in the 10% and in the 90% quantiles, i.e. those firms experiencing dramatically low or particularly high growth rates of knowledge properties. Firms characterized by significantly high growth rates of coherence or cognitive distance are more prone to direct their future search behavior towards strategies leading to the smoothing of such indicators. This is much consistent with the lifecycle interpretation of such properties (Krafft, Quatraro, Saviotti, 2009).

Such results are to be considered as preliminary, and some extensions would be particularly usueful. First of all, it would be useful to carry out such an analysis on a wider dataset, including firms from different countries as well as allowing from longer time series. Moreover, it would be also important to cluster the regressions according to different firms size classes, as well as to implement a temporal disaggregation of the sample. On the whole, the results provided in this paper open up a new avenue to the analysis of persistence, and allows for a better understanding of firms' search strategies, rather than simply observing whether a firm has introduced an innovation or not.

6 References

Acs et al., 2002, Patents and Innovation Counts as Measures of Regional Production of New Knowledge, *Research Policy*, 2002, 31(7), 1069-1085.

Alfranca, O., Rama, R., von Tunzelmann, N. (2002), A patent analysis of global food and beverage firms: The persistence of innovation, *Agribusiness* 18, 349 – 368.

Allen, R. C. (1983). Collective invention. *Journal of Economic Behavior and Organization* 4: 1-24.

Antonelli, C. (1999), *The Microdynamics of Technological Change*, Routledge, London.

Antonelli, C., Crespi, F., Scellato, G., (2012), Patterns of persistence in productivity growth. The Italian evidence, *Structural Change and Economic Dynamics*, forthcoming.

Arthur, W.B., 2009, *The Nature of Technology. What It Is and How It Evolves*. New York, Free Press.

Attaran, 1985, Industrial diversity and economic performance in U.S. areas. *The Annals of Regional Science* 20, pp. 44-54?

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.

Bloch, H. and Metcalfe, J.S., 2011. Complexity in the Theory of the Developing Firm. in *Handbook on the Economic Complexity of Technological Change,* Ed C. Antonelli, Cheltenham: Edward Elgar.

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., Coad, A., Jacoby, N. and Secchi, A. (2010). Corporate Growth and Industrial Dynamics: Evidence from French Manufacturing. *Applied Economics*, 43, 103-116.

Bottazzi, G. and Secchi, A., 2006, Explaining the Distribution of Firms Growth Rates, Rand Journal of Economics, 37, 234–263.

Breschi S., Lissoni, F. and Malerba, F., 2003, Knowledge relatedness in firm technological diversification, *Research Policy*, 32, 69-97.

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.

Cefis, E. (2003), Is there persistence in innovative activities? *International Journal of Industrial Organization* 21, 489-515.

Cefis, E., Ciccarelli, M. (2005), Profit differentials and innovation, *Economics of Innovation and New Technology* 14, 43-61.

Cefis, E., Orsenigo, L. (2001), The persistence of innovative activities. A cross-countries and cross-sectors comparative analysis, *Research Policy* 30, 1139-1158.

Coad, A., 2010, Exploring the processes of firm growth: evidence from a vector auto-regression, *Industrial and Corporate Change,* 19, 1677-1703.

Coad, A. (2009). *The Growth of Firms: a Survey of Theories and Empirical Evidence*. Edward Elgar, Cheltenham, UK and Northampton, MA, USA.

Coad, A. (2007), A Closer Look at Serial Growth Rate Correlation, Review of Industrial Organization, 31(1), 69–82.

Coad, A. and Hoelzl, W. (2009). On the autocorrelation of growth rates: Evidence for micro, small and large firms from the Austrian service industries, 1975-2004. *Journal of Industry Competition and Trade*, 139-166.

Colombelli A., von Tunzelmann G.N. (2011), The persistence of innovation and path dependence in Antonelli C. (eds.) *Handbook on the Economic Complexity of Technological Change,* Edward Elgar, Cheltenham, pp. 105-120.

Dosi, G., 1982, Technological paradigms and technological trajectories, *Research Policy*, 11, 147–162.

Fleming, L., 2001, Recombinant Uncertainty in Technological Search, *Management Science* 47(1), 117-132.

Fleming, L., Mingo, S. and Chen, D., 2007, Collaborative brokerage, generative creativity and creative success, *Administrative Science Quarterly*, 52, 443-475.

Frenken, K. and Nuvolari, A., 2004, Entropy Statistics as a Framework to Analyse Technological Evolution, in J. Foster and W. Hölzl (eds), *Applied Evolutionary Economics and Complex Systems*. Edward Elgar: Cheltenham, U.K. and Northampton, Mass.

Frenken, K., von Oort, F. and Verburg, T., 2007, Related Variety, Unrelated Variety and Regional Economic Growth, *Regional Studies*, 41(5), 685-97.

Geroski, P., Van Reenen, J., Walters, C. (1997), How persistently do firms innovate?, *Research Policy* 26, 33-48.

Griliches, Z., 1990, Patent statistics as economic indicators: a survey, *Journal of Economic Literature*, 28, 1661-1707.

Griliches, Z., 1979, Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*. 10 92-116.

Jaffe, A., 1986, Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value, *American Economic Review*, 76(5), 984-1001.

Jaffe, A., 1989, Real Effects of Academic Research, *American Economic Review*, 79(5), 957-70.

von Hayek, F.A., (1937). Economics and Knowledge, *Economica* 4: 33-54.

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.

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.

Krafft, J., Quatraro, F. and Saviotti, P.P., 2011, The knowledge base evolution in biotechnology: A social network analysis, *Economics of Innovation and New Technology*, 20, 445-477.

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.

Kauffman, 1993, *Origins of order: Self-Organization and selection in evolution*, Oxford University Press, Oxford.

Lane, D.A., 2011. Complexity in the Theory of the Developing Firm. in *Handbook on the Economic Complexity of Technological Change,* Ed C. Antonelli, Cheltenham: Edward Elgar.

Lane, D.A., van Der Leeuw, S.E., Pumain, D., West, G. (eds.), 2009, *Complexity perspectives in innovation and social change*. Springer, Berlin.

Latham, W.R., Le Bas, C. (2006), *The economics of persistent innovation: An evolutionary view*, Springer, Berlin.

Malerba, F., Orsenigo, L., Petretto, P. (1997), Persistence of innovative activities sectoral patters of innovation and international technological specialization, *International Journal of Industrial Organization* 15, 801-826.

Mansfield, E., 1980, Basic research and productivity increase in manufacturing, *American Economic Review*, 70, 863-73.

March, 1991, Exploration and exploitation in organizational learning, *Organization Science*, 2(1), 71-87.

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.

Nesta and Saviotti, 2005, Coherence of the Knowledge Base and the Firm's Innovative Performance: Evidence from the U.S. Pharmaceutical Industry. *Journal of Industrial Economics.* 53(1) 123-42.

Nesta, L. and Saviotti, P.P., 2006, . Firm Knowledge and Market Value in Biotechnology. *Industrial and Corporate Change*. 15(4) 625-52.

Nesta, L. 2008, Knowledge and productivity in the worlds largest manufacturing corporations. *Journal of Economic Behavior and Organization*. 67 886-902.

Nelson, R., Winter, S. (1982), *An Evolutionary Theory of Economic Change*, Harvard University Press, Cambridge, MA.

Nightingale, P., 1998, A cognitive model of innovation, *Research Policy*, 27, 689-709.

Nooteboom, B., 2000, *Learning and innovation in organizations and economies*, Oxford: Oxford University Press.

Pavitt, K., 1985, Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics* 7, 77-99.

Peters, B. (2009), Persistence of innovation: Stylized facts and panel data evidence, *The Journal of Technology Transfer* 34, 226-243.

Quatraro, F. (2012). *The Economics of Structural Change in Knowledge*. London and New York, Routledge.

Raymond, W., Mohnen, P., Palm, F.C., Schim Van Der Loeff, S. (2006), Persistence of innovation in Dutch manufacturing: Is it spurious? CESifo Working Paper Series No. 1681.

Roper, S., Hewitt-Dundas, N. (2008), Innovation persistence: Survey and case-study evidence, *Research Policy* 37, 149-162.

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.A., 1912, *The Theory of Economic Development*, Harvard University Press, Cambridge.

Schumpeter, J.A., 1942, *Capitalism, Socialism and Democracy*, Harper and Row, New York.

Teece, D., Pisano, G. (1994), The dynamic capabilities of firms: An introduction, *Industrial and Corporate Change* 3, 537-555.

Teece et al., 1994, Understanding Corporate Coherence: Theory and Evidence, *Journal of Economic Behavior and Organization*, 23(1), 1-30.

Theil, 1967, *Economics and Information Theory*. Amsterdam: North Holland, University Press: Oxford.

Von Hippel, E. 1988. *The sources of innovation*. Oxford: Oxford University Press.

Weitzmann, 1998, Recombinant growth, *Quarterly Journal of Economics,* 113, 331-360.

APPENDIX A – The properties of knowledge structure

Knowledge Variety

We decided to measure variety in firms' knowledge base 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).

Such index was introduced to economic analysis by Theil (1967). Its earlier applications aimed at measuring the diversity degree of industrial activity (or of a sample of firms within an industry) against a uniform distribution of economic activities in all sectors, or among firms (Attaran, 1985; Frenken et al., 2007; Boschma and Iammarino, 2009).

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, which we will exploit in our analysis, is its multidimensional extension. Consider a pair of events (X_j, Y_m), and the probability of co-occurrence of both of them p_{jm}. A two dimensional (total) entropy measure can be expressed as follows (firm and time subscripts are omitted for the sake of clarity):

$$H(X,Y) = \sum_{j=1}^{q} \sum_{m=1}^{w} p_{jm} \log_2 \left(\frac{1}{p_{jm}} \right) \qquad \text{(A1)}$$

If one considers p_{jm} to be the probability that two technological classes j and m co-occur within the same patent, then the measure of multidimensional entropy focuses on the variety of co-occurrences of technological classes within firms' patents portfolios.

Moreover, the total index can be decomposed in a "within" and a "between" part anytime the events to be investigated can be aggregated in 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. It can be easily shown that the decomposition theorem holds also for the multidimensional case. Hence if one allows $j \in S_g$ and $m \in S_z$ (g = 1,…,G; z = 1,…, Z), we can rewrite *H(X,Y)* as follows:

$$H(X,Y) = H_Q + \sum_{g=1}^{G} \sum_{z=1}^{Z} P_{gz} H_{gz} \qquad \text{(A2)}$$

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

$$H_Q = \sum_{g=1}^{G} \sum_{z=1}^{Z} P_{gz} \log_2 \frac{1}{P_{gz}} \qquad \text{(A2a)}$$

$$P_{gz} = \sum_{j \in S_g} \sum_{m \in S_Z} p_{jm} \quad \text{(A2b)}$$

$$H_{gz} = \sum_{j \in S_g} \sum_{m \in S_z} \frac{p_{ij}}{P_{gz}} \log_2 \left(\frac{1}{p_{jm} / P_{gz}} \right) \quad \text{(A2c)}$$

Following Frenken et al. (2007), we can refer to between-group and within-group entropy respectively as *unrelated technological variety* (*UTV*) and *related technological variety* (*RTV*), while total information entropy is referred to as *general technological variety (TV)*. The distinction between related and unrelated variety is based on the assumption that any pair of entities included in the former generally are more closely related, or more similar to any pair of entities included in the latter. This assumption is reasonable when a given type of entity (patent, industrial sector, trade categories etc.) is organized according to a hierarchical classification. In this case each class at a given level of aggregation contains "smaller" classes, which, in turn contain yet "smaller" classes. Here, small refers to a low level of aggregation.

We can reasonably expect then that the average pair of entities at a given level of aggregation will be more similar than the average pair of entities at a higher level of aggregation. Thus, what we call related variety is measured at a lower level of aggregation (3 digit class within a 1 digit macro-class) than unrelated variety (across 1 digit macro-classes). This distinction is important because we can expect unrelated (or inter-group) variety to negatively affect productivity growth, while related (or intra-group) variety is expected to be positively related to productivity growth. Moreover, the evolution of total variety is heavily influenced by the relative dynamics of related and unrelated variety, such that if unrelated variety is dominant the effects of total variety on productivity growth can be expected to be negative, while the opposite holds if related technological variety dominates the total index (Krafft, Quatraro, Saviotti, 2011).

Knowledge Coherence

Third, we calculated the *coherence* (*R*) of firms' knowledge base, defined as the average complementarity of any technology randomly chosen within the firm's portfolio 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 the firm level. First of all, one should calculate the weighted average relatedness WAR_i of technology *i* with respect to all other technologies present within the sector. Such a measure builds upon the measure of technological relatedness τ, which is introduced in Appendix A. Following Teece et al. (1994), WAR_j is defined as the degree to which technology *j* is related to all other technologies $m \neq j$ within the firm *i*, weighted by patent count P_{mit}:

$$WAR_{jit} = \frac{\sum_{m \neq j} \tau_{jm} P_{mit}}{\sum_{m \neq i} P_{mit}} \quad \text{(A3)}$$

Finally the coherence of knowledge base within the firm is defined as weighted average of the WAR_{jit} measure:

$$R_{it} = \sum_{j \neq m} WAR_{jit} \times \frac{P_{jit}}{\sum_j P_{jit}} \qquad \text{(A4)}$$

This measure captures the degree to which technologies making up the firm's knowledge base are complementary 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.

Cognitive Distance

We finally implement a measure of knowledge similarity, as proxied by cognitive distance (Nooteboom, 2000), which is 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-cented correlation of the vectors V_{lk} and V_{jk}. The similarity of technologies l and j can then be defined as follows:

$$S_{lj} = \frac{\sum_{k=1}^{n} V_{lk} V_{jk}}{\sqrt{\sum_{k=1}^{n} V_{lk}^2} \sqrt{\sum_{k=1}^{n} V_{jk}^2}} \qquad \text{(A5)}$$

The idea underlying the calculation of this index is that two technologies j and i 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:

$$d_{lj} = 1 - S_{lj} \qquad \text{(A6)}$$

Once the index is calculated for all possible pairs, it needs to be aggregated at the firm 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 i, i.e. the average distance of i from all other technologies.

$$WAD_{lt} = \frac{\sum_{j \neq l} d_{lj} P_{jit}}{\sum_{j \neq l} P_{jit}} \tag{A7}$$

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:

$$CD_t = \sum_l WAD_{lit} \times \frac{P_{lit}}{\sum_l P_{lit}} \tag{A8}$$

Table 1 - Sectoral distribution of sampled firms

Industry	ISIC 4 code	Freq.	Percent	Cumul
Agriculture, forestry and fishing	A	5	0.15	0.15
Mining and quarrying	B	5	0.15	0.30
Manufacturing	C	2556	76.60	76.90
Electricity, gas, steam and air conditioning supply	D	5	0.15	77.05
Water supply; sewerage, waste management and remediation activities	E	15	0.45	77.49
Construction	F	82	2.46	79.95
Wholesale and retail trade; repair of motor vehicles and motorcycles	G	289	8.66	88.61
Transportation and storage	H	20	0.60	89.21
Accommodation and food service activities	I	5	0.15	89.36
Information and communication	J	35	1.05	90.41
Financial and insurance activities	K	32	0.96	91.37
Real estate activities	L	95	2.85	94.22
Professional, scientific and technical activities	M	153	4.58	98.80
Administrative and support service activities	N	28	0.84	99.64
Human health and social work activities	Q	6	0.18	99.82
Arts, entertainment and recreation	R	2	0.06	99.88
Other service activities	S	4	0.12	100.00
Total		3337	100	

Table 2 - Size distribution of sampled firms (employees)

Size Class		Freq.	Percent	Cumul.
Micro	(<20)	783	22.38	22.38
Small	(20-99)	1,357	38.78	61.16
Medium	(100-199)	527	15.06	76.22
Large	(>199)	832	23.78	100
	Total	3,499	100	

Table 3 - Descriptive Statistics of Knowledge Properties

	N	mean	min	max	p10	p25	p50	p75	p90
Kn. Capital	8784	0.331	-0.112	2.472	0.067	0.129	0.264	0.545	0.615
CD	1535	0.041	-1.961	1.880	-0.252	-0.076	0.005	0.130	0.377
RKV	5237	0.007	-3.478	3.382	-0.006	0.000	0.000	0.000	0.001
UKV	3166	0.008	-1.355	1.217	-0.041	0.000	0.000	0.001	0.070
KV	2400	-0.003	-0.500	0.510	-0.101	-0.013	0.000	0.009	0.094
Kn. Coherence	2231	-0.069	-2.804	2.575	-0.657	-0.265	-0.013	0.102	0.487

All variables are expressed in growth rates, before normalization.

Table 4 - Correlation Matrix

	Kn. Coherence	Kn. Cap.	CD	RKV	UKV	KV
Kn. Coherence	1.000					
Kn. Cap.	**0.058**	1.000				
	0.006					
CD	**-0.064**	0.018	1.000			
	0.017	0.473				
RKV	0.007	**0.286**	-0.010	1.000		
	0.733	0.000	0.710			
UKV	0.019	**0.481**	-0.012	0.031	1.000	
	0.424	0.000	0.684	0.079		
KV	0.039	0.520*	0.052	**0.301**	**0.550**	1.000
	0.136	0.000	0.098	0.000	0.000	

All variables are expressed in normalized growth rates according to equation (2).
Figures in bold indicates correlation coefficients significant at 5%.

Table 5 - Econometric results (overall estimation, LAD regression)

Dep. Var.	(1) K. Cap.	(2) Koh	(3) TV	(4) RTV	(5) UTV	(6) CD	(7) Kdiv
Const.	-0.0550***	0.0221**	-0.0187***	0.000776	-0.0145***	0.00834	0.00432***
	(0.00242)	(0.00904)	(0.00230)	(0.00250)	(0.00117)	(0.00611)	(0.000141)
β_1	0.504***	-0.203***	0.00593	0.0106	0.0218***	-0.227***	-0.0146***
	(0.0106)	(0.0344)	(0.00729)	(0.0181)	(0.00786)	(0.0606)	(0.00348)
β_2	0.0752***	-0.148***	0.0223***	0.0357**	0.0333***	-0.0130	-0.0105***
	(0.0138)	(0.0298)	(0.00538)	(0.0153)	(0.0103)	(0.0356)	(0.00230)
β_3	-0.0169***	-0.0382	-0.00600	0.0634***	0.0482***	-0.0142	-0.0105***
	(0.00649)	(0.0263)	(0.0130)	(0.0212)	(0.00638)	(0.0244)	(0.00126)
Observations	3,758	867	872	2,292	1,549	556	3,758

Bootstrap Standard errors in parentheses.
All variables are expressed in normalized growth rates according to eq. (2).
*** p<0.01, ** p<0.05, * p<0.1

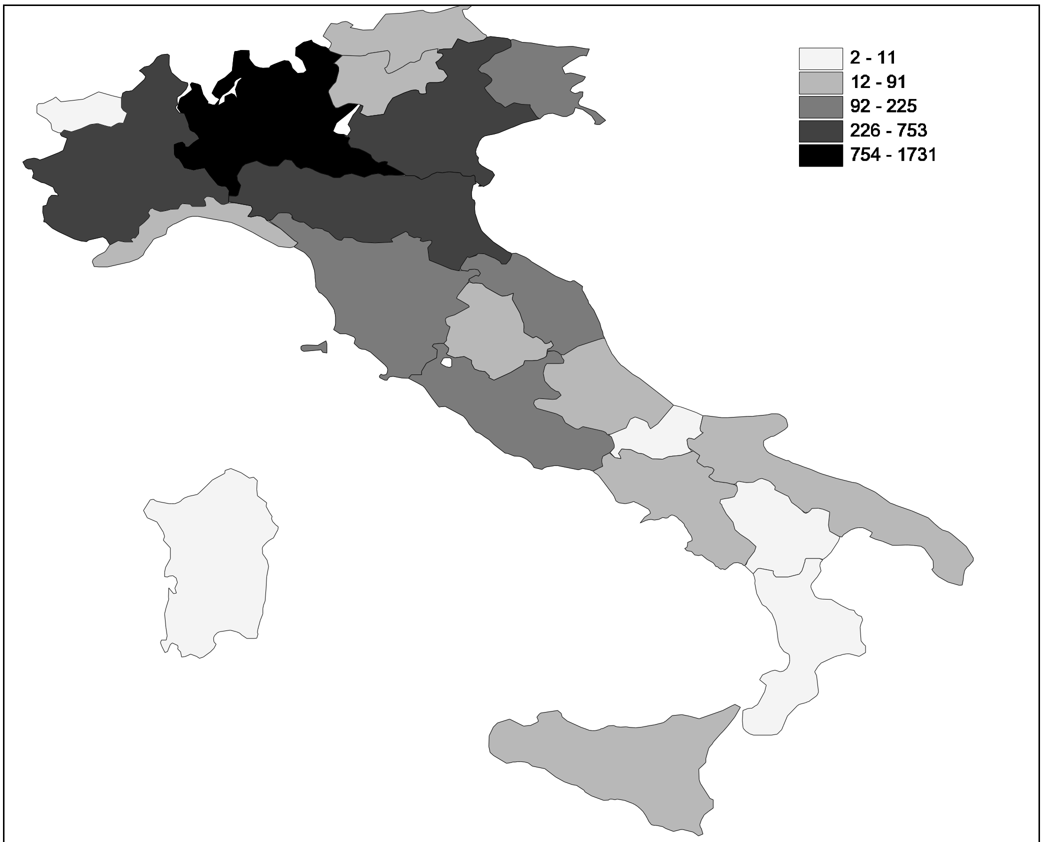
Table 6 - Econometric results (quantile regression)

	Dep. Var.	(1) K. Cap.	(2) Koh	(3) TV	(4) RTV	(5) UTV	(6) CD	(7) Kdiv
q10	*Const.*	-0.0969***	-0.520***	-0.239***	-0.0367***	-0.118***	-0.222***	-0.248***
		(0.00395)	(0.0321)	(0.0183)	(0.00728)	(0.0150)	(0.0177)	(0.0207)
	β_1	0.495***	-0.490***	-0.00116	0.0150	0.123*	-0.303**	-0.354***
		(0.0209)	(0.0657)	(0.150)	(0.00921)	(0.0669)	(0.132)	(0.0278)
	β_2	0.0902***	-0.421***	0.0997	0.00700	0.153***	-0.0586	-0.334***
		(0.0165)	(0.114)	(0.121)	(0.00595)	(0.0371)	(0.0522)	(0.0311)
	β_3	-0.0825***	-0.145	-0.442***	0.00416	0.114***	-0.0183	-0.334***
		(0.00941)	(0.0883)	(0.0598)	(0.00948)	(0.0191)	(0.0791)	(0.0373)
q25	*Const.*	-0.0642***	-0.215***	-0.0754***	-0.0229***	-0.0317***	-0.0845***	-0.0155***
		(0.000602)	(0.0209)	(0.00795)	(0.000650)	(0.00188)	(0.00903)	(0.000513)
	β_1	0.554***	-0.330***	-0.00307	0.0115	0.0252*	-0.283***	-0.0484***
		(0.00794)	(0.0507)	(0.0636)	(0.00954)	(0.0133)	(0.0718)	(0.00582)
	β_2	0.0202***	-0.281***	0.0943***	0.0129***	0.0389***	0.00544	-0.0484***
		(0.00539)	(0.0629)	(0.0309)	(0.00229)	(0.0111)	(0.0463)	(0.00891)
	β_3	-0.0137**	-0.0608	-0.0988**	0.00932	0.0353***	-0.00901	-0.0484***
		(0.00633)	(0.0380)	(0.0484)	(0.00589)	(0.00703)	(0.0290)	(0.00932)
q50	*Const.*	-0.0550***	0.0221***	-0.0187***	0.000776	-0.0145***	0.00834	0.00432***
		(0.00208)	(0.00827)	(0.00164)	(0.00234)	(0.00112)	(0.00635)	(0.000168)
	β_1	0.504***	-0.203***	0.00593	0.0106	0.0218**	-0.227***	-0.0146***
		(0.00919)	(0.0394)	(0.00541)	(0.0124)	(0.0101)	(0.0692)	(0.00302)
	β_2	0.0752***	-0.148***	0.0223***	0.0357*	0.0333***	-0.0130	-0.0105***
		(0.00971)	(0.0467)	(0.00694)	(0.0216)	(0.0128)	(0.0351)	(0.00177)
	β_3	-0.0169***	-0.0382	-0.00600	0.0634***	0.0482***	-0.0142	-0.0105***
		(0.00565)	(0.0256)	(0.0109)	(0.0228)	(0.00571)	(0.0304)	(0.00255)
q75	*Const.*	-0.0326***	0.162***	-0.000979	0.00454***	0.00580***	0.0857***	0.0187***
		(0.00104)	(0.0164)	(0.00141)	(0.000449)	(0.00160)	(0.00824)	(6.01e-05)
	β_1	0.618***	-0.249***	0.00955*	0.00877	0.0135	-0.232***	-0.0195**
		(0.0169)	(0.0765)	(0.00501)	(0.0126)	(0.0155)	(0.0600)	(0.00824)
	β_2	0.0188*	-0.171***	0.0169**	0.0364	0.0669***	-0.0668*	-0
		(0.0112)	(0.0525)	(0.00782)	(0.0248)	(0.0171)	(0.0405)	(0.000224)
	β_3	0.0185***	-0.00981	0.0352***	0.0769***	0.0268***	-0.0118	0
		(0.00336)	(0.0258)	(0.00947)	(0.00996)	(0.0103)	(0.0464)	(0.00392)
q90	*Const.*	-0.0189***	0.384***	0.0374***	0.0275***	0.0287***	0.202***	0.171***
		(0.00267)	(0.0400)	(0.0116)	(0.00384)	(0.00426)	(0.0194)	(0.0240)
	β_1	0.745***	-0.458***	-0.00515	0.0140	0.0485*	-0.323***	-0.155***
		(0.0262)	(0.0881)	(0.0756)	(0.00926)	(0.0262)	(0.0847)	(0.0409)
	β_2	0.0213	-0.389***	-0.0115	0.0318*	0.0752***	-0.0848	0.0908***
		(0.0179)	(0.0997)	(0.0361)	(0.0183)	(0.0211)	(0.0584)	(0.0301)
	β_3	0.0226**	-0.155**	-0.00273	0.0736***	0.0471**	-0.0186	0.203***
		(0.00886)	(0.0695)	(0.0310)	(0.0160)	(0.0207)	(0.0920)	(0.0321)
	Obs.	3,758	867	872	2,292	1,549	556	3,758

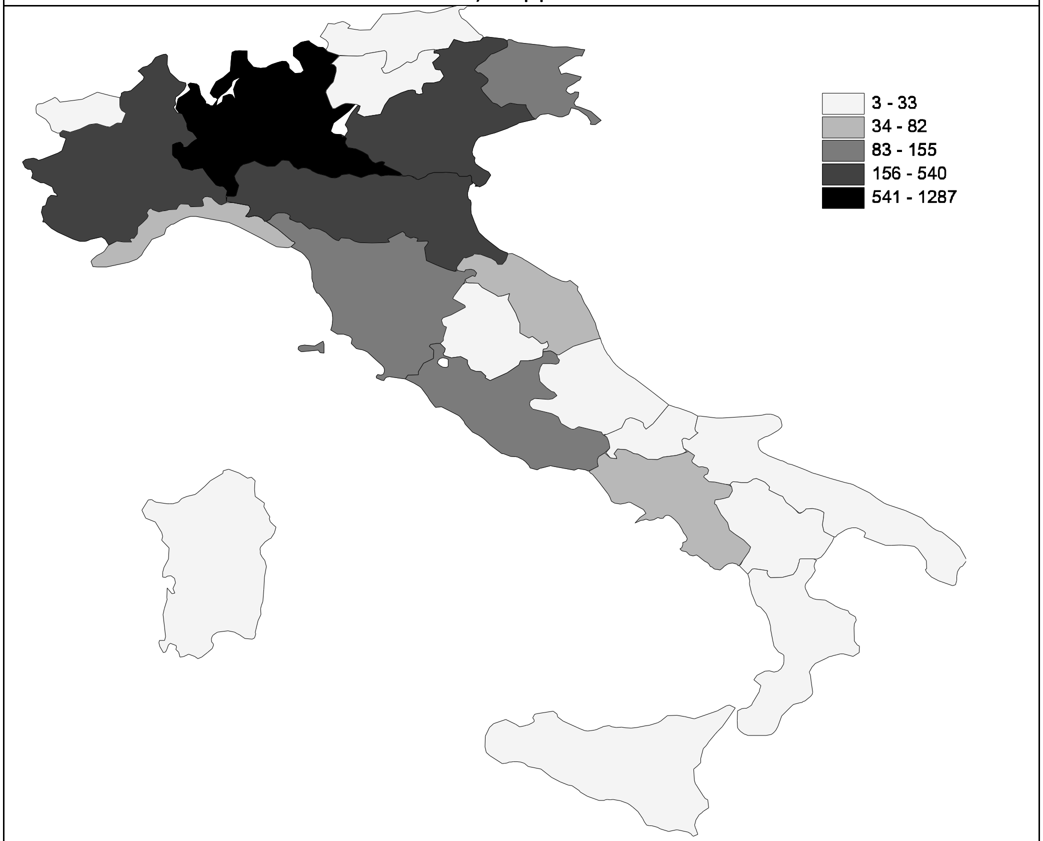
Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Figure 1 - Regional distribution of firms and applicants



a) applicants



b) firms

Figure 2 - Kernel Density Distribution of the Properties of Knowledge Structure (normalized growth rates)

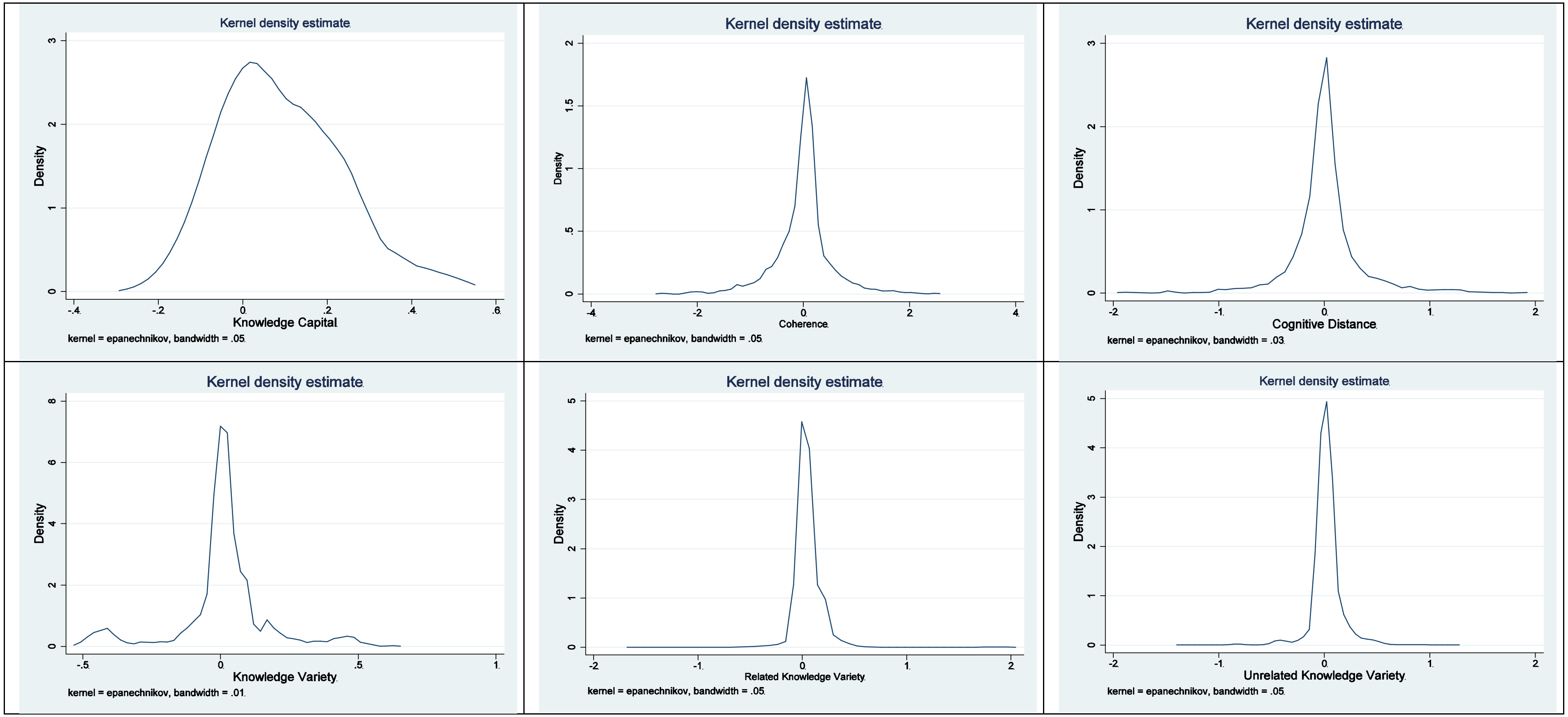
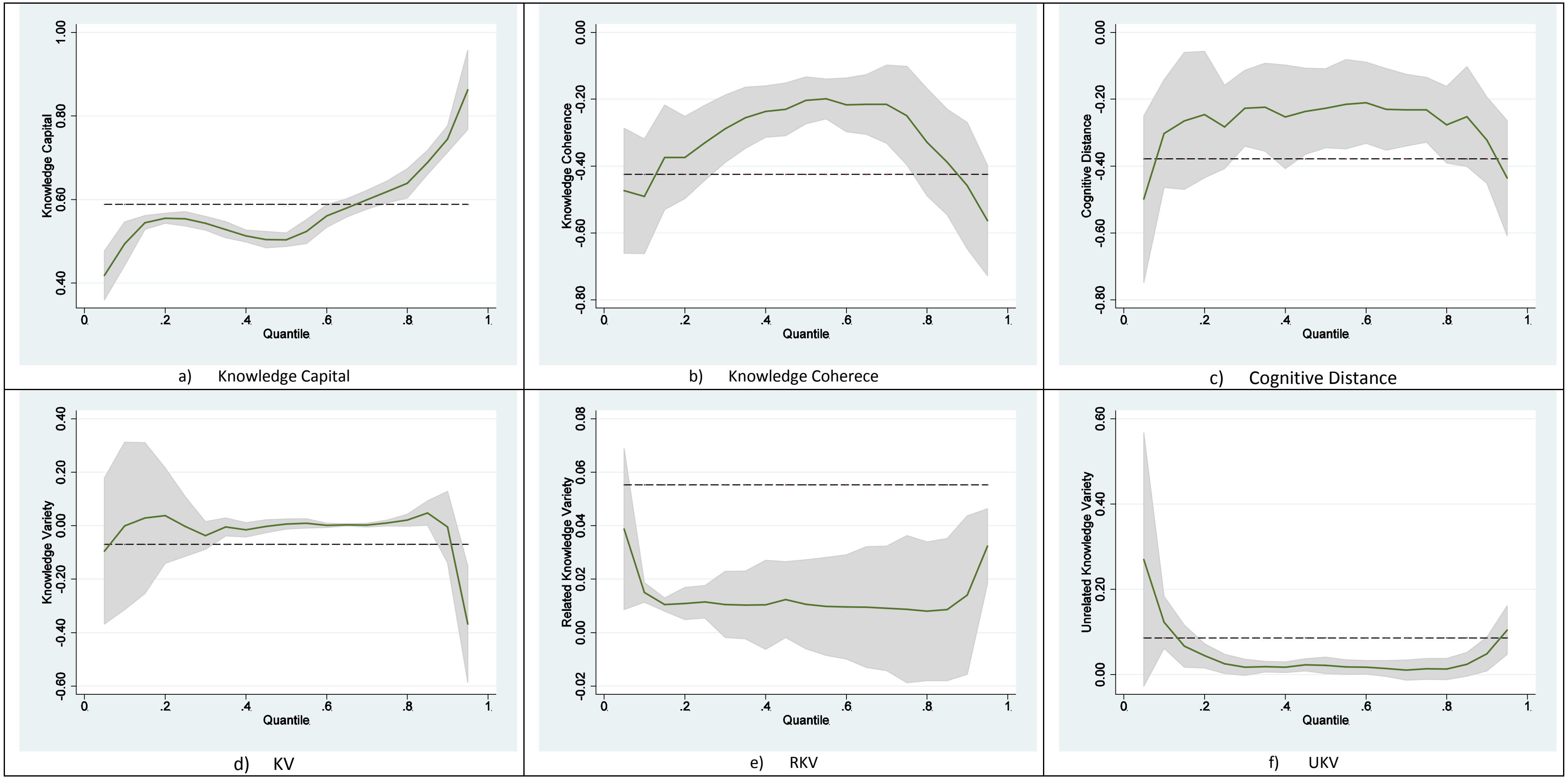


Figure 3 - Regression quantiles for knowledge structure properties autocorrellation coefficients, with 95% confidence intervals



a) Knowledge Capital
b) Knowledge Coherece
c) Cognitive Distance
d) KV
e) RKV
f) UKV