CIRCLE

LUND UNIVERSITY

Papers in Innovation Studies

Paper no. 2015/6

Knowledge base combinations and innovation performance in Swedish regions

Markus Grillitsch (Markus.Grillitsch@circle.lu.se)
CIRCLE, Lund University

Roman Martin (Roman.Martin@circle.lu.se)
CIRCLE, Lund University

Martin Srholec (Martin.Srholec@circle.lu.se)
CIRCLE, Lund University & CERGE-EI, Charles University and Economics Institute of the Academy of Sciences of the Czech Republic

This is a pre-print version of a paper that has been submitted for publication to a journal.

This version: February 2015

Centre for Innovation, Research and Competence in the Learning Economy (CIRCLE)
Lund University
P.O. Box 117, Sölvegatan 16, S-221 00 Lund, SWEDEN
http://www.circle.lu.se/publications

WP 2015/6

Knowledge base combinations and innovation performance in Swedish regions

Markus Grillitsch, Roman Martin and Martin Srholec

Abstract

The literature on geography of innovation suggests that innovation outcomes depend on the type of knowledge base employed by firms. While knowledge bases are distinct categories with regards to the nature and the rational of knowledge creation, existing studies also stress that innovation usually involves more than one knowledge base. In fact, new ideas often occur when analytical, synthetic and symbolic knowledge intertwines. It remains unclear, though, which combinations of knowledge bases are most conducive to innovation at the level of the firm, and how this is influenced by the knowledge bases available in the regional milieu. Therefore the contribution of this paper is threefold: i) to measure knowledge bases of firms and their regional heterogeneity in a more comprehensive way than the existing empirical literature has been able to do so far, ii) to quantitatively assess the impact of combinations of knowledge bases on innovation output, iii) to analyze the interplay between firm- and region-level knowledge bases (and combinations thereof) in generating innovations. Empirically, the paper applies econometric analysis on firm- and region-level data from Sweden. The knowledge base of firms is captured using detailed occupational data derived from linked employer-employee datasets that is merged at the firm-level with information from Community Innovation Surveys. The empirical analysis reveals in a quantitative way the extent to which the knowledge base combinations affect innovativeness of firms.

JEL codes: O30, O31, R10

Keywords: Knowledge bases, knowledge combination, regions, innovation performance, microdata, cross-level interaction, Sweden

1. Introduction

Scholars dealing with the geography of innovation have been preoccupied with the question on why and how innovation and knowledge creation are influenced by the regional context. Considerations on this question have been advanced in particular in the literature on industrial districts (Brusco 1986; Becattini 1989; Bellandi 1989), innovative milieus (Aydalot 1986; Camagni 1991; Maillat et al. 1995), learning regions (Asheim 1996; Morgan 1997; Hassink 2001) and regional innovation systems (Cooke, Uranga, and Etxebarria 1998; Cooke, Heidenreich, and Braczyk 2004; Asheim and Gertler 2005), in which innovation is understood as a result of interactive learning processes taking place not only within the boundaries of the firm, but also with various actors outside the firm, governed by an institutional framework and confined to a specific geographical area.

Recent literature on the geography of innovation deals with the question which types of knowledge are employed and exchanged in these interaction processes. One increasingly prominent conceptual framework to study knowledge dynamics between firms and other organizations is the distinction between analytical, synthetic and symbolic knowledge bases (Laestadius 1998; Asheim and Gertler 2005; Asheim, Coenen, and Vang 2007). Knowledge bases are distinct theoretical categories with regards to the nature and the rationale of knowledge creation, however, the existing empirical research on this topic emphasizes that innovative firms often combine different types of knowledge (Moodysson, Coenen, and Asheim 2008; Manniche 2012). While the existing evidence is predominantly based on intensive case studies (e.g. Moodysson 2008; Martin and Moodysson 2011; Strambach and Klement 2012), econometric research on the role of knowledge base combinations for innovation is lacking[1].

This paper sets out to fill that gap with an extensive econometric analysis on the interplay of firm internal and regional knowledge bases and their contribution to innovation performance of firms. We address the question which types of firm internal knowledge base combinations are most conducive to innovation, and which types of regional knowledge base combinations are most favorable for innovation in which types of firms. In doing so, we analyze whether specializing or combining different types of knowledge bases in-house leads to better innovation outcomes, and whether firms benefit more if they are located in a regional environment that possesses a specialized or a more diversified set of knowledge bases.

Empirically, we draw on a large firm-level dataset derived from five waves of Community Innovation Surveys (CIS) merged with detailed georeferenced occupational data from linked employee-employer datasets in Sweden. This occupational data allows us to measure knowledge bases more reliably and accurately than existing studies have been able to do. We estimate tobit models, in which the innovation output of firms given by sales of products new

[1] Econometric studies on knowledge bases are scarce, and the few notable exceptions deal with knowledge networks, but have no specific focus on innovation outcomes and measure knowledge bases rather indirectly (e.g. Asheim, Ebersberger, and Herstad 2012; Herstad, Aslesen, and Ebersberger 2014).

to the market is the function of knowledge bases of firms, regions and their interactions, while accounting for other relevant characteristics.

The remainder of the paper is organized as follows. Section 2 reviews the literature on differentiated knowledge bases with focus on the most recent works on knowledge base combinations. We build on the implicit argument that firms that combine different types of knowledge and modes of innovation tend to perform better. Moreover, we take up the literature on agglomeration economies and knowledge spillovers, and question whether regional specialization or diversity in knowledge bases is more conducive to innovation. Section 3 explains the data and methodology. Section 4 describes the econometric model and explores the relationship between knowledge base combinations and the innovation performance of firms. Section 5 concludes and provides implications for policy concerned with regional development.

2. Theoretical framework: Knowledge bases and the geography of innovation

The notion of knowledge bases provides a differentiated perspective on the nature of knowledge used by firms in the innovation process (Laestadius 1998; Asheim and Gertler 2005; Asheim, Coenen, and Vang 2007). The concept distinguishes between three epistemologically different types of knowledge, that is, analytical, synthetic and symbolic. This distinction is grounded on universal categories of knowledge stemming from ancient Greek philosophy, specifically the notions of episteme, technê and art. Analytical knowledge refers to theoretical knowledge that is applied to understand and explain features of the natural world. It is mostly related to scientific principles and competences. The synthetic knowledge base refers to knowledge that is practical and applied for the purpose of creating goods to attain functional goals. It is mostly associated with engineering skills. The symbolic knowledge base has been introduced to account for the growing importance of design and aesthetics in many products and services. It is mostly concerned with creativity, design and cultural values.

Grounded on these three categories, the knowledge base concept has been applied to study industry specific differences in the geography of innovation (e.g. Plum and Hassink 2011; Aslesen and Freel 2012; Martin and Moodysson 2013). As an industry typology, knowledge bases can be regarded as alternative to traditional industry classifications that are based on product categories (e.g. NACE and SNI), or on characteristics of innovating companies, as it is the case with the taxonomy developed by Pavitt (1984). Traditional industry classification systems have limitations especially when it comes to capturing emerging and transforming industries, and industries that cross over traditional product categories. Building on three universal categories of knowledge, the knowledge base typology can be seen as more generic and applicable to study innovation in a broad range of regions, sectors and companies.

Studies conducted at the level of industries reveal that the geography of innovation differs considerably by the dominant knowledge base (Martin and Moodysson 2013). Analytical industries interact often with research organizations. As science-based innovation relies on codified knowledge that is abstract, and universally valid, the exchange of analytical knowledge is less restricted to spatial distance (Knorr-Cetina 1999; Moodysson 2008). In synthetic industries, knowledge exchange predominantly occurs along the supply chain and between users and producers, often within communities of practice (Wenger 1998; Gertler 2008). Innovation is driven by learning by doing, using and interacting, and the most important knowledge form is tacit, which implies that relatively little interaction takes place over far geographical distance (Moodysson, Coenen, and Asheim 2008; Herstad, Aslesen, and Ebersberger 2014). Innovation in symbolic industries is even more governed by the local context. Symbolic industries innovate within short-term projects, and companies change their connections frequently (Grabher 2002). The importance of cultural knowledge and the project-based organization of innovation implies that knowledge exchange takes place primarily within locally configured networks (Martin and Moodysson 2011).

While knowledge bases have been a fruitful device to study industry specific differences in the geography of innovation, it should not be neglected that the knowledge base typology is a theoretically derived categorization. In practice, innovation hardly ever involves only one type of knowledge base, rather, it often includes a combination of different knowledge types obtained from different sources (Trippl, Tödtling, and Lengauer 2009; Grillitsch and Trippl 2014; Herstad, Aslesen, and Ebersberger 2014). Also, strong heterogeneity between firms can often outweigh the role of sectors in explaining different patterns of innovation (Srholec and Verspagen 2012). This calls for a more nuanced view on knowledge processes taking place on the level of the firm.

In line with that, Moodysson et al. (2008) find that concrete innovation projects in the life science industry consist of a mix of analytical and synthetic modes of knowledge creation, with different sensitivity to spatial distance. Manniche (2012) analyses the type of knowledge bases involved in a range of innovation projects, and finds that knowledge interactions are typically characterized by one knowledge base, whereas the entire biography of an innovation very often involves analytical, synthetic and symbolic knowledge. Martin and Moodysson (2011) study knowledge bases in a typical innovation project of a new media company, and find that the problem solving sequence includes analytical, synthetic and symbolic challenges, while symbolic knowledge defines the basis for the firm’s competitiveness. Comparing data from 15 case studies, Tödtling and Grillitsch (2014) find that the combinations of firm competencies and external knowledge sourcing relate to the dominant knowledge base in the industry and lead to different types of innovation.

Furthermore, Zukauskaite and Moodysson (2013) find that even though the food sector is dominated by synthetic modes of innovation, some food companies branch out in new product segments by combining synthetic and analytical knowledge. This is in line with Strambach and Klement (2013), who study innovation biographies in the automotive industry, and argue that combinatorial knowledge dynamics can play an important role in creating new development paths, and thereby diminishing the risk of negative lock-in effects. Even though their empirical evidence is on the level of an innovation project, they suggest that dynamics at the sectorial and regional level will follow a similar pattern. In this vein, Martin and Trippl (2014) study how knowledge bases co-evolve with the development of a cluster, and find that clusters can renew themselves by drawing on a variety of knowledge bases obtainable in the regional milieu.

All these studies point in the direction that even though analytical, synthetic and symbolic knowledge are distinct epistemological categories, they are hardly employed exclusively and detached from each other. In fact, the notion of knowledge bases suggests that innovation is a result of diverse knowledge inputs. Innovation results not only from science-based knowledge but also from engineering know-how and arts-based experience. It remains unclear, however, whether analytical, synthetic and symbolic knowledge contribute to the same extent to innovation outcomes, and which knowledge base combinations are most conducive to innovation. Are firms that are clearly dominated by one knowledge base more prone to innovation, as they can channel their resources and specialize in a particular field? Or, alternatively, are firms that are capable of combining different types of knowledge more innovative, as they have the necessary breadth of competences to deal with diverse knowledge inputs? The first research question can be formulated as follows:

Which types of firm internal knowledge bases and combinations thereof are most conducive to innovation?

Another key issue addressed in this paper involves knowledge interdependencies between the firm and the regional milieu. Firms seldom innovate in isolation, but usually source knowledge from other actors in the innovation system (Moulaert and Sekia 2003; Asheim and Gertler 2005). The central argument for the role of the region is that the spatial and functional integration of innovation activities generates positive effects for co-located firms. These effects that are often referred to as agglomeration economies are beyond the control of the individual firm and result from the presence and collective action of other firms in the region (Malmberg 1996; Parr 2002). They can arise from co-location of firms in the same industry, i.e. localization economies (Marshall 1920; Arrow 1962; Romer 1986), or in different industries, i.e. urbanization economies (Jacobs 1969). Transcending this dichotomy, the literature on related variety suggests that sectorial diversity is important, though knowledge can spill over most effectively if a certain degree of similarity exists between sectors (Frenken, Van Oort, and Verburg 2007; Boschma and Iammarino 2009).

The knowledge spillover debate shows that agglomeration economies consist of various benefits stemming from traded and untraded interdependencies (Storper 1995) and localized capabilities (Maskell and Malmberg 1999), and go beyond cost-savings from shared infrastructure and access to spatially constrained resources, as it was emphasized in traditional location theory. Today, agglomeration economies are regarded as the capacity to foster innovation through knowledge exchange in the local milieu. One key to explaining spatial clustering of innovation activities is the important role of geographical proximity for interactive learning (Malmberg 1996; Malmberg and Maskell 2002). The region is seen as arena for the emergence of social relationships and common norms and values, which facilitate interaction and collaboration. Innovation is seen as embedded in a particular social, institutional and spatial context, and mutual trust between various actors in the regional milieu positively affects their innovative performance (Asheim and Gertler 2005).

Transferring the discussion on spillovers to the realm of knowledge bases, the question arises to what extend the innovation performance of firms is contingent on the knowledge bases available in the regional environment. Existing studies on knowledge bases observe different degrees of diversity and specialization among regions, with most regions being dominated by one knowledge base, and only few regions being characterized by a balanced mixed of analytical, synthetic and symbolic knowledge (Asheim and Hansen 2009; Martin 2012). It remains unclear, however, whether regional diversity or specialization in knowledge bases is more conducive to innovation, and which type of regional knowledge base configuration is most beneficial for innovation in which type of firms.

Following the localization economies argument, it is reasonable to expect that firms perform best if they are located in a region with similar knowledge specialization. Being embedded in a region with the same knowledge base will permit them to interact locally with other firms that work with similar problems and use similar skills and competencies, which increases their innovative scale. But then again, following the urbanization economies argument, it is reasonable to expect that firms are more innovate if they are located in a region with diverse knowledge bases. This will allow them to access dissimilar knowledge input from other actors in the region, and thereby widen their scope for innovation. From this, the competing theses follow that either regional specialization or regional diversity in knowledge bases is most beneficial for firm innovation.

Moreover, the question arises which types of regional knowledge base combinations are most beneficial for innovation performance in which types of firms. One can expect that firms with a synthetic knowledge base benefit most from being located in a region with a strong analytical knowledge base, as they can upgrade their innovation activities through scientific knowledge input (Jensen et al. 2007). But it is also reasonable to expect that the presence of a strong symbolic knowledge base in the region outweighs the importance of scientific knowledge, as design values and improved user experience are increasingly important for innovation in many sectors of the economy (Walsh 1996; Scott 2004). The interplay between regional knowledge base configurations and the innovation performance of firms is the second key issue addressed in the paper, which leads to the following research question:

Which types of regional knowledge base combinations are most conducive to innovation in which types of firms?

3. Data and Methodology

The data used for this study merges the Community Innovation Survey (CIS) with individual and firm registry data from the Statistical Office of Sweden (SCB). The registry data is used to identify firm-level and region-level knowledge bases through occupational data. Occupational data relates to the type of work an individual is performing and the skills and education usually required for this type of work. Occupational data is available for each individual aged 16 and over registered in Sweden on 31 December each year. In line with Asheim and Hansen (2009) and Martin (2012), we select occupations that are likely to be involved in innovation activities, and that can be clearly attributed to one of the three knowledge bases. Only few studies on knowledge bases use occupational data and if so, exploit this data on a three-digit level and aggregate it for regions (e.g. Asheim and Hansen 2009; Martin 2012). By using the most detailed four-digit data at the firm-level, we identify knowledge bases and their combinations more accurately than the existing literature has been able to do so far (for the list of occupations, see Appendix Table A1).

The database on firm- and municipal level knowledge bases is merged with five waves of the CIS, which is conducted every second year from 2004 to 2012 following the Oslo Manual (OECD 2005). The CIS data covers innovation activities of firms in the last three years before the date of the survey, thus the reference periods are 2002-04, 2004-06, 2006-08, 2008-10 and 2010-12. The sample for the CIS is stratified by firm size and sector based on SCBs firm register and includes firms with 10 employees or more. Answering the survey is compulsory by law. The combination of five waves of the CIS increases the size of the sample and thereby the regional representativeness of the micro data. Approximately half of the firms are observed repeatedly, which allows us to exploit the panel structure of the data.

The dependent variable for innovation output refers to the percentage of new to the market innovations in total turnover in the final year of the reference period. Firm-level knowledge bases are calculated as the share of employees with an analytical, synthetic or symbolic knowledge base in total employment over the three-year reference period. Size of the firm is measured as log of the number of employees in the initial year of the reference period. Firms were asked whether they are affiliated to an enterprise group, from which we derive a dummy with value 1 if the firm is a part of a group. Next, there is information about the geographic markets in which the firms sell their goods or services, from which we obtain a dummy with value 1 if the firm delivers abroad. Finally, industry dummies are derived from a classification in seven broad sectors based on two-digit NACE, rev. 1.1 categories.[2]

Table 1 presents descriptive statistics about the firm-level variables. After omitting observations with missing records, the pooled micro sample consists of 20,482 observations (3,105 in 2002-04, 3,093 in 2004-06, 4,409, 2006-08 4,329 in 2008-10 and 5,546 in 2010-12). About 3.8% of the firms' sales were generated by innovations new to the market. By far the most common knowledge base maintained by the firms in-house is synthetic, followed with a large gap by symbolic and analytical. Small and medium size firms are well represented, as the median size of the firm is about 26 employees and only around 11% of the sample consists of large firms with more than 250 employees. More than two-thirds of the firms are affiliated to a group and about three-fifths of them are exporters.

[2] The sectors are constructed using the following two-digit NACE. rev 1.1 categories: manufacturing (10-35); mining and utilities (5-10 and 35-41); wholesale and retail (45-49); transportation (49-55); information and communication (58-64); financial and insurance activities (64-68); professional, scientific and technical activities (69-77). Due to the very detailed spatial structure, we cannot account for more detailed industry classifications.

Table 1 Firm-level variables

Variable	Observations	Mean	Std. Dev.	Min	Max
New to the market innovations in total turnover (%)	20,482	3.83	12.02	0	100.00
Firm analytical knowledge (%)	20,482	0.74	4.72	0	100.00
Firm synthetic knowledge (%)	20,482	12.33	20.93	0	100.00
Firm symbolic knowledge (%)	20,482	2.04	9.29	0	100.00
Number of employees (log)	20,482	3.59	1.40	0	10.21
Member of a group (dummy)	20,482	0.68	0.47	0	1.00
Foreign sales (dummy)	20,482	0.58	0.49	0	1.00

The regional variables are constructed based on the number of individuals working in each municipality in total, by knowledge base and by occupation during the three years reference period of the respective CIS wave. In addition to individuals within the municipal borders, the regional variable also accounts for knowledge spillovers from other municipalities. Spillovers from other municipalities are given by:

(1) $S_{xr} = \sum_{s=1}^{n} I_{xs} e^{-\lambda t_{rs}}$

where S_{xr} denotes the spillovers for variable x in municipality r. I_{xs} stands for the number of individuals who have the characteristic x in other municipalities s. The potential knowledge spillovers from other municipalities $s = 1,\dots,n$ is diminished by applying an exponential distance-decay function $e^{-\lambda t_{rs}}$. t_{rs} denotes the travel distance in minutes by car between the municipalities r and s[3]. The distance values are multiplied by λ, capturing how sensitive knowledge spillovers are to time distance. In the regressions, we use different values for λ in order to test for robustness. All region-level variables refer to municipalities and are described by:

(2) $M_{xr} = I_{xr} + S_{xr}$

whereby M_{xr} denotes the region-level variable for accessibility to individuals with characteristic x in municipality r and is the sum of I_{xr}, the number of individuals with characteristic x in municipality r, and the respective spillovers S_{xr}. On this basis, we derive variables for regional knowledge base concentrations by calculating the share of individuals with an analytical, synthetic or symbolic knowledge base in the total regional workforce.

Furthermore, we are interested in the potential of firms to combine different knowledge bases locally. This potential depends on the regional presence of analytical, synthetic and symbolic knowledge, thus relates to the diversity between knowledge bases. In addition, we look at the potential of firms to combine different skills and competences belonging to the same knowledge base. This relates to the diversity of occupations within a specific knowledge base. Between and within diversity of knowledge bases is captured with an entropy measure that

[3] The time-distance measure has been provided by the Swedish Transport Authority upon our request.

has correspondingly been used in studies on related and unrelated variety (Frenken, Van Oort, and Verburg 2007).[4]

By definition, we classify four digit occupations exclusively under the three knowledge bases and thus can use these two levels to decompose diversity between and within knowledge bases. In our case, therefore, the share *SH* of knowledge base *b* in the total of all three knowledge bases, where *b=1,…,B,* can be derived by summing up the shares *sh* of all occupations *o*, where *o=1,…,O*, which are classified under the respective knowledge base K_b.[5]

(3) $$SH_b = \sum_{o \epsilon K_b} sh_o$$

Between diversity (*BD*), or the potential for combinations between knowledge bases, is given by:

(4) $$BD = \sum_{b=1}^{B} SH_b \log\left(\frac{1}{SH_b}\right)$$

Within diversity (*WD*), or the potential for combinations within knowledge bases, is given by:

(5) $$WD = \sum_{b=1}^{B} SH_b H_b$$

where:

(6) $$H_b = \sum_{o \epsilon K_b} \frac{sh_o}{SH_b} \log\left(\frac{1}{sh_o / SH_b}\right)$$

Population density is the log of the number of inhabitants divided by the area. We include population density to account for general urbanization (dis)economies in the sense of factors related to urbanization which are not accounted for by the knowledge base measures, and which can be either positive such as access to finance, access to markets and culture, or negative, due to various congestion effects such as high real estate prices, crime rates, and traffic jams (Parr 2002). Finally, we include border dummies for where the firm is located, differentiating between Swedish inland, the east north coast, the east south coast, the west

[4] The use of the entropy measure to capture between and within diversity of knowledge bases is different from measuring related and unrelated variety. While the related and unrelated variety measures are typically based on industry sector codes, the knowledge base typology cuts across industry classifications, since different knowledge bases can occur even in closely related sectors (e.g. a scientist and an engineer might work in the same industry, even though having different knowledge bases). Hence, the distinction between knowledge bases cut across the dichotomy of related and unrelated variety.

[5] Engineers and technicians of the same type are combined. This applies to the following occupation codes: 2131 and 2139; 2142 and 3112; 2143 and 3113; 2144 and 3114; 2145 and 3115; 2146 and 3116; 2147 and 3117; 2131 and 2139. A list of all occupation codes is provided in annex A1.

coast, the Norwegian border, and the Finnish border, because only domestic spillovers are taken into account.[6]

Table 2 presents descriptive statistics for the regional variables. Sweden is divided into 290 municipalities, which differ significantly in size, population, and population density. The municipalities in the urban centers around Stockholm, Gothenburg and Malmö are of much smaller size than the municipalities in the less populated northern part of Sweden. The smallest municipality is confined to less than 9 km^2 whereas the biggest extends to over 19,000 km^2. The lowest population is only 2,500 inhabitants, which is in strong contrast to Stockholm municipally with over 750,000 inhabitants. In order to account for these disparities, the regional variables consider spillovers based on the time-distance to other municipalities as explained above. Again, synthetic knowledge is most frequent, followed by symbolic knowledge and analytical knowledge. As analytical knowledge concentrates not only in firms, but also in universities and other public organizations, the differences are not as large as at the firm-level.

Table 2 Region-level variables

Variable	Observations	Mean	Std. dev.	Min	Max
Regional analytical knowledge concentration (%)	290	0.83	0.84	0.00	5.77
Regional synthetic knowledge concentration (%)	290	5.42	2.46	0.77	13.32
Regional symbolic knowledge concentration (%)	290	1.15	0.71	0.26	4.18
Regional total knowledge concentration (%)	290	7.40	3.56	1.04	16.86
Regional knowledge between diversity (eq 4)	290	0.70	0.14	0.31	0.99
Regional knowledge within diversity (eq 5)	290	1.82	0.15	1.26	2.14
Population density (log)	290	3.46	1.55	-1.38	7.16

Note: Reported is the average for all five CIS waves at $\lambda = 0.100$.

Figures 1-6 illustrate the spatial distribution of the regional variables by dividing the municipalities into quantiles. The lowest quantile is shaded in the lightest grey whereas the highest quantile is shaded in the darkest grey. Overall, we find strong concentrations of all three knowledge bases in the main urban centers including the wider region of Stockholm/Uppsala, Gothenburg as well as in the Malmö/Lund area (see Figures 1-3). Yet, there are differences between these three regions. Symbolic and analytical knowledge is concentrated especially in Stockholm/Uppsala, Gothenburg is characterized by a high level of synthetic knowledge whereas Malmö/Lund scores high in particular as regards analytical knowledge. The part of Sweden bordering Norway features, with a few exceptions, a relatively low concentration of all three knowledge bases. In the northeastern part of Sweden, we find that Umeå region is characterized by a high concentration of analytical knowledge. Synthetic knowledge dominates in the central part of Sweden, with the exception of Linköping, where also a high concentration of analytical knowledge can be observed. The

[6] For instance, there are likely to be strong cross-border knowledge spillovers between Sweden and Denmark, in particular in the Copenhagen-Malmö area, which are not reflected in the data. Hence, the border dummies are included to at least partly control for this source of bias.

between diversity is especially high in the wider Stockholm/Uppsala area, in Malmö/Lund as well as in Umeå (see Figure 4). The within diversity, in contrast, is relatively low in these regions, but high on the coast south of Stockholm, the coast north of Malmö, parts of central Sweden, and in the greater Gothenburg area (see Figure 5). The total knowledge concentration turns out to be particularly high in the three main urban centers, but also in some smaller regions such as Linköping and Umeå (see Figure 6).

Figure 1
Regional analytical concentrations

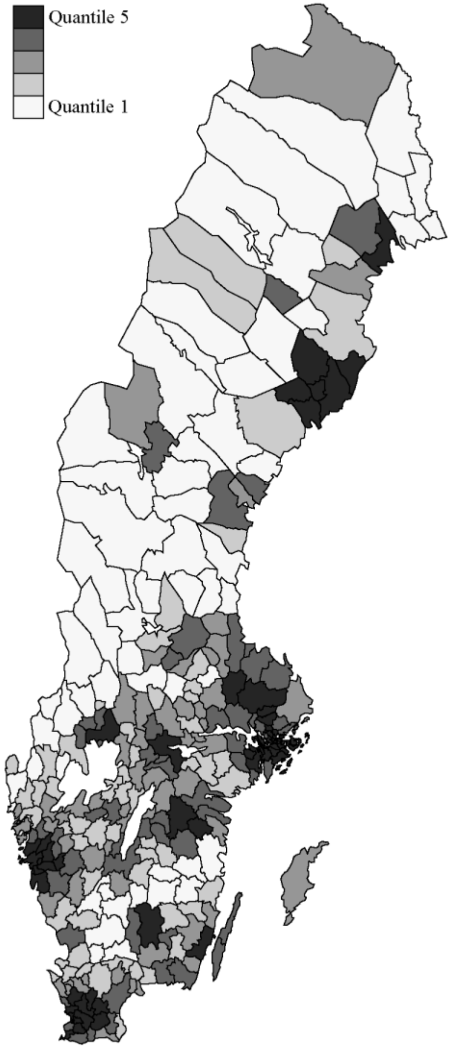


Figure 2
Regional synthetic concentrations

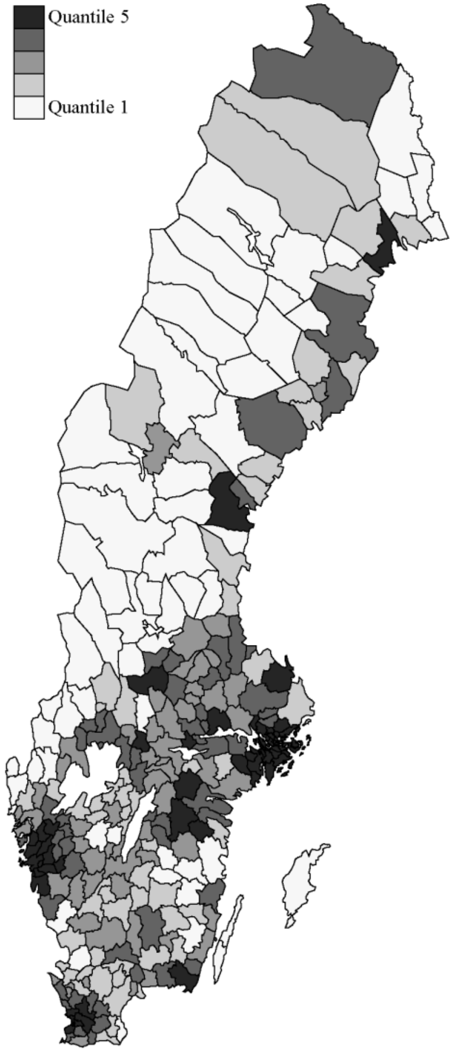


Figure 3
Regional symbolic concentrations

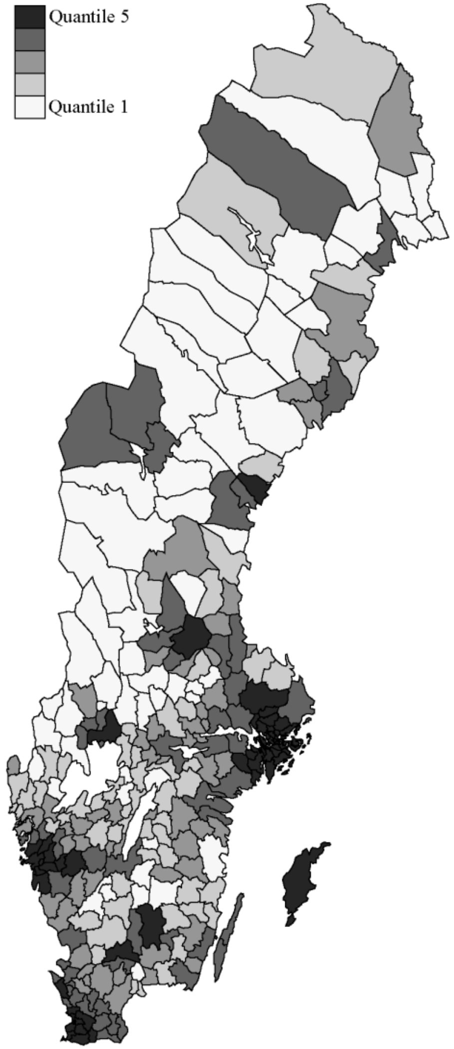


Figure 4
Between Diversity

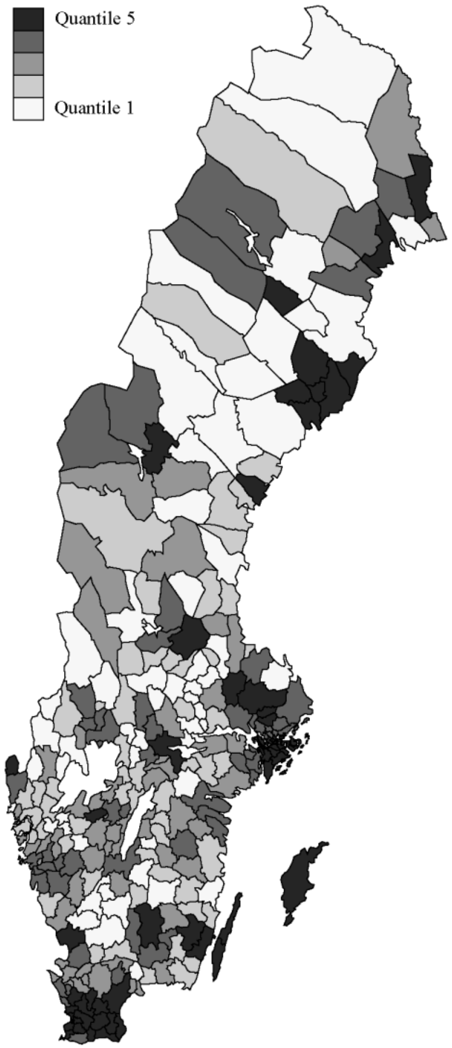


Figure 5
Within Diversity

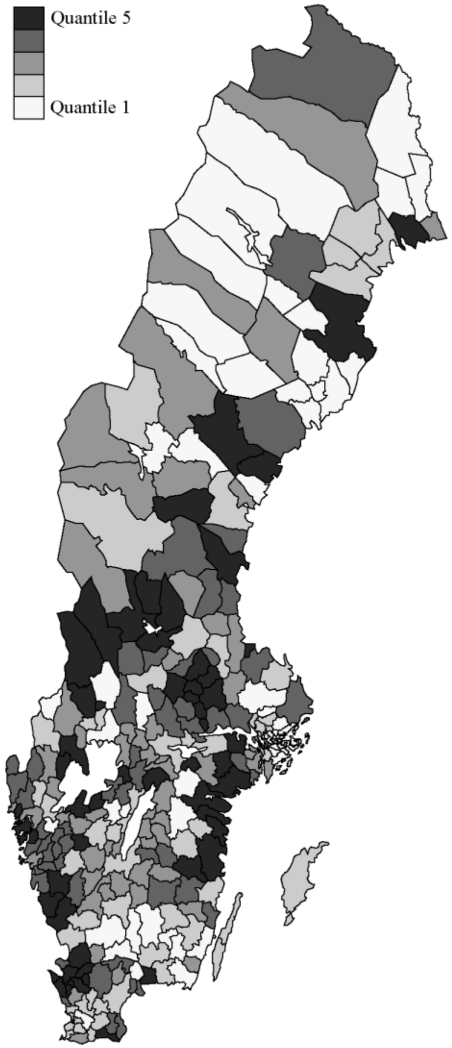
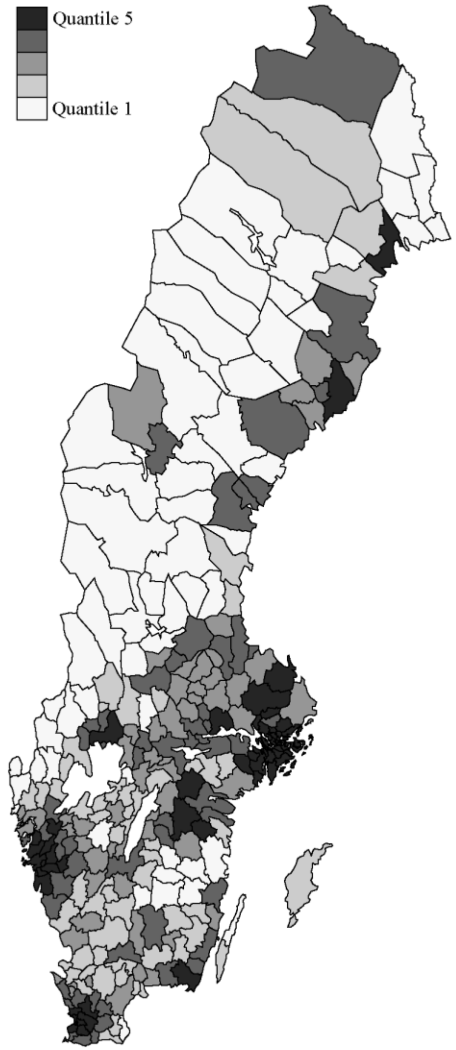


Figure 6
Regional total knowledge concentrations



4. Empirical Analysis

Economic geographers and regional scientists have been long interested in comparative regional analyses. Such regional comparisons often aim at examining various facets of agglomeration economies, regional institutions and policies, where the aggregate level is the appropriate unit of analysis. However, if the prime interest is in how regional factors affect a problem that is rooted in micro behavior, such as the innovativeness of firms, it may be problematic to project statistical inferences discovered at a higher level onto a lower level. Aggregated analyses that assume relationships observed at regional level equally hold for the firm-level may suffer from the so-called "ecological fallacy" (van Oort et al. 2012).

A complex contextual problem, such as the interplay between firm and regional knowledge bases on the one hand and innovation in firms on the other, cannot be fully understood at any single level of analysis (Beugelsdijk 2007; Srholec 2010). The success of a firm in innovation can be described as driven by interactions of firm internal and regional knowledge bases, and in particular the regional concentration, diversity and specialization of knowledge bases. However, as the firm and not the region is the unit of the analysis that innovates, the firm is at the center of our analytical framework. Hence, the general specification of the econometric model to be estimated relates a dependent variable at the firm-level to predictor variables at firm- and region-levels as follows:

(7) $$innov_{it} = \alpha KB_{jt} + \beta X_{jt} + \gamma kb_{it} + \delta x_{it} + \varphi w_k + \omega z_t + \varepsilon_{it}$$

where i is a firm, j is a region, k is an industry and t is time, so the innovation output of a firm ($innov_{it}$) is the function of the regional knowledge base (KB_{jt}), control characteristics of the region (X_{jt}), the knowledge base of the firm (kb_{it}), control characteristics of the firm (x_{it}), other relevant controls represented by industry (w_k), temporal shocks (z_t) and errors (ε_{it}).

The focal point is the estimated impact of KB_{jt}, kb_{it} and their cross-level interactions, which refer to the accessibility, combinations and diversity of knowledge bases. The vector of regional control variables X_{jt} includes the population density (in logs) and the border dummies, the vector x_{it} of firm-level controls consists of size given by the number of employees (in logs), the dummy for being affiliated to a group and the dummy for exporters, the battery of industry dummies w_k detects differences in technological opportunities and the set of time dummies z_t accounts for shocks occurring in specific periods.

Because the dependent variable, namely the percentage of new to the market innovations in turnover, is truncated by 0 from below and 100 from above, we use tobit models. To facilitate the interpretation of the results, we standardize the regional predictors by deducting the mean and dividing by the standard deviation, so these variables enter the estimate with mean of zero and standard deviation equal to one. After this transformation, we can directly compare the magnitude of the estimated regional effects because these predictors have a common scale of units of standard deviation. Marginal effects for the unconditional expected value of the dependent variable are reported; predictors are fixed at their means. Stata 13 has been used to perform the estimates (for more details see Stata 2013a, pp. 2391-2397).

Table 3 gives the base results using ordinary pooled tobit to estimate the model, thus the idiosyncratic errors ε_{it} are conventionally assumed to be i.i.d. N(0, σ_ε^2). Standard errors are clustered at the municipal level in order to admit the actual variation of interest. As benchmark, we allow for spillovers of the regional characteristics across municipal borders with λ = 0.100, which implies a relatively high decay. Nevertheless, the results are shown to be robust to the decay parameter, unless spillovers reaching significantly beyond immediate neighbours are allowed for. The specification of the model remains the same, except of the KB_{jt} variables, which are changing in order to test the impact of the different regional knowledge base characteristics.

Table 3 **The relationships of firm-level and regional knowledge bases (λ = 0.100) and innovativeness of firms (% of new to the market innovations in total turnover), pooled tobit**

	(1) Regional knowledge base accessibility	(2) Regional knowledge base diversity	(3) Regional within diversity by knowledge base
Regional analytical knowledge concentration (%)	0.269 (0.086)***	..	..
Regional synthetic knowledge concentration (%)	-0.025 (0.105)	..	..
Regional symbolic knowledge concentration (%)	0.085 (0.113)	..	..
Regional total knowledge concentration (%)	..	0.295 (0.164)*	0.336 (0.158)**
Regional knowledge between diversity (eq 4)	..	0.239 (0.108)**	0.341 (0.110)***
Regional knowledge within diversity (eq 5)	..	-0.024 (0.104)	..
Regional analytical knowledge within diversity (eq 6)	..	..	0.195 (0.105)*
Regional synthetic knowledge within diversity (eq 6)	..	..	-0.038 (0.101)
Regional symbolic knowledge within diversity (eq 6)	..	..	-0.020 (0.120)
Population density (log)	0.009 (0.217)	-0.186 (0.201)	-0.362 (0.216)*
Firm analytical knowledge (%)	0.081 (0.020)***	0.083 (0.019)***	0.082 (0.020)***
Firm synthetic knowledge (%)	0.049 (0.005)***	0.049 (0.005)***	0.049 (0.005)***
Firm symbolic knowledge (%)	-0.003 (0.007)	-0.004 (0.007)	-0.003 (0.008)
Number of employees (log)	0.289 (0.052)***	0.289 (0.052)***	0.289 (0.053)***
Member of a group (dummy)	0.415 (0.146)***	0.413 (0.146)***	0.414 (0.146)***
Foreign sales (dummy)	2.783 (0.152)***	2.787 (0.155)***	2.787 (0.155)***
Period dummies	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes
Border dummies	Yes	Yes	Yes
F	53.12***	55.13***	49.01***
AIC	58,321.97	58,325.70	58,324.17
BIC	58,536.01	58,539.74	58,554.06
Log (pseudo)likelihood	-29,133.98	-29,135.85	-29,133.08
Number of firms	10,110	10,110	10,110
Number of observations	20,482	20,482	20,482

Note: Marginal effects for the unconditional expected value of the dependent variable are reported; for binary variables the marginal effects refer to discrete change from 0 to 1; regional variables are standardized; standard errors clustered at the regional level are in brackets; ***, **, and * indicate significance at the 1, 5, and 10 percent level.

In the first column of Table 3, we consider the impact of the accessibility to analytical, synthetic and symbolic knowledge on the innovativeness of firms. On the one hand, the analytical knowledge base comes out with a positive and highly statistically significant coefficient, hence confirming the thesis that the regional access to this kind of knowledge is directly beneficial for firms. On the other hand, however, the regional synthetic and symbolic knowledge bases do not seem to make much difference, which indicates that their connection to firm-level innovation is somewhat more complex, as further vindicated by the cross-level interactions below.

Even though, moreover, the distinction between the knowledge bases is conceptually clear, they turn out to be more difficult to differentiate empirically, because workers carrying the various types of knowledge tend to collocate in the same areas. Not surprisingly, therefore, the three variables measuring the knowledge base concentrations tend to be correlated in the range from 0.50 to 0.70, which represents a challenge, when estimating their individual impacts. Nevertheless, the problem of multicollinearity is not serious enough to undermine the main findings, because the coefficient of analytical knowledge is highly statistically significant, and because the results come out qualitatively similar, if the variables are included in the model separately.

In the second column of Table 3, we examine the impact of the regional knowledge base diversity using the entropy indicators delineated in the previous section. The main outcome is that a location in regions with high between diversity, hence with a high potential for firms to combine the three knowledge bases, is conducive to the innovativeness of firms, while a location in regions that are only diversified within a particular kind of knowledge does not seem to pay off. It does not mean, of course, that under special conditions there cannot be outliers in this respect; however, the overall pattern in the data is that in order to facilitate innovation a region needs to nurture different kinds of knowledge bases. Generally speaking, the results strongly back the argument that regional diversity in knowledge bases positively affects the innovation performance of firms.

Given the magnitude of the marginal effects, all else equal to the mean, a firm that operates in a region with knowledge base between diversity one standard deviation above the mean, which roughly corresponds to Malmö, is estimated to achieve an about 0.24 percentage points higher share of innovative sales as compared to a firm located in a region, which exhibits average conditions. Moreover, a firm is estimated to gain just thanks to the broadly diversified knowledge base about 0.46 percentage points if located in Umeå, 0.45 percentage points in Uppsala and 0.30 percentage points in Stockholm; all three of which feature in the top decile as regards between diversity. At the first glance, this might not seem that much. But if compared to the sample mean of innovative sales of 3.83 percent, the estimated impact actually represents an increase by about one eighth in the top regions, which is a tangible contribution to the innovativeness of local firms.

Between diversity of knowledge bases tends to be negatively correlated with within diversity. Only three out of 30 municipalities in the top decile in terms of between diversity at the same time score above the mean on within diversity. Overall, only 65 municipalities score above the mean in both and merely two exceptions that prove the rule, namely Ystad and Simrishamn, score simultaneously in the top quartiles. Even the main urban centers of Stockholm and Malmö provide high between diversity but score relatively low on within diversity, thus tend to be rather specialized within. So the prevailing pattern is that a region excels in one or the other kind of diversity but very rarely in both of them.[7]

From this follows that a high between diversity of regional knowledge bases does not tend to materialize from developing everything across the board but rather by combining access to a broad between diversity with specialization in selected segments of each knowledge base. In other words, the most favorable environment for innovation in practice can be found in regions that have clearly defined strengths within each of the analytical, synthetic and symbolic knowledge bases, thus those regions not chasing too many rabbits at once, but those that facilitate combination between specific bits of the different knowledge bases. Hence, in this respect the results call for smart specialization within the different knowledge bases, while allowing for broad combinations between them.

In the third column of Table 3, we examine in more detail the impact of within diversity by splitting the overall measure derived from Equation (5) into the underlying three sub-indices delineated separately for each knowledge base in Equation (6). The results reveal that the inconclusive coefficient of the overall within diversity detected in the previous estimate conflates a weakly statistically significant positive impact of analytical knowledge within diversity and slightly negative but statistically insignificant impacts of the within diversity of synthetic and symbolic knowledge bases. In addition, the between diversity coefficient increases noticeably and turns out statistically significant at 1% level, so the main finding holds even stronger. From this result follows a refinement that the most beneficial environment for innovation is a strong between diversity of knowledge bases derived specifically from combining synthetic and symbolic specialization with analytical diversity, although the latter coefficient suffers from a relatively large margin of error, so this conclusion should not be overstated.

So far we have not discussed the firm-level knowledge base variables. All else equal to average, ten percentage points increase of analytical and synthetic knowledge that is at firm's disposal in-house is estimated to lead to about 0.8 and 0.5 percentage points increase in innovative sales, respectively. Thus the impact of the former is noticeably higher than the latter, which again highlights the essential role of analytical knowledge, even inside of the firms. Both of these coefficients come out highly statistically significant. Internal symbolic knowledge does not seem to be important, though as shown later, plays a role in combination with analytical knowledge. It is also noteworthy that the results of the firm-level knowledge variables are remarkably stable, regardless of the regional variables.

[7] Note that this is not the artefact of how the diversity measures are computed, as maximum between diversity can be achieved either by even distribution across the base categories, which at the same time entails maximum within diversity, or by concentrating everything in a single based category in each group, which in turn entails minimum within diversity. So by principle any combination of between and within diversity is feasible.

As far as the remaining control variables are concerned, the results are largely in line with the expectations. Population density comes out with a weakly statistically significant coefficient in the last estimate only, the negative sign of which indicates that essential aspects of urbanization economies have been successfully captured by the knowledge base variables, so what remains left for this variable to account for are the adverse congestion effects. All three structural features of the firm, namely the size, group and export variables, are as expected positive and highly statistically significant across the board, hence confirmed to be highly relevant to account for.

Appendix Table A2 reports the robustness of the estimates to higher and lower values of the spillover decay parameter λ. Only the results of the main variables of interest are presented for the sake of saving space, where the rows correspond to the columns in the previous table. A lower decay parameter implies a wider distribution of the respective knowledge in space, which evens out the regional differences, and in turn leads to higher correlation of the between and within diversity indexes. For example, the correlation coefficient is -0.13 for λ = 0.150, -0.19 for λ = 0.125, -0.30 for λ = 0.100, -0.45 for λ = 0.075, -0.64 for λ = 0.050 and already as much as -0.77 for λ = 0.025. Hence, λ < 0.050 is not considered due to multicollinearity concerns.[8]

Overall, the main results appear robust to the decay parameter. Interestingly, however, the estimated marginal effect of analytical knowledge accessibility decreases with λ, while the opposite tendency has been detected for both the between diversity and the within diversity of analytical knowledge. In other words, the relevance of the diversity measures seem to be confined to smaller areas, hence more local in nature, while the analytical knowledge accessibility impacts tend to spread much wider. Also the between diversity coefficient becomes statistically insignificant at the conventional levels in the second specification with low λ, as the multicollinerity starts to kick in. However, the results remain in favor of supporting that between diversity of regional knowledge bases is what primarily matters for innovation.

Admittedly, it is advisable to control not only for observable firm-level characteristics on which firm potentially sort out themselves in space, such as most prominently their internal knowledge bases, but also for their unobservable individual characteristics that may matter for the innovation output. For example, the latter include various time-invariant characteristics, which are not available in the data and/or difficult to measure, thus not properly accounted for in the estimate, such as the entrepreneurial spirit, latent capabilities or

---

[8] It should be further noted that λ = 0.150 leads to accessibility figures that are close to the original municipal level variables, i.e. without considering any spillovers altogether. More specifically, the correlation coefficients between the accessibility based on λ = 0.150 and the municipal variables are 0.70 and 0.63 for between and within diversity, respectively. The main difference is in the broader Stockholm area, where the municipalities are very small with short distances between each other and at the same time where knowledge bases are concentrated, so the spillovers remain strong even with a high decay.

risk profiles of firms. If these characteristics are not taken into account, the estimated coefficients may be biased, because of picking up their impacts.

Finally, therefore, we repeat the estimates with the help of panel data methods, namely the random effects tobit estimator. Note that estimating a fixed effects model that would be generally preferable is not a real option for us here, because the number of periods in the sample is rather limited, and thus some of the key predictors are nearly time-invariant. Hence, for this purpose, the composite error term is decomposed in two elements, i.e. $\varepsilon_{it} = \mu_i + \nu_{it}$, which include unobserved individual effects (μ_i) and other time-variant unobserved variables (ν_{it}), where μ_i, are conventionally assumed to be i.i.d. $N(0,\sigma_{\mu}^{2})$ independently of ν_{it} (for more details see Stata 2013b, pp. 430-437).

Table 4 provides the results. At the bottom of the table, the estimated unobserved individual effects (μ_i) are reported, i.e. the so-called variance components, and consequently ρ, which is the proportion of the total variance attributed to the individual component. If ρ is close to zero, the unobserved individual effects do not account for the outcome and thus the panel estimator is not more efficient than the pooled estimator, which is clearly not the case. A likelihood-ratio test has been performed whether ρ is different from zero confirms that the unobserved individual variance is quite substantial. From this follows that the random effects tobit is more efficient than ordinary pooled tobit.

Nevertheless, the random effects tobit estimator might not be consistent, if the underlying orthogonality assumptions do not hold, which is quite likely in this model. Not much could have been done about this directly, because valid instrumental variables are not available, which is admittedly a chronic problem for empirical research on innovation. As a crude indication to which extent this is a problem, it is instructive to compare results of the pooled and panel estimators. If the difference is negligible, the bias is likely to be small and vice versa. Fortunately, a cursory comparison reveals that the results give a qualitative very similar picture regardless of the estimator. The estimated coefficients come out very similar in magnitude and the levels of statistical significance are nearly identical. Hence, this source of bias seems to be largely inconsequential, which is reassuring.

Table 4 The relationships of firm-level and regional knowledge bases (λ = 0.100) and innovativeness of firms (% of new to the market innovations in total turnover), tobit with firm random effects

	(1) Regional knowledge base accessibility	(2) Regional knowledge base diversity	(3) Regional within diversity by knowledge base
Regional analytical knowledge concentrations (%)	0.289 (0.070)***	..	..
Regional synthetic knowledge concentrations (%)	-0.100 (0.119)	..	..
Regional symbolic knowledge concentrations (%)	0.071 (0.113)	..	..
Regional total knowledge concentrations (%)	..	0.229 (0.138)*	0.281 (0.130)**
Regional knowledge between diversity (eq 4)	..	0.235 (0.107)**	0.343 (0.107)***
Regional knowledge within diversity (eq 5)	..	-0.057 (0.110)	..
Regional analytical knowledge within diversity (eq 6)	..	..	0.178 (0.097)*
Regional synthetic knowledge within diversity (eq 6)	..	..	-0.041 (0.100)
Regional symbolic knowledge within diversity (eq 6)	..	..	-0.017 (0.133)
Population density (log)	0.070 (0.229)	-0.148 (0.163)	-0.321 (0.193)*
Firm analytical knowledge (%)	0.081 (0.013)***	0.084 (0.013)***	0.084 (0.013)***
Firm synthetic knowledge (%)	0.047 (0.004)***	0.047 (0.004)***	0.048 (0.004)***
Firm symbolic knowledge (%)	-0.005 (0.008)	-0.005 (0.008)	-0.005 (0.008)
Number of employees (log)	0.240 (0.054)***	0.239 (0.054)***	0.239 (0.054)***
Member of a group (dummy)	0.433 (0.156)***	0.432 (0.157)***	0.434 (0.156)***
Foreign sales (dummy)	2.547 (0.153)***	2.551 (0.153)***	2.550 (0.153)***
Period dummies	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes
Border dummies	Yes	Yes	Yes
Wald χ^2	1,212.39***	1,207.71***	1,210.64***
$\sigma(\mu)$	19.899 (0.513)***	19.900 (0.513)***	19.885 (0.513)***
ρ	0.393	0.393	0.392
AIC	57,564.20	57,569.63	57,570.37
BIC	57,786.16	57,791.59	57,808.19
Log (pseudo)likelihood	-28,754.10	-28,756.81	-28,755.19
Number of firms	10,110	10,110	10,110
Number of observations	20,482	20,482	20,482

Note: Marginal effects for the unconditional expected value of the dependent variable are reported; for binary variables the marginal effects refer to discrete change from 0 to 1; regional variables are standardized; standard errors are in brackets; ***, **, and * indicate significance at the 1, 5, and 10 percent level.

Next, we turn to the connection between in-house combinations of knowledge bases and the innovativeness of firms. Table 5 reports the marginal effects of firm internal analytical, synthetic and symbolic knowledge bases for firm innovativeness at certain levels of the respective other firm internal knowledge bases. This analysis addresses the research question about which types of firm internal knowledge bases and combinations thereof are most conducive to innovation. In order to investigate this question, we add interaction terms for the firm-level shares of analytical, synthetic and symbolic knowledge to the base models presented in the first column of Tables 3 and 4.

As regards the interplay between analytical and synthetic knowledge, we find that the marginal effect of analytical knowledge augments for higher levels of synthetic knowledge (column 1) and likewise, a higher marginal effect of synthetic knowledge is shown for higher levels of analytical knowledge (column 3). The results provide evidence for a substantial synergy effect between analytical and synthetic knowledge. All else being equal, a 10 percentage points increase of analytical knowledge leads to a 0.7 percentage points increase in sales of innovative products for firms that have a 5% share of synthetic knowledge whereas it leads to a 2.2 percentage points increase for firms with a 50% share of synthetic knowledge.

Further, the analysis reveals strong synergies between analytical and symbolic knowledge. Symbolic knowledge has no effect for firms without analytical knowledge. In sharp contrast, a 10 percentage points increase in symbolic knowledge is associated with a 13 percentage points increase in sales of innovative products for firms with a 50% share of analytical knowledge (column 5). Furthermore, the effect of analytical knowledge on firm innovativeness increases for higher levels of symbolic knowledge (column 2). However, no synergies can be observed between synthetic and symbolic knowledge. Symbolic knowledge remains insignificant at different levels of synthetic knowledge (column 6). Also synthetic knowledge has only a small marginal effect on the innovativeness of firms at different levels of symbolic knowledge. Due to the large standard error, this effect becomes insignificant at higher levels of symbolic knowledge (column 4).

Table 5 Marginal effects including interaction terms of firm-level knowledge bases

	Marginal effects of firm-level analytical knowledge		Marginal effects of firm-level synthetic knowledge		Marginal effects of firm-level symbolic knowledge	
at share of firm-level knowledge	(1) synthetic	(2) symbolic	(3) analytical	(4) symbolic	(5) analytical	(6) synthetic
Pooled tobit						
0%	0.062***	0.065***	0.047***	0.048***	-0.007	-0.002
	(0.015)	(0.019)	(0.005)	(0.005)	(0.009)	(0.009)
5%	0.073***	0.128***	0.061***	0.050***	0.060***	-0.001
	(0.0161)	(0.022)	(0.006)	(0.006)	(0.018)	(0.007)
25%	0.127***	0.384***	0.137***	0.059***	0.476***	0.008
	(0.029)	(0.078)	(0.037)	(0.022)	(0.128)	(0.0184)
50%	0.221***	0.709***	0.287**	0.069	1.381***	0.0241
	(0.063)	(0.170)	(0.114)	(0.044)	(0.407)	(0.051)
75%	0.347***	1.042***	0.496**	0.080	2.733***	0.047
	(0.113)	(0.284)	(0.228)	(0.067)	(0.830)	(0.098)
Tobit with firm random effects						
0%	0.064***	0.071***	0.046***	0.047***	-0.007	-0.004
	(0.015)	(0.015)	(0.004)	(0.004)	(0.009)	(0.008)
5%	0.074***	0.116***	0.057***	0.048***	0.041*	-0.003
	(0.014)	(0.021)	(0.006)	(0.005)	(0.022)	(0.008)
25%	0.120***	0.296***	0.121***	0.053***	0.335**	0.003
	(0.020)	(0.010)	(0.031)	(0.016)	(0.159)	(0.016)
50%	0.198***	0.519**	0.244***	0.060*	0.971**	0.014
	(0.048)	(0.208)	(0.089)	(0.033)	(0.468)	(0.040)
75%	0.301***	0.740**	0.413**	0.066	1.922**	0.029
	(0.092)	(0.327)	(0.176)	(0.050)	(0.934)	(0.076)

Note: Marginal effects for the unconditional expected value of the dependent variable are reported. Standard errors are in parentheses (clustered at the municipal level in the pooled tobit); ***, ** and * indicate significance at the 1%, 5% and 10% level. All other variables are kept at the mean.

Finally, we investigate interaction effects between firm- and region-level knowledge bases. Table 6 presents the marginal effects of the regional knowledge bases for firm innovativeness at certain levels of firm internal analytical, synthetic or symbolic knowledge.Similarly to the exercise above, we add the interaction terms between firm-level and region-level knowledge base shares to the base models presented in the first column of Tables 3 and 4.

On the one hand, regional concentrations of analytical knowledge have not only a prevailing positive effect on firm innovativeness, as already presented above, the marginal effects even increase for higher shares of firm-level analytical, synthetic or symbolic knowledge. The marginal effects are significant for all three knowledge base types with the exception of high levels of symbolic knowledge, for which standard errors are large. Not surprisingly, the firms with strong internal knowledge bases, thus with a high absorptive capacity, benefit more from the analytical knowledge available regionally.

On the other hand, a regional concentration of synthetic knowledge is negatively related to innovation performance of firms that have a high internal share of synthetic knowledge. As the traditional manufacturing regions are characterized by a high share of synthetic knowledge, this result hints to possible lock-in effects for local firms dominated by a synthetic knowledge base. In contrast, regional concentrations of synthetic knowledge have a significant positive effect on the innovativeness of firms with high levels of analytical knowledge. This indicates that firms dominated by an analytical knowledge base are relatively more innovative in traditional manufacturing regions. In other words, combining synthetic and analytical knowledge at the firm and regional levels, respectively, create positive synergies.

Table 6 Marginal effects including interaction terms between knowledge bases at the level of the firms and regions

	Marginal effects of regional analytical knowledge			Marginal effects of regional synthetic knowledge			Marginal effects of regional symbolic knowledge		
at share of firm-level knowledge	(1) analytical	(2) synthetic	(3) symbolic	(4) analytical	(5) synthetic	(6) symbolic	(7) analytical	(8) synthetic	(9) symbolic
Pooled Tobit									
0%	0.231**	0.179**	0.223**	-0.109	0.068	-0.035	0.113	0.109	0.102
	(0.090)	(0.080)	(0.096)	(0.114)	(0.112)	(0.117)	(0.115)	(0.099)	(0.116)
5%	0.262***	0.200**	0.253***	0.206	0.022	-0.103	0.028	0.106	0.098
	(0.083)	(0.083)	(0.082)	(0.159)	(0.111)	(0.109)	(0.121)	(0.105)	(0.113)
25%	0.402***	0.305***	0.373**	1.733**	-0.244**	-0.372	-0.388	0.084	0.078
	(0.117)	(0.106)	(0.153)	(0.755)	(0.124)	(0.239)	(0.317)	(0.134)	(0.127)
50%	0.626**	0.481***	0.521	4.306**	-0.769***	-0.706	-1.093	0.024	0.055
	(0.258)	(0.178)	(0.323)	(1.804)	(0.230)	(0.488)	(0.757)	(0.193)	(0.183)
75%	0.908**	0.711**	0.667	7.701**	-1.543***	-1.036	-2.028	-0.079	0.031
	(0.456)	(0.305)	(0.503)	(3.301)	(0.447)	(0.759)	(1.393)	(0.283)	(0.255)
Tobit with firm random effects									
0%	0.260***	0.202***	0.252***	-0.177	0.002	-0.107	0.075	0.070	0.067
	(0.075)	(0.077)	(0.077)	(0.125)	(0.121)	(0.127)	(0.118)	(0.107)	(0.120)
5%	0.271***	0.225***	0.275***	0.092	-0.049	-0.183	0.013	0.069	0.066
	(0.079)	(0.075)	(0.077)	(0.148)	(0.121)	(0.131)	(0.130)	(0.111)	(0.119)
25%	0.313	0.333***	0.364**	1.497***	-0.326**	-0.488	-0.314	0.059	0.064
	(0.221)	(0.085)	(0.172)	(0.482)	(0.150)	(0.312)	(0.339)	(0.136)	(0.174)
50%	0.362	0.511***	0.476	4.099***	-0.855***	-0.865	-0.925	0.029	0.062
	(0.562)	(0.163)	(0.333)	(1.182)	(0.288)	(0.609)	(0.855)	(0.202)	(0.298)
75%	0.402	0.741**	0.586	7.755***	-1.620***	-1.240	-1.786	-0.024	0.060
	(1.060)	(0.312)	(0.497)	(2.302)	(0.540)	(0.923)	(1.636)	(0.318)	(0.435)

Note: Marginal effects for the unconditional expected value of the dependent variable are reported. Standard errors are in parentheses (clustered at the municipal level in the pooled tobit); ***, ** and * indicate significance at the 1%, 5% and 10% level. All other variables are kept at the mean.

5. Conclusions

Overall, the results confirm that the knowledge base typology provides a fruitful analytical framework to differentiate between various types of knowledge that enter into the innovation process. While analytical, synthetic and symbolic knowledge bases are distinct categories with regards to the nature and the rationale for knowledge creation, the estimates strongly support the combinatorial thesis proposed by Manniche (2012) and Strambach and Klement (2012) that innovation is stimulated by cross-fertilization between more than one knowledge base. More specifically, the key findings are fourfold.

First, as far as the firm internal combinations are concerned, the econometric results are in line with the qualitative findings of Moodysson (2008) and Martin and Moodysson (2011) that firms use different types of knowledge in innovation. In particular, we find strong synergies between analytical and synthetic as well as analytical and symbolic knowledge bases. Symbolic knowledge turns out to be a catalyst of innovation only if combined with analytical knowledge. The results, therefore, provide a more differentiated picture about which combinations matter most and refine the argument typically made in the knowledge base literature, see, for instance, Asheim, Boschma, and Cooke (2011), that all three knowledge bases are equally relevant. Even though this thesis might be valid from an epistemological perspective, it does not hold true with regards to the impact on innovation outcome.

Second, the results back the argument made by Asheim and Gertler (2005) and others that these complementary knowledge bases are not only present within the firm but also sourced from their regional milieu. In this respect, the main finding is that between diversity, that is, the regional potential for broad combinations across the three knowledge bases, trumps within diversity, that is, the regional potential to combine different skills and competences within one knowledge base. Nevertheless, the underlying evidence does not collapse into a "more of everything" kind of advice, which is arguably not very insightful for policy guidance, as the resources at hand are always limited. Instead, the results call for a strategy of smart specialization within the knowledge bases, while allowing for diversified combinations between them. Regional policy should aim at promoting interfaces between the knowledge bases. This resonates well with platform policies, as advocated by Asheim, Boschma, and Cooke (2011) or smart specialization policies (Foray 2009; Foray, David, and Hall 2011), which are likely to stimulate regional innovation and growth, if carefully customized to the specific regional context.

Third, this paper sheds new light on how innovation performance is influenced by the interplay between knowledge bases at the level of the firm and those available in the region. The results indicate that access to a strong analytical knowledge base in the region, provided typically by a strong science and higher education system, has a direct positive impact, while regional synthetic and symbolic knowledge only makes a difference for innovation performance depending on firm internal knowledge bases. One finding that meets the eye is that there is a tendency for adverse effects, hence a lock-in situation, if both firms as well as other regional actors specialize in synthetic knowledge. This finding resonates well with literature on old industrial regions (Grabher 1993; Tödtling and Trippl 2004; Hassink 2005), which again calls for a strategy of knowledge base diversification, especially in manufacturing regions. Moreover, symbolic knowledge regardless of whether firm internal or present within a region boosts innovation in combination with other knowledge bases but not on its own. Hence, firms need a solid analytical and/or synthetic knowledge base in order to benefit from the injection of symbolic knowledge for innovation.

Admittedly, policy makers need to understand the cross-level interactions of knowledge bases if they are to be successful at promoting innovation. On one hand, policies should strengthen firm internal knowledge bases that enable firms to benefit more from the knowledge available in the region. On the other hand, policy should furnish firms with the type of knowledge that they are lacking internally in order to forge the most productive combinations. After all, the regional conditions have a tangible impact, but at the same time much depends on what firms are capable of doing themselves. One can at least partly compensate one for the other, but the most powerful forces boosting innovation are unleashed with their joint effects.

It is fully acknowledged that knowledge bases are still measured imperfectly by occupational data. Nevertheless, the information on the type of work an individual is performing, and the level of skill usually required to perform this work, is probably as close as it gets given the existing data for inferring on knowledge bases. A more fine-grained account of the underlying knowledge bases remains the sanctuary of detailed case studies, though; we need quantitative evidence on this topic in order to forge the much needed synergy between different methodological approaches that is often unexploited in the literature. Another major limitation is that despite best efforts to reduce the omitted variables problem, there are inherent endogeneity issues that could not be tackled directly, because valid instruments are not available in the data, which is admittedly a chronic problem for empirical research on innovation. Hence, one needs to be careful to infer on causality. It remains a challenge for future research to address these caveats as soon as even richer data become available in the future.

Acknowledgements: Financial support from VINNOVA Core Funding of Centers for Innovation Systems Research project 2010-01370 on 'Transformation and Growth in Innovation Systems: Innovation Policy for Global Competitiveness of SMEs and R&I Milieus' is gratefully acknowledged. An earlier version of the article was presented at the workshop "Special Issue of Economic Geography: Combinatorial Knowledge Bases, Regional Innovation and Development Dynamics", CIRCLE, Lund, May 13th-14th, 2014, The article has benefited from comments and suggestions from Koen Frenken and Sverre Herstad. All the usual caveats apply.

References

Arrow, K.J. 1962. Economic Welfare and the Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, ed. R. Nelson, 609-626. Princeton: Princeton University Press.

Asheim, B.T. 1996. Industrial districts as 'learning regions': A condition for prosperity. *European Planning Studies* **4:379-400**.

Asheim, B.T.; L. Coenen; and J. Vang. 2007. Face-to-face, buzz, and knowledge bases: sociospatial implications for learning, innovation, and innovation policy. *Environment and Planning C: Government and Policy* **25:655-670**.

Asheim, B.T.; B. Ebersberger; and S.J. Herstad. 2012. MNCs between the local and the global: knowledge bases, proximity and distributed knowledge networks. In *Innovation and institutional embeddedness of multinational companies*, ed. M. Heidenreich, 77-104. Cheltenham: Edward Elgar Publishing.

Asheim, B.T. and M.S. Gertler. 2005. The Geography of Innovation: Regional Innovation Systems. In *The Oxford Handbook of Innovation*, ed. J. Fagerberg; D.C. Mowery; and R.R. Nelson, 291-317. Oxford: Oxford University Press.

Asheim, B.T. and H.K. Hansen. 2009. Knowledge Bases, Talents, and Contexts: On the Usefulness of the Creative Class Approach in Sweden. *Economic Geography* **85:425-442**.

Aslesen, H.W. and M. Freel. 2012. Industrial Knowledge Bases as Drivers of Open Innovation? *Industry and Innovation* **19:563-584**.

Aydalot, P. 1986. *Milieux innovateurs en Europe*. Paris: GREMI.

Becattini, G. 1989. Sectors and/or Districts: Some Remarks on the Conceptual Foundations of Industrial Economics. In *Small Firms and Industrial Districts in Italy*, ed. E.J. Goodman; J. Bamford; and P. Saynor, 123–135. London and New York: Routlege.

Bellandi, M. 1989. The Industrial District in Marshall. In *Small Firms and Industrial Districts in Italy*, ed. E.J. Goodman; J. Bamford; and P. Saynor, 136–152. London and New York: Routlege.

Beugelsdijk, S. 2007. The Regional Environment and a Firm's Innovative Performance: A Plea for a Multilevel Interactionist Approach. *Economic Geography* **83:181-199**.

Boschma, R. and S. Iammarino. 2009. Related Variety, Trade Linkages, and Regional Growth in Italy. *Economic Geography* **85:289-311**.

Brusco, S. 1986. Small firms and industrial districts: the experience of Italy. In *New firms and regional development in Europe*, ed. D. Keeble and E. Wever, 184-202. London: Croom Helm.

Camagni, R. 1991. Innovation networks: spatial perspectives / edited by Roberto Camagni on behalf of GREMI (Groupe de recherche européen sur les milieux innovateurs). London and New York: Belhaven Press.

Cooke, P.; M. Heidenreich; and H.-J. Braczyk. 2004. Regional Innovation Systems: The Role of Governance in a Globalized World. New York: Routledge.

Cooke, P.; M.G. Uranga; and G. Etxebarria. 1998. Regional systems of innovation: an evolutionary perspective. *Environment and Planning A* **30:1563-1584**.

Foray, D. 2009. Understanding smart specialisation. In *The Question of R&D Specialisation, Perspectives and policy implications*, ed. D. Pontikakis; D. Kyriakou; and R. Van Bavel, 19-28: Joint Research Center, European Commission, Brussels.

Foray, D.; P.A. David; and B.H. Hall. 2011. Smart specialisation, From academic idea to political instrument, the surprising career of a concept and the difficulties involved in its implementation. *MTEI Working Paper, Management of Technology & Entrepreneurship Institute, École Polytechnique Fédéderale de Lausanne*.

Frenken, K.; F. Van Oort; and T. Verburg. 2007. Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies* **41:685-697**.

Gertler, M.S. 2008. Buzz without being there? Communities of practice in context. In *Community, Economic Creativity, and Organization*, ed. A. Amin and J. Roberts, 203-226. Oxford: Oxford University Press.

Grabher, G. 1993. The weakness of strong ties; the lock-in of regional development in the Ruhr area. In *The Embedded Firm: On the Socioeconomics of Industrial Networks*, ed. G. Grabher, 255-277. London & New York: Routledge.

Grabher, G. 2002. The Project Ecology of Advertising: Tasks, Talents and Teams. *Regional Studies* **36:245-262**.

Grillitsch, M. and M. Trippl. 2014. Combining Knowledge from Different Sources, Channels and Geographical Scales. *European Planning Studies* **22:2305-2325**.

Hassink, R. 2001. The Learning Region: A Fuzzy Concept or a Sound Theoretical Basis for Modern Regional Innovation Policies? *Zeitschrift für Wirtschaftsgeographie* **45:219-230**.

Hassink, R. 2005. How to unlock regional economies from path dependency? From learning region to learning cluster. *European Planning Studies* **13:521-535**.

Herstad, S.J.; H.W. Aslesen; and B. Ebersberger. 2014. On industrial knowledge bases, commercial opportunities and global innovation network linkages. *Research Policy* **43:495-504**.

Jensen, M.B.; B. Johnson; E. Lorenz; and B.-Å. Lundvall. 2007. Forms of knowledge and modes of innovation. *Research Policy* **36:680-693**.

Knorr-Cetina, K. 1999. *Epistemic cultures. How the sciences make knowledge*. Cambridge: Harvard University Press.

Laestadius, S. 1998. Technology Level, Knowledge Formation and Industrial Competence in Paper Manufacturing. In *The Micro Foundations of Economic Growth*, ed. G. Eliasson and C. Green, 212-226. Ann Arbour: University of Michigan Press.

Maillat, D.; B. Lecoq; F. Nemeti; and M. Pfister. 1995. Technology District and Innovation: The Case of the Swiss Jura Arc. *Regional Studies* **29:251-263**.

Malmberg, A. 1996. Industrial geography: agglomeration and local milieu. *Progress in Human Geography* **20:392-403**.

Malmberg, A. and P. Maskell. 2002. The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering. *Environment and Planning A* **34:429-449**.

Manniche, J. 2012. Combinatorial Knowledge Dynamics: On the Usefulness of the Differentiated Knowledge Bases Model. *European Planning Studies* **20:1823-1841**.

Marshall, A. 1920. *Principles of Economics*. London: Macmillan.

Martin, R. 2012. Measuring Knowledge Bases in Swedish Regions. *European Planning Studies* **20:1569-1582**.

Martin, R. and J. Moodysson. 2011. Innovation in symbolic industries: The geography and organization of knowledge sourcing. *European Planning Studies* **19:1183-1203**.

Martin, R. and J. Moodysson. 2013. Comparing knowledge bases: on the geography and organization of knowledge sourcing in the regional innovation system of Scania, Sweden. *European Urban and Regional Studies* **20:170-187**.

Martin, R. and M. Trippl. 2014. Regional Variety in Knowledge Bases and the Evolution of the ICT Cluster in Southern Sweden. *Paper presented at The Geography of Innovation Conference in Utrecht, January 2014*.

Maskell, P. and A. Malmberg. 1999. Localised learning and industrial competitiveness. *Cambridge Journal of Economics* **23:167-185**.

Moodysson, J. 2008. Principles and Practices of Knowledge Creation: On the Organization of "Buzz" and "Pipelines" in Life Science Communities. *Economic Geography* **84:449-469**.

Moodysson, J.; L. Coenen; and B.T. Asheim. 2008. Explaining spatial patterns of innovation: analytical and synthetic modes of knowledge creation in the Medicon Valley life-science cluster. *Environment and Planning A* **40:1040-1056**.

Morgan, K. 1997. The Learning Region: Institutions, Innovation and Regional Renewal. *Regional Studies* **31:491 - 503**.

Moulaert, F. and F. Sekia. 2003. Territorial innovation models: a critical survey. *Regional Studies* **37:289-302**.

OECD. 2005. *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition*. Paris: OECD Publishing.

Parr, J.B. 2002. Agglomeration economies: ambiguities and confusions. *Environment and Planning A* **34:717-731**.

Pavitt, K. 1984. Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy* **13:343-373**.
Plum, O. and R. Hassink. 2011. Comparing knowledge networking in different knowledge bases in Germany. *Papers in Regional Science* **90:355-371**.
Romer, P.M. 1986. Increasing Returns and Long-Run Growth. *The Journal of Political Economy* **94:1002-1037**.
Scott, A.J. 2004. Cultural-Products Industries and Urban Economic Development: Prospects for Growth and Market Contestation in Global Context. *Urban Affairs Review* **39:461-490**.
Srholec, M. 2010. A Multilevel Approach to Geography of Innovation. *Regional Studies* **44:1207-1220**.
Srholec, M. and B. Verspagen. 2012. The Voyage of the Beagle into innovation: explorations on heterogeneity, selection, and sectors. *Industrial and Corporate Change* **21:1221-1253**.
Stata. 2013a. Stata base reference manual, release 13. College Station, Texas: A Stata Press Publication.
Stata. 2013b. Stata longitudinal-data/panel-data reference manual, release 13. College Station, Texas: A Stata Press Publication.
Storper, M. 1995. The Resurgence of Regional Economies, Ten Years Later. *European Urban and Regional Studies* **2:191-221**.
Strambach, S. and B. Klement. 2013. Exploring plasticity in the development path of the automotive industry in Baden-Württemberg: the role of combinatorial knowledge dynamics. *Zeitschrift für Wirtschaftsgeographie* **57:67–82**.
Tödtling, F. and M. Grillitsch. 2014. Types of Innovation, Competencies of Firms, and External Knowledge Sourcing—Findings from Selected Sectors and Regions of Europe. *Journal of the Knowledge Economy* **5:330-356**.
Tödtling, F. and M. Trippl. 2004. Like Phoenix from the Ashes? The Renewal of Clusters in Old Industrial Areas. *Urban Studies* **41:1175-1195**.
Trippl, M.; F. Tödtling; and L. Lengauer. 2009. Knowledge Sourcing Beyond Buzz and Pipelines: Evidence from the Vienna Software Sector. *Economic Geography* **85:443-462**.
van Oort, F.G.; M.J. Burger; J. Knoben; and O. Raspe. 2012. Multilevel Approaches and the Firm-Agglomeration Ambiguity in Economic Growth Studies. *Journal of Economic Surveys* **26:468-491**.
Walsh, V. 1996. Design, innovation and the boundaries of the firm. *Research Policy* **25:509-529**.
Wenger, E. 1998. *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.
Zukauskaite, E. and J. Moodysson. 2013. Multiple Paths of Development: Knowledge Bases and Institutional Characteristics of the Swedish Food Sector. *CIRCLE Electronic Working Paper Series* **WP2013/46**.

Appendix

Table A1: **Occupation groups with analytical, synthetic and symbolic knowledge base**

Occupations group (SSYK 96)	
Analytical occupations	
2111	Physicists and astronomers
2112	Meteorologists
2113	Chemists
2114	Geologists and geophysicists
2121	Mathematicians and related professionals
2122	Statisticians
2131	Computer systems designers, analysts and programmers with PhD degree*
2139	Computing professionals not elsewhere classified
2211	Biologists, botanists, zoologists and related professionals
2212	Pharmacologists, pathologists and related professionals
2213	Agronomists and related professionals
2310	College, university and higher education teaching professionals
Synthetic occupations	
2131	Computer systems designers, analysts and programmers without PhD degree*
2142	Civil engineers
2143	Electrical engineers
2144	Electronics and telecommunications engineers
2145	Mechanical engineers
2146	Chemical engineers
2147	Mining engineers, metallurgists and related professionals
2148	Cartographers and surveyors
2149	Architects, engineers and related professionals not elsewhere classified
3111	Chemical and physical science technicians
3112	Civil engineering technicians
3113	Electrical engineering technicians
3114	Electronics and telecommunications engineering technicians
3115	Mechanical engineering technicians
3116	Chemical engineering technicians
3117	Mining and metallurgical technicians
3118	Draughtspersons
3119	Physical and engineering science technicians not elsewhere classified
Symbolic occupations	
2141	Architects, town and traffic planners
2431	Archivists and curators
2451	Authors, journalists and other writers
2452	Sculptors, painters and related artists
2453	Composers, musicians and singers
2454	Choreographers and dancers
2455	Film, stage and related actors and directors
2456	Designer
3131	Photographers and image and sound recording equipment operators
3471	Decorators and commercial designers
3472	Radio, television and other announcers
3473	Street, night-club and related musicians, singers and dancers
3474	Clowns, magicians, acrobats and related associate professionals
3476	Stage managers, prop masters, etc.

Note: The Swedish classification of occupational groups SSYK 96 builds on the International Standard Classification of Occupations ISCO-88.

* Investigating job descriptions, we found that the category “Computer systems designers, analysts and programmers” was on the border line between analytical and synthetic knowledge. We therefore combined the occupational code with educational information and classified those with PhD education as analytical and those without as synthetic.

Table A2: The relationship of regional knowledge base and innovativeness of firms by the spillover decay parameter (λ), pooled tobit

	(1) λ = 0.150		(2) λ = 0.125		(3) λ = 0.075		(4) λ = 0.050	
Regional knowledge base accessibility:								
Regional analytical knowledge concentration (%)	0.251	(0.075)***	0.260	(0.080)***	0.281	(0.099)***	0.285	(0.129)**
Regional synthetic knowledge concentration (%)	-0.020	(0.100)	-0.016	(0.102)	-0.054	(0.109)	-0.068	(0.121)
Regional symbolic knowledge concentration (%)	0.038	(0.094)	0.066	(0.103)	0.081	(0.122)	0.066	(0.131)
Regional knowledge base diversity:								
Regional knowledge between diversity (eq 4)	0.263	(0.105)**	0.256	(0.106)**	0.220	(0.112)**	0.167	(0.121)
Regional knowledge within diversity (eq 5)	-0.005	(0.105)	-0.013	(0.104)	-0.032	(0.109)	-0.073	(0.119)
Regional within diversity by knowledge base:								
Regional knowledge between diversity (eq 4)	0.356	(0.115)***	0.361	(0.115)***	0.311	(0.103)***	0.267	(0.102)***
Regional analytical knowledge within diversity (eq 6)	0.180	(0.104)*	0.198	(0.106)*	0.173	(0.104)*	0.120	(0.109)
Regional synthetic knowledge within diversity (eq 6)	-0.039	(0.104)	-0.044	(0.102)	-0.020	(0.104)	-0.031	(0.124)
Regional symbolic knowledge within diversity (eq 6)	-0.043	(0.118)	-0.035	(0.120)	0.001	(0.120)	0.049	(0.121)

Note: Marginal effects for the unconditional expected value of the dependent variable are reported are reported; for binary variables the marginal effects refer to discrete change from 0 to 1; regional variables are standardized; standard errors clustered at the regional level are in brackets; ***, **, and * indicate significance at the 1, 5, and 10 percent level.