Papers in Evolutionary Economic Geography

20.15

Related variety, recombinant knowledge and regional innovation.
Evidence for Sweden, 1991-2010

Mikhail Martynovich and Josef Taalbi

Utrecht University Human Geography and Planning

RELATED VARIETY, RECOMBINANT KNOWLEDGE AND REGIONAL INNOVATION. EVIDENCE FOR SWEDEN, 1991-2010

Mikhail Martynovich
mikhail.martynovich@keg.lu.se
Department of Human Geography, Lund University

Josef Taalbi
josef.taalbi@ekh.lu.se
Department of Economic History, Lund University

Abstract

This study investigates how related variety in the regional employment mix affects the innovation output of a region. Departing from the idea of recombinant innovation, previous research has argued that related variety enhances regional innovation as inter-industry knowledge spillovers occur more easily between different but cognitively similar industries. This study combines a novel dataset and related variety measures based on network theory, which allows a more nuanced perspective on the relationship between related variety and regional innovation. The principal novelty of the paper lies in employing new data on product innovations commercialised by Swedish manufacturing firms between 1970 and 2013. In this respect, it allows a direct measure of regional innovation output as compared to patent measures, usually employed in similar studies. The second contribution of this paper is that we employ network-topology based measures of related variety that allow us to measure relatedness as the recombination rather than direct flow of knowledge. We argue that this measure comes closer to the notion of innovation as spurred by recombination and show that this measure is a superior predictor of innovation activity.

Keywords: related variety, relatedness, innovation, network analysis

JEL codes: L16, O31, R11, R12

1. Introduction

It is widely acknowledged that (re)combining previously unconnected ideas in a new manner is key to the production of new knowledge and technological innovation (Arthur 2007). Stemming from Schumpeter's famous 'Neue Kombinationen' (Schumpeter 1934), this idea has been conceptualised in the notion of recombinant innovation as 'the way that old ideas can be reconfigured in new ways to make new ideas' (Weitzman 1998, p. 333). Recombination underlines the importance for firms to tap into heterogeneous knowledge sources to enhance their creative potential and innovation ability (Rodan and Galunic 2004).

As innovation processes draw on knowledge that is mainly sourced locally (Almeida and Kogut 1999, Breschi and Lissoni 2009), firms located in regions with diversified industry structures have better preconditions for recombinant innovation, mirroring the diversification argument put forward by Jacobs (1969). Bringing this argument further, regional industry composition should be an important determinant of regional innovation output.

Recent contributions within evolutionary economic geography have demonstrated, however, that not all regional diversity is equally conducive to innovation since not all existing knowledge can be easily (re)combined (Noteboom 2000, Boschma et al. 2009). To support efficient knowledge exchange and recombinant innovation, the regional variety of industries should be related, that is, consist of industries with cognitively close knowledge bases. On the other hand, fundamentally different knowledge bases in unrelated industries make inter-industry knowledge transfer difficult (Noteboom 2000). Put forward by Frenken et al. (2007), this hypothesis has inspired a large number of studies, which (with some reservations) provide empirical support to the claim that related variety fosters growth in regional employment, value added, and productivity (for a literature review, see Content and Frenken 2016).

While many of these studies conclude that related variety facilitates knowledge spillovers at a regional level, thus providing tentative evidence for its relationship with innovation, not much work has been done on investigating the impact of related variety on regional innovation performance as such. Notable exceptions include studies on regional patenting in Sweden (Tavassoli and Carbonara 2014) and US (Castaldi et al. 2015), as well as investigations into firm innovation activity in Italy (Antonietti and Cainelli 2011) and Norway (Aarstad et al. 2016). We contribute to this literature by estimating the knowledge production function for Swedish regions.

The principal novelty of the paper lies in employing a recently developed innovation database (SWINNO) containing extensive information about product innovations commercialised by Swedish manufacturing firms between 1970 and 2013 (Sjöö et al. 2014, Kander et al. 2019). This allows us to employ a more accurate measure of regional innovation output as compared to patent measures, usually employed in the previous studies.

Secondly, while previous studies employ cross-section data or short panels, our investigation period spans the 20-year period between 1991 and 2010. Given that industrial structures of regions change rather slowly over time (Firgo and Mayerhofer 2017), this allows us to arrive to a more accurate estimation of the relationship between related variety and innovation output of regions.

Another important aspect is that we construct new measures of related variety based on relatedness as recombination of knowledge. The conventional approach derives measures of related variety from the hierarchical structure of official industry classifications. Its fundamental weakness is that, by assuming cognitive similarity to exist only between industries sharing some digits in industry classifications, it underestimates the span of knowledge spillover channels between industries (Firgo and Mayerhofer 2017, Kuusk and Martynovich 2018). Furthermore, as industry classifications are fixed, such measures of related variety disregard potential intertemporal dynamics in the relatedness linkages between various industries in the process of technological development (Martynovich 2016, Kuusk and Martynovich 2018), which makes them inappropriate for studying the impact of industry structure on regional innovation output over time.

Instead, we follow Neffke and Henning (2013) in assuming that the existence of extensive labour flows between two industries signals that they are related and derive related variety measures from the economy-wide directed network of inter-industry worker flows. However, we take this approach further by considering not only *directly related* industries, but also those which are related through triadic closure, i.e. indirectly connected through another industry. Specifically, we construct relatedness measures based on the *recombination of skills*. We argue that taking into account indirect recombination of skills brings us closer to the original theory in which innovation is spurred by recombination of knowledge and allows a more nuanced perspective on the relationship between regional industry structure and innovation output.

We find broad support for the notion that related variety matters for innovation output performance across Swedish regions. Our contribution is twofold. First, our study makes an empirical contribution to the literature by providing the set of related variety measures that outperform those conventionally employed in the literature. Second, we demonstrate that estimating regional knowledge production functions based on the notion of recombinant innovation require the measures that capture the recombination of knowledge explicitly. In that respect, we suggest that previous efforts of investigating the role of related variety for regional innovation were based on a restricted notion of relatedness.

2. Previous research

2.1. Related variety and regional innovation

The idea that regional industrial diversity is conducive to innovation has its roots in the theory of agglomeration economies, particularly with respect to localised knowledge spillovers. As these come in different flavours, much research in the field of geography of innovation (often referred to as 'MAR vs. Jacobs' debate) has focused on whether specialisation or diversity enhances local innovation.

Here, MAR refers to the theories of Marshall, Arrow, and Romer that suggest knowledge spillovers to take place predominantly between similar economic activities and give rise to localisation economies as a source of regional innovation and growth. In contrast, Jacobs (1969) claims that regional growth is fuelled by the recombinant process of knowledge generation that builds on a pre-existing variety of knowledge that is combined in new ways. In their concept of nursery cities, Duranton and Puga (2001), Duranton and Puga (2004) support this view, concluding that heterogeneity of workers and firms is the foundation of the mechanisms behind agglomeration economies, identified as sharing, matching, and learning. Empirical evidence on the issue has remained largely inconclusive (Beaudry and Schiffauerova 2009, de Groot et al. 2016), leading to the claims that theoretical notions of specialisation and diversity are too simplistic to capture the relationship between regional industry structure and regional performance (Content and Frenken 2016).

In an attempt to resolve this empirical controversy, Frenken et al. (2007) suggested to decompose industrial diversity into variety in cognitively similar ("related variety") and distant ("unrelated variety") industries. They proposed that related variety fosters regional performance as inter-industry knowledge spillovers occur more easily between different but cognitively connected industries. As pointed by Boschma (2005, p. 64), "a not too great cognitive distance between firms (in terms of competencies and skills) enables effective communication and thus learning, while a not too small cognitive distance avoids lock-in, especially when access to dissimilar bodies of knowledge is required in product innovation". On the other hand, "at a certain point cognitive distance becomes so large as to preclude sufficient mutual understanding needed to utilize those opportunities" (Noteboom et al. 2007, p. 1017). In this respect, it is the optimal level of cognitive proximity that allows for interaction and innovation (Noteboom 2000, Hassink et al. 2014). Related variety captures the concept of cognitive distance between the extremes which allows economic actors to identify the potential for recombining complementary knowledge and resources and eventually facilitate innovation activity (Aarstad et al. 2016).

The related variety hypothesis has inspired a large number of empirical studies on the relationship between related variety and regional economic performance. These studies have provided the general support for the claim that related variety fosters growth in regional employment, value added, and productivity (Content and Frenken 2016). Since innovation is related to changes in both productivity and employment, this may be indicative of existence of the relationship between related variety and the innovative performance of regions. While the studies directly investigating the latter relationship have not been plentiful, they have informed the research with some important aspects of the relationship between related variety and innovation output.

A study of regional patenting in Sweden by Tavassoli and Carbonara (2014) has demonstrated that knowledge variety per se does not substantially affect regional innovativity, but it has a robust positive impact if it is a 'related' variety. Studying US regional patent data, Castaldi et al. (2015) conclude that related variety raises the likelihood of innovations in general, while unrelated variety raise the likelihood of breakthrough innovations. Miguelez and Moreno (2018) arrive to a similar conclusion investigating patenting activity in European NUTS-2 regions. Using Norwegian Community Innovation Survey data, Aarstad et al. (2016) show that the probability of a firm to engage in innovation activity increases with the degree of related variety in a region in which the firm operates. In case of Swedish firms, Wixe (2018) confirms this result, with an important qualification that related variety plays a more pronounced role in innovative activities of firms located in metropolitan regions.

Following the common conclusion in all of the above-mentioned studies that related variety fosters innovation, we formulate the following hypothesis:

> **Hypothesis 1 (H1):** There is a positive association between related variety in the regional industry structure and regional innovation output.

There are, however, as suggested by Castaldi et al. (2015) and Miguelez and Moreno (2018), important nuances to this relationship, in that some types of innovation may rather be associated with unrelated variety. Radical innovations from the firm perspective typically require distant search, exploration, while other types of improvement result from local search (see March 1991). In other words, explorative innovation may require the recombination of unrelated forms of knowledge, while exploitation, building upon established knowledge, may be associated with related variety. We hence put forward the following hypothesis:

> **Hypothesis 2 (H2):** There is a negative or no association between related variety in the regional industry structure and regional explorative innovation output.

2.2. Skill similarity as a measure of relatedness between industries

A common feature of all previous studies on related variety and regional innovation is that they employ the measure of related variety based on the hierarchical structure of official industry classifications. The fundamental weakness of this approach is that by assuming cognitive similarity to exist only within the same group of the industry classification it underestimates potential channels for knowledge spillovers (Firgo and Mayerhofer 2017). Also, such approach ignores the dynamic nature of relatedness, that is the temporal changes in relatedness linkages between industries over time (Kuusk and Martynovich 2018). Therefore, we employ a set of relatedness measures based on the mobility of workers across industry boundaries.

A defining feature of contemporary labour markets is that skilled workers possess highly specialised human capital as they invest heavily in education and training to acquire skills that allow them to carry out certain job tasks (Neffke et al. 2017). The specificity of human capital affects job switch decisions (Elliott and Lindley 2006) as workers tend to switch to jobs where they incur fewer shortages (skills to be acquired) and smaller redundancies (skills not needed anymore) in their human capital (Neffke and Nedelkoska 2010, Neffke et al. 2017). In other words, when switching jobs workers tend to maximise the opportunity to reuse their skills. This may be achieved if workers search for employment in industries that value skill portfolios similar to those in the previous employment. Therefore, one may expect that an overlap in skill requirements between industries facilitates the inter-industry worker reallocation, while the lack of such overlap constrains it. Thus, labour flows should be larger between industries with bigger overlap in skill requirements – or, skill related industries. Flipping this argument, Neffke and Henning (2013) suggested that the existence of extensive labour flows between two industries may signal that these industries are skill related. That is, relatedness may be inferred from the analysis of labour flows.

Skill relatedness captures directly the scope for knowledge spillover between industries since knowledge is less likely to be beneficially exchanged between economic activities requiring very different skills and know-how (Noteboom 2000). Since, in the knowledge economy, human capital is a key input to the production process and a crucial source of firm competitiveness (Porter 1987, Spender and Grant 1996), the skill relatedness indicator is more directly connected to this crucial element of economic life than the previous indicators. Hence, skill relatedness reveals new aspects of inter-industry linkages complementing the earlier export or production co-occurrence and input-output-based relatedness measures.

A recent investigation by Neffke et al. (2017) demonstrates that skill relatedness based on labour flows: (1) does not reflect the industrial composition of local economies but is a more generic measure; (2) is not simply a reflection of worker reallocation from shrinking to growing industries; and, (3) outperforms alternative relatedness measures in predicting entry and growth of local industries. In other words, skill relatedness is a powerful predictor for understanding economic dynamics.

However, the literature on skill-relatedness also has limitations. In recent work, some authors (Janssen 2015, Boschma 2017) have noted that the literature has so far ignored indirect connections between industries, such as triadic closure (indirectly connected through common neighbours) or through indirect paths (indirectly connected through a chain of related industries). Broadly in line with these suggestions, we note that taking indirect linkages between industries into account comes closer to the notion that related variety is based on the *recombination* of knowledge present in diverse industries (Frenken et al. 2007). Hence, a potentially important step is to investigate how knowledge recombines, rather than just looking at direct skill flows. Our new approach to recombinant knowledge flows is detailed in section 4.2.

3. Innovation output in Sweden, 1970-2013

3.1. SWINNO – Database of Swedish Innovations

Our regional innovation output data is based on a longitudinal micro-database containing detailed information about over 4 000 Swedish innovations commercialised between 1970 and 2013. The underlying data collection instrument is the literature-based innovation output method (LBIO) (Kleinknecht and Bain 1993).

While innovation indicators such as R&D expenditure and patents have been used for a long time and are available for a large set of countries and long time periods, these measures are only indirectly measuring innovation. To deal with some of these issues, LBIO was developed in the 1990s. The major advantage in relation to other innovation measures, is that the LBIO method captures actual innovations, while avoiding self-reporting bias of survey-based output indicators. Since the 1990s, there is a sizeable number of studies based on this approach (Acs and Audretsch 1990, Kleinknecht and Bain 1993, Santarelli and Piergiovanni 1996, Alegre-Vidal et al. 2004, Saarinen 2005, Villar et al. 2012, Sjöö 2014, Sjöö et al. 2014, Taalbi 2014, Kander et al. 2019). Apart from Sweden, there are LBIO databases for the US, Netherlands, Spain, and Finland, although only Finland has a comparable long-run ambition and broad industry coverage (Saarinen 2005, Makkonen and van der Have 2013).

A few studies have compared the performance of patents and LBIO as regional innovation indicators (Acs et al. 2002, Gössling and Rutten 2007, Makkonen and van der Have 2013), all of which have found that patents and innovation counts from LBIO have reasonably high correlation coefficients in cross-sectional data. Meanwhile, concerns remain to what extent patents or other innovation indicators can be used as proxies in longitudinal settings.[1]

SWINNO was constructed by scanning about 8,600 articles in 15 Swedish trade journals, covering the manufacturing industry and ICT services (Sjöö 2014, Sjöö et al. 2014, Taalbi

[1] Findings from the Swedish database, 1970-2013, suggest EPO and national patent applications to have sizeable cross-sectional correlations at the municipality level but performing more modestly in fixed-effects panel regressions Taalbi, J. (2020). Regional innovation and patenting in the long run. A note on Sweden, 1970-2013. Mimeo..

2014, Taalbi 2017, Taalbi 2017). All journals were required to have an editorial mission to report on technological development of the industry. The edited sections of the journals were searched for innovations, defined as a significantly improved good, process or service that is commercialised on a market. As a demarcation, only innovations developed by Swedish companies were included. While commercialised new goods, processes or services are included, the method typically does not capture in-house process innovations.

SWINNO contains several variables based on detailed innovation biographies, which are summarised in Table 1 below.

Table 1. SWINNO variables (excerpt)

Variable name	Variable description	Variable coding
Year of commercialisation	Commercialization year of the innovation. This is either the year as mentioned in the sources, or, if no year is mentioned, the year of the earliest article that reports that the innovation is for sale.	Year
Product code	Product code of the innovation according to SNI2002 (national implementation of NACE Rev. 1.1).	Five-digit industry code
Location	Location of the innovating firm	Municipality code
Degree of novelty	Degree of novelty from the firm perspective	1=entirely new (exploration)[2] 2=major improvement (exploitation)[3] 3=incremental improvement (exploitation)[4]

The location of innovation is defined in this study as the location of an innovating firm. This data stems in part from the innovation biographies available from trade journal articles. Almost 60% of all innovations are directly linked to a specific municipality. The remaining innovations are not attached to a geographical location in the primary material. In these cases, the innovation is matched with location based on information about firms' history, active persons and patents related to innovation. In cases where the firm, often a new business, is established in only one city, this location was used. For large corporate groups, historical sources (annual reports and literature) are used to match the specific innovation or inventor to a subdivision or subsidiary. In the remaining cases, when this has not been possible, an assessment was made of the type of innovation and the connection to a particular development

[2] An innovation is graded as *totally new* if the firm ventured into a new field of technology and the innovation required a significant reconfiguration of the firm's knowledge base, including cases where the firm was started to introduce the innovation. 44.9% of innovations in the dataset are classified as totally new.

[3] An innovation is classified as a *major improvement* if the innovation was developed from the current knowledge base of the firm but was described in the article as a significant advance. 43.65% of innovations are classified as major improvements.

[4] An innovation is classified as an *incremental improvement* to the firm, if the innovation was a new version of a previous innovation. 11.46% of innovations are classified as incremental improvements.

area. Through these multiple steps, 96% of all innovations are linked to a particular municipality, with a fairly stable degree of coverage over time (Table 2).

Table 2. Geocoded innovations per decade and in total

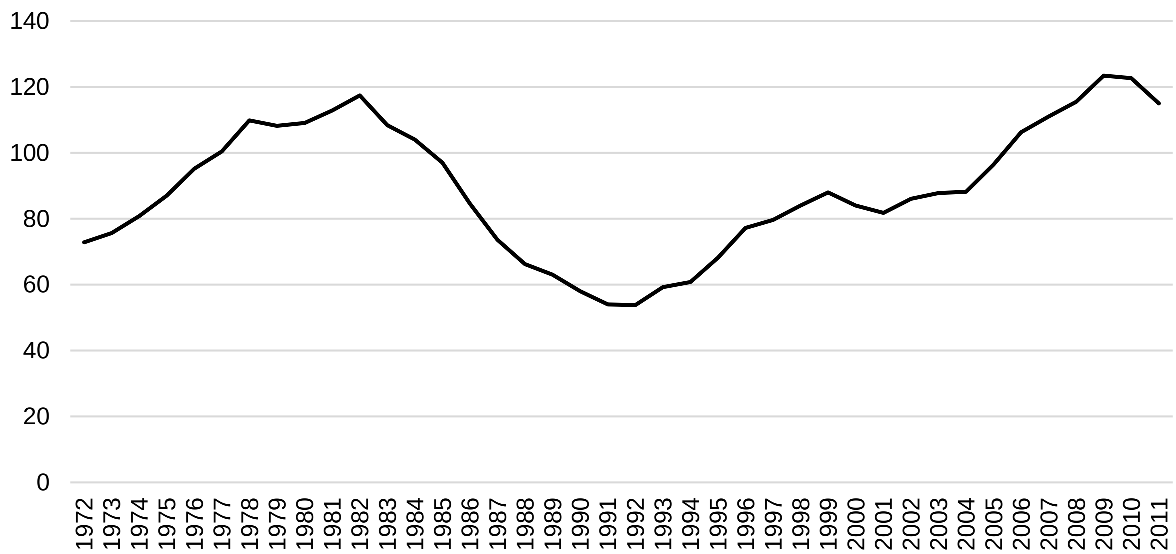
	Geocoded	Number of innovations	Share geocoded
1970-1979	866	931	93%
1980-1989	955	992	96%
1990-1999	667	708	94%
2000-2009	985	1,006	98%
2010-2013	460	466	99%
Total	3,933	4,103	96%

For investigation purposes, 290 Swedish municipalities are merged into 90 local labour markets (as of 2000) to form the spatial units employed in the subsequent analysis. This unit of analysis is appropriate as knowledge flows in Sweden were demonstrated to transcend municipal borders, while being bounded within functional regions (Andersson and Karlsson 2007). This, among other things, implies that a large part of spatial dependence is internalised within these spatial units (Tavassoli and Carbonara 2014).

3.2. Innovation in Sweden: Descriptive statistics

The total count of innovation output of the Swedish economy over the period between 1970 and 2013 is reported in Figure 1.

Figure 1. National innovation count, 1970-2013, 5-year moving average

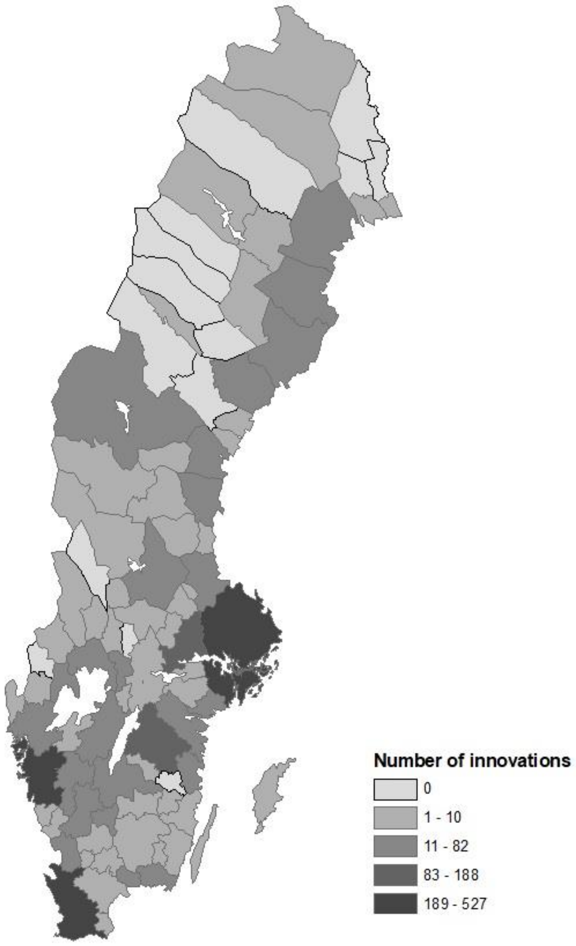


A striking feature of innovation activity in Sweden is a clear cyclical pattern with two periods of increase in innovation output (1970-1982 and 1994-2009) interrupted by a period during which a decline in the number of commercialized innovations is observed. These waves of innovation output can in part be connected phases of expansion related to ICTs (Taalbi 2018): a first one related to industry automation and the second related to the further expansion of technologies based on digital technologies, Internet and telecommunication infrastructure.

The first wave of innovation can also be related to creative responses to the oil crisis of the 1970s (Taalbi 2014, Taalbi 2017).

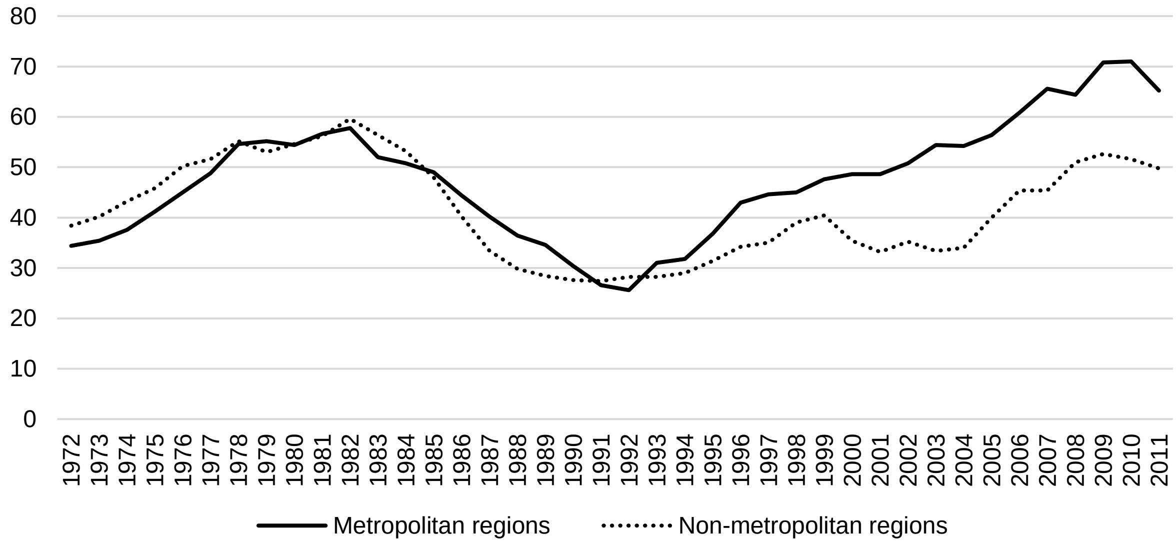
The regional distribution of innovation output in Sweden demonstrates a clear pattern of geographical concentration (Figure 2), which is in line with findings of previous studies investigating innovation in Sweden using other indicators (Tavassoli and Carbonara 2014, Wixe 2018). In other words, innovation performance of Swedish regions demonstrates a clear pattern of overdispersion (the innovation count sample variance is 44 times the sample mean). Four regions – Stockholm, Gothenburg, Malmö, and Linköping – are performing much better than others, of which the former three are the Swedish metropolitan areas and the latter is an industrial region with a long history of innovation activity. At the same time, 14 regions do not have any innovation output throughout the whole investigation period. Such concentrated geographical pattern of regional innovation output implies that spatial dependency is not a significant concern (Tavassoli and Carbonara 2014).

Figure 2. Regional distribution of innovation, total count,1970-2013



An interesting pattern emerges when we consider the spatial distribution of innovation activity over time. Figure 3 reports innovation output over the considered time period divided between metropolitan and non-metropolitan regions. During the expansion in the 1970s and decline in the 1980s, there were no significant differences between the two groups of regions. However, from the early 1990s one can observe a decoupling: metropolitan regions consistently outperform non-metropolitan regions in the innovation output. Moreover, it is not just any innovations, but explorative, radical innovations (from the firm perspective) which became increasingly produced in metropolitan regions, while the distribution of major and incremental improvements across regions remained more equal (see figures A1.1-A1.3 in Appendix 1). This reflects a wave of disruptive technologies, including digital technologies, but also entry of entirely new firms. While many factors may have contributed to that, we posit that the increased complexity of innovation and importance of knowledge recombination could play a role. Metropolitan regions are characterised by more diversified industry structures, which provides a fertile environment for recombinant innovation and ventures into new fields. A broader connection between innovation counts and related variety would support this notion.

Figure 3. Innovation count by region type, total count, 1970-2013, 5-year moving average



4. Relating innovation output to regional industry structure: methodological considerations

We perform an econometric analysis of Swedish regional innovation performance conditional on regional industry structure in 90 local labour markets between 1991 and 2010. The dependent variable is regional innovation count derived from SWINNO.

Our explanatory and control variables are derived from LISA (*Longitudinal Integration Database for Health Insurance and Labour Market Studies*) that is a longitudinal linked employer-employee database covering all individuals registered in Sweden (SCB 2016). It connects each employed individual to the establishment of main employment (denoted by the identity number of establishment) for which the industry affiliation is known. In this respect, the data allows constructing a very precise measure of regional industrial structures based on regional employment. Annual data cover the period between 1991 and 2010.

Classification of economic activities is based on the Swedish Standard Industrial Classification (SNI), which is the Swedish implementation of the Statistical Classification of Economic Activities in the European Community (NACE). During the period of observation two versions of SNI are employed: SNI92 (based on NACE Rev. 1) and SNI2002 (based on

NACE Rev. 1.1). Establishing unambiguous links between these two classification schemes was necessary to ensure the data consistency over time. As SNI2002 was a result of a minor revision of SNI92, this was solved by manually merging the activity classifications at the five-digit level. The resulting classification included 746 five-digit industries, which were further aggregated into 505 four-digit industries.

4.1. Modelling framework

We investigate the determinants of regional innovation output using the regional knowledge production function (Acs et al. 2002), which in general form is specified as:

$$Y_r = \alpha X_r^{\beta}, \tag{1}$$

where Y_r is the innovation output in region r and X_r is the vector of innovation inputs within the region r.

Given the specific features of the dependent variable – a count variable with high overdispersion (see Figure 2) – we apply the negative binomial regression model to estimate the relationship between regional innovation and regional related variety. It is specified in the following way:

$$\Pr(Y_{rt} = \widetilde{y_{rt}} | RV_{rt}, C_{rt}, \sigma_{rt}) = \frac{e^{-\lambda_{rt}} \times (\lambda_{rt})^{\widetilde{y_{rt}}}}{\widetilde{y_{rt}}!} \tag{2}$$

where

$$y_{rt} = 0,1,2,3 \ldots$$

and

$$\lambda_{rt} = \exp(\beta_1 RV_{rt} + \beta_2 C_{rt} + u_t) \times \exp(\sigma_{rt})$$

where Y_{rt} – is innovation counts in region *r* over *4* years between t where t represents one of the periods 1991-1994, 1995-1998, 1999-2002, 2003-2006 and 2007-2010, RV_{rt} – the matrix of related variety variables, C_{rt} – the matrix of control variables, and u_t represents region-invariant unobservable time effects.

In order to investigate the relationship between the quality of regional innovation output and related variety, we specify the following generalised linear models:

$$share\_new_{rt} = \beta_1 RV_{rt} + \beta_2 C_{rt} + u_t + \varepsilon_{rt} \tag{3}$$

$$share\_major_{rt} = \beta_1 RV_{rt} + \beta_2 C_{rt} + u_t + \varepsilon_{rt} \tag{4}$$

$$share\_incerem_{rt} = \beta_1 RV_{rt} + \beta_2 C_{rt} + u_t + \varepsilon_{rt} \tag{5}$$

where $share\_new_{rt}$, $share\_major_{rt}$, $share\_increm_{rt}$ are, respectively, the shares of totally new innovations (explorative innovation), major improvements and incremental improvements (exploitative innovation) in the regional innovation output in region *r* over *4* years.

A 3-year period panel model on innovation counts for sub-periods is preferred over use of the full (year-by-year) panel structure of the data because the key variables on industrial composition change rather slowly over time, implying a very low year-by-year variance within regions (Firgo and Mayerhofer 2017). Furthermore, a year-by-year panel only identifies effects of changes in industrial composition on regional innovation that take place immediately but not longer-run effects. All variables in RV_{rt} and C_{rt} consist of the values for the first year (t) in each sub-period to mitigate endogeneity concerns.

In the estimation of the model presented above, we employ the random-effect (RE) estimator for two features of our data. First, RE estimator is more efficient than fixed-effects (FE) when most of variation in the data consists of the between-variation rather than the within-variation Second, FE estimator may wrongfully include the impact of those variables which exhibit only slight changes over time (Tavassoli and Carbonara 2014).

4.2. Measuring related variety

After structuring the regional employment profiles into 505 industries, we calculate the regional degree of related variety according to three approaches: entropy-based measure as proposed by Frenken et al. (2007), regional skill relatedness measure based on Neffke and Henning (2013) and Fitjar and Timmermans (2017), and finally measures that take into account the overall topology of directed worker flow network (including indirect connections).

4.2.1. Entropy-based measure of related variety

To calculate the entropy-based measure in the spirit of Frenken et al. (2007) the regional industry structure is analysed by making a distinction between two levels of industry aggregation: four-digit industry and two-digit industry. Each four-digit industry i belongs to one of two-digit industry classifications S_g, where $g \in \{1,2,\dots,G\}$. After that, the share of regional employment in each two-digit industry class g in each time period t Q_{grt} is calculated as a sum of its four-digit constituents q_{irt}: $Q_{grt} = \sum_{i \in S_g} q_{irt}$.

The related variety index is then calculated as the weighted entropy index for lower level four-digit industry codes in each of two-digit industry classes, indicating the diversity within the lower levels:

$$RV_{rt}^{Entropy} = \sum_{g=1}^{G} Q_{grt} H_{grt} \tag{6}$$

where

$$H_{grt} = \sum_{i \in S_g} \frac{q_{irt}}{Q_{grt}} \log_2 \left(\frac{1}{q_{irt}/Q_{grt}} \right).$$

The inherent assumption behind this indicator is that four-digit industries are related when they share the two-digit industry class. Such a definition of relatedness disregards potential linkages between industries belonging to different industry groups, thus underestimating related variety at a regional level and, thereby, possibly misrepresenting opportunities for recombinant innovation. Besides, as industry classifications are fixed, linkages between industries are not allowed to emerge/disappear over time. This implies that the only source of variation in the indicator is the structural change at a regional level, while the evolution of industry ties (e.g., as a result of technological change) is not accounted for. This limits the explanatory power of entropy-based related variety measure in long-term setting. Nevertheless, widespread use of this indicator in the similar studies motivates its inclusion to our analysis for comparison purposes.

4.2.2. Regional skill-relatedness

The weaknesses of the entropy-based measure of related variety stem from the highly restrictive underlying definition of which industries are related. To deal with these issues, it is necessary to *reveal* relatedness between industries empirically without imposing any *ex ante* assumptions on structure of industry linkages. Measures based on revealed relatedness should, at least in principle, better capture related variety at a regional level and its evolution over time.

As follows from discussion in Section 2.2, industry relatedness may be inferred from analysis of inter-industry worker flows. More specifically, excessive exchange of labour between two industries signals overlapping skill requirements between them and indicates that these industries are related (Neffke and Henning 2013). Let F_{ijt} be an observed (actual) flow of labour between industries i and j at time t and $\widehat{F_{ijt}}$ – an expected flow of labour between them derived from some general characteristics of both industries[5]. Then the values of ratio of observed to predicted flows

$$SR_{ijt}^{Skill} = \frac{F_{ijt}}{\widehat{F_{ijt}}},$$

that are significantly larger than 1 indicate that industries i and j are skill-related[6]. A detailed description of procedures of calculating expected labour flows and determining the significance of skill-relatedness indicator are provided in Appendix 2.

To account for a substantial inter-temporal variation in relatedness between industries (Kuusk and Martynovich 2018), we construct separate $N \times N$ industry relatedness matrices for five time periods: 1991-1994, 1995-1998, 1999-2002, 2003-2006, 2007-2010. In these matrices, cell values indicate the degree of 'closeness' of each couple of industries.

From constructed linkage metrics, we derive the regional measure of related variety, by weighting the metrics according to Fitjar and Timmermans (2017):

$$RV_{rt}^{Skill} = \frac{(\sum_{i=1}^{N}\left(\frac{\sum_j s_{ijrt}}{2}\right)\sqrt{q_{irt}})/N_{rt}}{(\sum_{i=1}^{N}\sqrt{q_{irt}})/N_{rt}} \tag{7}$$

5 Note that these are calculated at the national level.

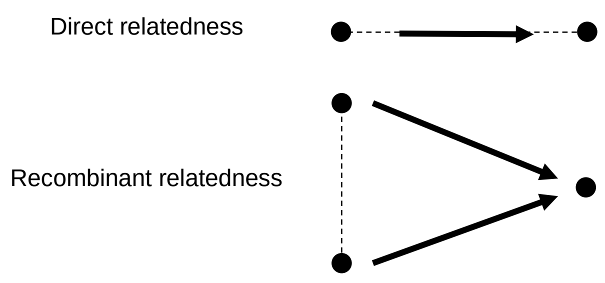
6 We impose no assumption on the symmetry of ties. That is, if industry i is related to industry j this does not necessarily imply that industry j is related to industry i.

where s_{ijrt} – a measure of inter-industry relatedness between industry i and other industries j ($i \neq j$) present in the region r in time period t; q_{irt} – a share of industry i in the regional employment in time period t; N_{rt} – a number of industries present in the region in time period t. Using the information on the presence of related ties between industries within the region, this indicator allows measuring the overall degree of related variety in the regional economy. In broad terms, it represents the (weighted) average number of related industries per each industry.

4.2.3. Topology of skill flow networks and regional related variety

Recently, some authors (Janssen 2015, Boschma 2017) have noted that the literature has so far ignored indirect connections, such as triadic closure (indirectly connected through common neighbours) or through indirect paths (indirectly connected through a chain of related industries) In addition to these suggestions, we note that taking indirect linkages between industries into account comes closer to the notion that related variety is based on the recombination of knowledge present in diverse industries (Frenken et al. 2007). While skill-relatedness is based on the direct flow of skilled labour between industries, we suggest that the notion of recombinant knowledge may rather require attention to indirect linkages between industries. In other words, industries that that have *similar skill-flow patterns* to other industries are likely to have recombinant knowledge and thus be conducive to innovation (see illustration in Figure 4). As a complement to the standard measures of related variety, we thus construct new measures of industry linkages by employing node similarity metrics from network theory.

Figure 4. Direct and recombinant (indirect) relatedness



To estimate these similarity measures, we note that skill-flows form a directed network having industries as nodes. We develop three measures to capture similarity using local information from the skill-flow network. A first measure uses the adjacency matrix with elements a_{ij} having value 1 if industries are connected, otherwise 0. The other two uses the weights of the skill-flows to separate strong and weak ties. The first, called the Dice metric, is based on the number of common neighbours of two nodes i and j. For a directed network the number of common neighbours is given by:

$$|\Gamma(i) \cap \Gamma(j)| = \sum_{k \in \Gamma(i) \cap \Gamma(j)} a_{ik} + a_{jk}$$

with $\Gamma(i)$ and $\Gamma(j)$ being the set of neighbouring nodes of i and j respectively, and k being an index of neighbouring nodes.

The Dice metric proceeds by normalising the common neighbours by taking into account the sum of node degrees defined as

$$s_{ij}^{Dice} = \sum_{k\in\Gamma(i)\cap\Gamma(j)} \frac{2(a_{ik} + a_{jk})}{(\sum_{k\in\Gamma(k)} a_{ik} + \sum_{k\in\Gamma(k)} a_{jk})} = \frac{2|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i)| + |\Gamma(j)|} \tag{8}$$

This metric ranges between 0 and 1, where 0 denotes two industries having no links to the same node and 1 denotes two industries having all links to the same set of nodes.

Since, most skill-flows are relatively small, we may prefer to take into account the weights of the network w_{ij}. The Jaccard similarity metric normalises common neighbours by taking into account the effect of neighbourhood size:

$$s_{ij}^{Jac} = \sum_{k\in\Gamma(i)\cap\Gamma(j)} \frac{w_{ik} + w_{jk}}{\sum_{l\in\Gamma(i)} w_{il} + \sum_{m\in\Gamma(i)} w_{jm}} \tag{9}$$

As before, the metric ranges between 0 and 1, where 0 denotes two industries having no links to the same node and 1 denotes two industries having all links to the same set of nodes.

As rare connections may be more indicative of where new links emerge, the Adamic-Adar metric normalises the metric by letting a common neighbour k be weighted by the rarity of relationships between other nodes and k (Adamic and Adar 2003):

$$s_{ij}^{AA} = \sum_{k\in\Gamma(i)\cap\Gamma(j)} \frac{w_{ik} + w_{jk}}{\log(1 + \sum_{m\in\Gamma(k)} w_{km})} \tag{10}$$

We aggregate these matrices into the regional related variety indicator by substituting s_{ij} in equation (7) by values calculated in (8)-(10).

4.3. Control variables

To measure the *absolute diversity* in regional employment mix we calculate the reverse Hirschman-Herfindahl index defined in the following way:

$$Diversity_{rt} = \frac{1}{\sum_{i=1}^{N} Q_{grt}{}^{2}}$$

where Q_{irt} – employment share of a two-digit industry i in region r in time period t.

Following van Oort et al. (2015), we include the Theil index (sum of location quotients of the SNI 2-digit industries weighted by their employment shares within a region) as a measure of *relative regional specialisation*. It is calculated as:

$$Theil_{rt} = \sum_{i=1}^{N} \frac{Q_{grt}}{Q_{gt}} * \ln\left(\frac{Q_{grt}}{Q_{gt}}\right)$$

where Q_{grt} – employment share of a two-digit industry g in region r in sub-period t; Q_{gt} – employment share of a two-digit industry i in national employment in time period t.

The difference between the two latter measure is that while the absolute diversity measure reflects the concentration of employment within a region, Theil index transforms the individual sectoral concentration measures in a generalised between-region specialisation measure.

Human capital effects on regional employment dynamics is captured by the share of regional workers with higher education (within the group of workers aged 25+):

$$human\_cap_{rt} = \frac{HE\_emp\_25+_{rt}}{emp\_25+_{rt}}$$

An important determinant of innovation at the regional level is also the degree of urbanisation: there is a consensus that the dynamism of large cities makes them motors of economic growth (Fujita et al. 1999, Duranton and Puga 2001). Urban agglomeration is also considered to lead to greater innovation (Iammarino 2005) and to lower barriers and costs of knowledge sharing and transmission across individual and firm networks (Storper and Venables 2004). We therefore capture the urbanisation externalities by the population density in the respective region. Additionally, we add a dummy for metropolitan regions (Stockholm, Gothenburg, and Malmö) to control for the presence of the regional decoupling between metropolitan and non-metropolitan regions in the regional innovation output.

Finally, it is shown that sectors differ in their propensity to innovate. First of all, service sector is less likely to produce innovations its knowledge production compared with manufacturing (Tavassoli and Carbonara 2014). In order to incorporate this argument, the paper includes location quotient of the manufacturing specialisation. It is calculated as follows:

$$LQ\_man_{rt} = \frac{q_{manrt}}{q_{mant}},$$

where q_{manrt} – employment share of manufacturing industries in region r in time period t; q_{mant} – employment share of manufacturing industries in national employment in time period t. Higher value of LQ_man denotes the higher concentration of manufacturing industries in the region as compared to the national average.

Second, even within manufacturing, sectors have shown different behaviour in terms of propensity to patent, because different sectors have different technology and innovation opportunities, and are thus characterised by different technological regimes. As for controlling this second point, the paper includes location quotient of the high-tech manufacturing sectors (LQ_HTman)[7]. To account for the potential supportive role of knowledge-intensive services for innovation in the manufacturing sector, we also calculate the locational quotient of the knowledge-intensive services employment in the regional employment mix (LQ_KIS)[8].

Descriptive statistics and correlations are provided in Table 3.

[7] High-tech manufacturing includes high-technology and medium-high-technology sectors as defined by OECD. This corresponds to the following two-digit sectors in NACE Rev. 1.1. (24, 29-34, 35 excluding 35.1)

[8] Knowledge-intensive services are defined according to the OECD definition. This corresponds to the following two-digit sectors in NACE Rev. 1.1. (61, 62, 64-67, 70-74, 80, 85, 92).

Table 3. Descriptive statistics and correlations

Variable	N	Mean	St. dev.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
1. Innovation count	450	3.849	13.494	1.000																
2. Share_new	252	0.406	0.366	0.081	1.000															
3. Share_major	252	0.464	0.386	-0.075	-0.808	1.000														
4. Share_increm	252	0.106	0.211	0.005	-0.195	-0.349	1.000													
5. RV_entropy	450	1.245	0.219	0.425	0.095	-0.145	0.098	1.000												
6. RV_skill	450	20.594	3.062	0.336	0.024	-0.113	0.167	0.762	1.000											
7. RV_Dice	450	35.331	13.417	0.323	-0.041	-0.021	0.159	0.704	0.652	1.000										
8. RV_Jac	450	53.738	14.994	0.372	0.031	-0.101	0.152	0.822	0.951	0.777	1.000									
9. RV_AA	450	132.318	41.590	0.456	0.056	-0.116	0.137	0.876	0.916	0.757	0.976	1.000								
10.Theil index	450	0.318	0.213	-0.241	-0.093	0.114	-0.048	-0.796	-0.669	-0.538	-0.696	-0.721	1.000							
11. Diversity	450	11.501	1.955	0.401	0.045	-0.072	0.091	0.601	0.522	0.431	0.542	0.602	-0.460	1.000						
12. Pop. density	450	24.170	29.041	0.748	0.062	-0.084	0.042	0.585	0.425	0.396	0.494	0.598	-0.366	0.472	1.000					
13. Human capital	450	0.187	0.053	0.499	0.085	-0.114	0.068	0.681	0.660	0.814	0.753	0.735	-0.606	0.242	0.475	1.000				
14. LQ_man	450	1.230	0.554	-0.245	-0.293	0.285	0.021	-0.452	-0.339	-0.172	-0.313	-0.337	0.628	-0.082	-0.198	-0.478	1.000			
15. LQ_Htman	450	0.955	0.746	-0.020	-0.156	0.202	-0.051	-0.193	-0.108	-0.041	-0.072	-0.065	0.197	-0.019	0.059	-0.169	0.582	1.000		
16. LQ_KIS	450	0.912	0.127	0.304	0.196	-0.196	-0.026	0.441	0.402	0.220	0.389	0.396	-0.631	-0.091	0.228	0.584	-0.852	-0.424	1.000	
17. Metro	450	0.033	0.180	0.838	0.069	-0.069	0.007	0.422	0.252	0.240	0.299	0.409	-0.231	0.374	0.867	0.393	-0.230	-0.043	0.262	1.000

5. Related variety and regional innovation in Sweden

5.1. Estimating the regional knowledge production function

Table 4 reports the results of estimating the regional knowledge production regions in Swedish regions between 1991 and 2010. In line with previous studies, we identify a positive relationship between related variety at the regional level and innovation output of a region for all five measures employed in the analysis: entropy-based related variety, regional skill relatedness, and three measures based on the topology of network of labour flows across industries. There are, however, important qualifications to this result.

First, in the restricted model specifications, where only related variety measures are included as explanatory variables, all five measures have a positive impact on regional innovation at the 1% level. Once we control for other characteristics of regional economies, significance drops to the 10% level for the entropy-based measure, while remaining highly significant in most other specifications. This suggests that the revealed measures of related variety, based on labour flows between industries, are better predictors of regional innovation output. This claim is supported by Akaike and Schwarz information criteria.

Second, within the group of revealed measures of related variety, measures based on node similarity metrics derived from skill-networks (except Dice) should be preferred over the regional skill relatedness measure which is derived from the direct flows of labour between each couple of industries. This follows from information criteria. This suggests that related variety measures accounting for the indirect linkages between industries are more suitable to capture the idea of recombinant innovation.

Third, an interesting result emerges once we consider the models with and without a dummy for metropolitan areas, which is included to control for the decoupling in innovation output between metropolitan and non-metropolitan regions during our period of investigation (see Figure 3). Including the dummy not only improves the significance of related variety measures in all models, but also increases coefficient magnitude. There are two implications of this. First, metropolitan and non-metropolitan regions are characterised by two different scales of knowledge production function. Second, the role of related variety is pronounced (and becomes even more so once we account for these two scales). This supports our conjecture that regional decoupling, observed in Figure 3 and Figures A.1.1-A.1.3., might be explained by the increased complexity of innovation and importance of knowledge recombination.

Control variables in general perform according to our expectations. Regions with higher population density (reflecting urbanisation externalities) and higher share of workers with higher education (reflecting regional human capital) perform better in terms of innovation output. In terms of specialisation of regional industrial portfolios, larger presence of manufacturing employment is associated with higher innovativeness as follows from the positive significant coefficient for manufacturing location quotient. This might stem from the fact that the SWINNO database does not capture process innovation, but may also indicate that manufacturing remains the backbone of innovation for regional economies. Moreover, not surprisingly, the presence of knowledge intensive services, which are identified as key nodes in knowledge networks that foster processes of knowledge creation and diffusion (Miozzo and Grimshaw 2005), has a positive impact on regional innovation output.

Table 4. Regional innovation output and related variety

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
RV_Entropy	5.021***	1.214*	1.256*												
	(0.539)	(0.729)	(0.704)												
RV_Skill				0.413***	0.109**	0.169***									
				(0.044)	(0.049)	(0.051)									
RV_Dice							0.122***	0.019	0.037**						
							(0.010)	(0.017)	(0.017)						
RV_Jac										0.103***	0.037***	0.053***			
										(0.008)	(0.012)	(0.012)			
RV_AA													0.034***	0.015***	0.017***
													(0.002)	(0.004)	(0.004)
Theil index		-0.113	-0.159		-0.139	-0.137		-0.160	-0.112		-0.009	0.054		0.073	0.080
		(0.212)	(0.210)		(0.199)	(0.192)		(0.212)	(0.208)		(0.207)	(0.200)		(0.210)	(0.205)
Diversity		0.194***	0.168***		0.182***	0.132**		0.210***	0.173***		0.173***	0.121**		0.165***	0.131***
		(0.052)	(0.052)		(0.052)	(0.052)		(0.051)	(0.050)		(0.052)	(0.051)		(0.052)	(0.051)
Population density		0.345***	0.236**		0.368***	0.187*		0.394***	0.230**		0.328***	0.124		0.289***	0.146
		(0.117)	(0.115)		(0.110)	(0.109)		(0.111)	(0.109)		(0.112)	(0.109)		(0.111)	(0.108)
Human capital		12.623***	11.677***		12.319***	10.726***		10.908***	8.068***		10.644***	8.400***		9.945***	8.569***
		(2.519)	(2.453)		(2.525)	(2.426)		(3.075)	(2.988)		(2.640)	(2.534)		(2.644)	(2.533)
LQ_man		1.366***	1.427***		1.335***	1.403***		1.298***	1.326***		1.253***	1.287***		1.206***	1.249***
		(0.352)	(0.332)		(0.351)	(0.320)		(0.353)	(0.325)		(0.348)	(0.313)		(0.339)	(0.313)
LQ_HTman		0.082	0.097		0.031	0.054		0.049	0.075		0.038	0.064		0.059	0.079
		(0.137)	(0.129)		(0.136)	(0.124)		(0.136)	(0.125)		(0.135)	(0.121)		(0.131)	(0.120)
LQ_KIS		4.173***	3.908***		3.933***	3.428***		4.526***	4.494***		4.052***	3.609***		4.211***	3.923***
		(1.317)	(1.282)		(1.316)	(1.261)		(1.334)	(1.276)		(1.307)	(1.242)		(1.285)	(1.231)
Metro			1.047***			1.392***			1.249***			1.482***			1.170***
			(0.371)			(0.369)			(0.363)			(0.354)			(0.336)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N regions	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
LR test vs pooled	146.257	435.573	454.443	140.694	434.068	477.617	225.715	432.022	482.032	240.500	447.946	513.417	286.450	468.105	523.137
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N obs	450	450	450	450	450	450	450	450	450	450	450	450	450	450	450
AIC	1352.113	1270.877	1264.481	1349.286	1268.584	1256.205	1318.777	1272.420	1262.803	1307.724	1264.309	1249.213	1297.623	1261.333	1251.330
BIC	1384.987	1332.516	1330.229	1382.160	1330.223	1321.953	1351.651	1334.059	1328.551	1340.598	1325.948	1314.961	1330.496	1322.971	1317.078

Note: Dependent variable in all models is regional innovation count. The table reports coefficient parameters with standard errors in parentheses. ***(**,*) indicate a significant difference from 1 at the 1% (5%, 10%) level. LR test vs pooled reports the outcome of the Chi-squared test comparing panel specification of the model vs. pooled specification. AIC – Akaike information criterion, BIC – Schwarz information criterion

5.2. Robustness check

To perform a robustness check, we would like to take advantage of the fact that two network topology measures – namely, Dice and Jaccard metrics – range between zero and one. Therefore, by a slight transformation of equation (7) we can derive a measure of dissimilarity of labour flows between different industries:

$$URV_{rt} = \frac{(\sum_{i=1}^{N} \left(\frac{\sum_j (1 - s_{ijrt})}{2}\right) \sqrt{q_{irt}})/N_{rt}}{(\sum_{i=1}^{N} \sqrt{q_{irt}})/N_{rt}} \quad (11)$$[9]

It is easy to show that

$$\frac{N_{rt}}{2} = URV_{rt} + RV_{rt}$$

which means that for Dice and Jaccard metrics we can decompose overall variety (N_r) into two components, related variety, and what we call unrelated variety. This allows us to compare the impact of related variety to that of unrelated variety to compare our results with the previous literature (Castaldi et al. 2015, Miguelez and Moreno 2018). For comparison reasons, we also include a traditional measure of unrelated variety as the entropy of the distribution of regional employment at the two-digit industry level:

$$URV_{rt}^{Entropy} = \sum_{i \in S_g} Q_{grt} \log_2 \left(\frac{1}{Q_{grt}}\right). \quad (11)$$[10]

After deriving these measures, we repeat the analysis, including the measures of unrelated variety alongside the measures of related variety (Table 5). The result of estimation confirm our conclusions.

Indeed, in restricted models all three measures of related variety have a positive and strongly significant impact on regional innovation output. The entropy-based measure of related variety turns insignificant once we control for other regional characteristics, while measures based on node similarity metrics remain significant. As in case of models reported in Table 4, related variety based on Jaccard metrics demonstrates the most consistent performance as a predictor of regional innovation output. When it comes to the impact of unrelated variety, the results are mixed: it is positive and significant in all restricted specifications, but turns insignificant in models with control variables (except entropy-based unrelated variety).

All in all, our results remain robust to the inclusion of unrelated variety measures.

[9] Notations are similar to those in equation (7).

[10] Notations are similar to those in equation (6).

Table 5. Regional innovation output and related variety (robustness check, excerpt)

	(1)	(2)	(3)	(4)	(5)	(6)
RV_Entropy	3.981*** (0.644)	0.953 (0.713)				
URV_Entropy	2.608*** (0.888)	3.659** (1.610)				
RV_Dice			0.070*** (0.022)	0.048* (0.025)		
URV_Dice			0.018** (0.007)	-0.004 (0.007)		
RV_Jac					0.041*** (0.014)	0.039*** (0.015)
URV_Jac					0.028*** (0.005)	0.010 (0.006)
Theil index		0.097 (0.233)		-0.113 (0.208)		0.147 (0.207)
Diversity		-0.013 (0.095)		0.170*** (0.050)		0.111** (0.052)
Population density		0.219* (0.113)		0.234** (0.110)		0.084 (0.111)
Human capital		12.948*** (2.491)		8.035*** (2.994)		8.577*** (2.522)
LQ_man		1.364*** (0.327)		1.322*** (0.326)		1.252*** (0.311)
LQ_HTman		0.061 (0.129)		0.076 (0.125)		0.066 (0.120)
LQ_KIS		3.672*** (1.271)		4.463*** (1.278)		3.656*** (1.239)
Metro		1.242*** (0.359)		1.348*** (0.406)		1.215*** (0.387)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
N regions	90	90	90	90	90	90
LR test vs pooled	153.396 (0.0000)	484.467 (0.0000)	257.158 (0.0000)	478.803 (0.0000)	331.476 (0.0000)	530.744 (0.0000)
N obs	450	450	450	450	450	450
AIC	1345.465	1261.327	1314.402	1264.486	1283.347	1248.840
BIC	1382.448	1331.184	1351.385	1334.344	1320.330	1318.697

Note: Dependent variable in all models is regional innovation count. The table reports coefficient parameters with standard errors in parentheses. ***(**,*) indicate a significant difference from 1 at the 1% (5%, 10%) level. LR test vs pooled reports the outcome of the Chi-squared test comparing panel specification of the model vs. pooled specification. AIC – Akaike information criterion, BIC – Schwarz information criterion

5.3. Related variety and quality of regional innovation

To investigate the hypothesis on the role of related variety for the quality of regional innovation we estimate generalised linear models – specified by equations (3)-(5) – where the dependent variable is the share of innovations that have a given degree of novelty, representing radical recombination of knowledge bases (entirely new) or the exploitation of existing knowledge bases (major improvement and incremental improvements). Table 6 reports the results.

Table 6. Quality of regional innovation and related variety

	Totally new innovations			Major improvements			Incremental improvements		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RV_Entropy	0.128			-0.293			0.113		
	(0.300)			(0.274)			(0.169)		
URV_Entropy	-0.999			0.768			0.448		
	(0.675)			(0.607)			(0.376)		
RV_Dice		-0.034***			0.028**			0.011	
		(0.013)			(0.013)			(0.008)	
URV_Dice		0.011**			-0.010**			-0.000	
		(0.004)			(0.004)			(0.003)	
RV_Jac			-0.022***			0.016**			0.007
			(0.008)			(0.008)			(0.005)
URV_Jac			0.006*			-0.007**			0.000
			(0.003)			(0.003)			(0.002)
Theil index	-0.024	0.049	0.029	-0.078	-0.128*	-0.123	0.063	0.062	0.064
	(0.098)	(0.084)	(0.085)	(0.086)	(0.077)	(0.077)	(0.054)	(0.047)	(0.047)
Diversity	0.056	0.014	0.020	-0.042	-0.010	-0.012	-0.021	-0.005	-0.007
	(0.044)	(0.023)	(0.023)	(0.040)	(0.022)	(0.022)	(0.024)	(0.013)	(0.013)
Population density	-0.056	-0.050	-0.035	0.002	0.001	-0.005	0.034	0.024	0.020
	(0.043)	(0.041)	(0.042)	(0.037)	(0.037)	(0.038)	(0.023)	(0.023)	(0.023)
Human capital	0.851	2.374*	2.537*	-0.940	-1.854	-1.917	-0.527	-1.472*	-1.350*
	(1.235)	(1.402)	(1.308)	(1.115)	(1.270)	(1.215)	(0.691)	(0.776)	(0.739)
LQ_man	-0.328***	-0.313***	-0.301**	0.329***	0.316***	0.319***	-0.015	-0.033	-0.042
	(0.123)	(0.119)	(0.120)	(0.112)	(0.112)	(0.113)	(0.069)	(0.068)	(0.068)
LQ_HTman	0.028	0.023	0.025	0.004	0.015	0.013	-0.024	-0.023	-0.023
	(0.051)	(0.048)	(0.048)	(0.045)	(0.043)	(0.043)	(0.028)	(0.026)	(0.027)
LQ_KIS	-0.625	-0.600	-0.372	0.557	0.450	0.350	0.079	0.175	0.018
	(0.547)	(0.539)	(0.541)	(0.514)	(0.516)	(0.525)	(0.313)	(0.310)	(0.315)
Metro	0.003	-0.255	-0.245	0.070	0.289*	0.273*	-0.016	0.020	0.021
	(0.156)	(0.182)	(0.181)	(0.127)	(0.163)	(0.162)	(0.082)	(0.100)	(0.099)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	252	252	252	252	252	252	252	252	252
R2_within	0.082	0.083	0.086	0.095	0.099	0.099	0.028	0.034	0.026
R2_between	0.398	0.434	0.430	0.345	0.371	0.360	0.137	0.120	0.152
R2_total	0.167	0.190	0.190	0.165	0.177	0.174	0.059	0.072	0.075

Note: Dependent variable in models (1)-(3) is the share of totally new innovations in the regional innovation output. Dependent variable in models (4)-(6) is the share of major improvements in the regional innovation output. Dependent variable in models (1)-(3) is the share of incremental improvements in the regional innovation output. The table reports coefficient parameters with standard errors in parentheses. ***(**,*) indicate a significant difference from 1 at the 1% (5%, 10%) level.

With regards to radical (explorative) innovations, our results confirm the results reported in previous studies (Castaldi et al. 2015, Miguelez and Moreno 2018): related variety is negatively associated with the most radical innovations, while unrelated variety has a positive relationship. This is true only for the related variety measures based on node similarity metrics while coefficients for entropy-based measures are insignificant. The observed outcome implies that when it comes to explorative innovations, which supposedly require recombination of highly unconnected ideas and/or technologies, higher degree of related variety in regional industry structure is not only less important but can even be damaging.

We, however, go beyond existing studies and also demonstrate the impact of related variety for less radical innovation. When it comes to major improvements, we observe the opposite picture to that of totally new innovation. That is, the share of major improvements is positively associated with related variety and negatively associated with unrelated variety. This suggests that less radical innovation benefits from the presence of more related regional industry structures, which makes knowledge exchange between industries easier.

Another interesting result comes from considering the role of manufacturing sector for innovations with different novelty degree. Extended presence of manufacturing has a negative impact on the share of totally new innovations, which points to the increased importance of service sector for generating innovations – either as a direct innovation generator or, more likely, as a support structure for radical innovation process. However, insignificant coefficient for location quotient for knowledge intensive services does not allow us to fully confirm this hypothesis. When it comes to major improvements, however, larger presence of manufacturing is a positive factor for regional innovation.

We also estimate a regression model for incremental improvements, which does not report any significant results. This might have to do with the fact that incremental improvements represent the minor share of innovations in SWINNO database. Besides, innovation indicator used already at the outset represents a selection of *significant* innovations, and hence, by construction, only captures incremental innovations inasmuch as they are incremental to the firm's knowledge base, but have some societal significance, for example, by implying a larger leap in performance.[11]

6. Concluding remarks

The present study has tested the related variety hypothesis by using an indicator of actual innovation output and by using entropy-based measure of related variety, regional skill-relatedness and network topology-based related variety measures capturing the recombination of knowledge. We find broad support for the notion that related variety matters for innovation output performance across Swedish regions. This is true regardless of what indicator of related variety is used.

More specifically, we demonstrate that measures of related variety based on the revealed relatedness between industries perform better in regressions explaining regional innovation output. Moreover, measures taking into consideration not only direct linkages between industries but also indirect similarities in the worker flows across industries are demonstrated to be superior predictors of regional innovation. In that respect, our study makes an empirical contribution to the literature by providing the set of related variety measures that outperform those conventionally employed in the literature.

The contribution, however, goes beyond empirical. Indeed, we demonstrate that estimating regional knowledge production functions based on the notion of recombinant innovation require the measures that capture the recombination of knowledge explicitly. In that respect, recombinant relatedness based on the node similarity metrics in the worker flow networks come closer to measuring recombinant relatedness, as they allow for the broader scope of potential linkages between industries as well as incorporate possibility for technological change by treating relatedness as a dynamic concept. In that respect, we demonstrate that previous efforts of investigating the role of related variety for regional innovation were based on a restricted notion of relatedness.

[11] For example, a new product may simply be an update of an earlier product that the firm has produced, hence incremental to the firm, but still imply a large leap in performance.

Looking at the patterns of exploration versus exploitation, related variety has a negative association with exploration (radical innovation) and positive relationship with exploitation, viz. innovations that build on previous knowledge within firms (compare to Castaldi et al. 2015, Miguelez and Moreno 2018). This has obvious corollaries for our understanding of regional transformation. Regions with unrelated variety in their industrial structure are better positioned for radical industrial renewal and absorbing disruptive technology shifts. Conversely, regions with strong related variety, synergistic but traditional industries, would tend to exploit established trajectories and knowledge bases. In this light, the decoupling of innovation activity observed between metropolitan and non-metropolitan regions (Figure 3 as well as Figures A1.1-A1.3 in Appendix 1) can be understood as driven by the greater capacity for unrelated variety and diversification of metropolitan areas. This is in line with hypotheses about technology shifts as driven by metropolitan areas (Henning et al. 2016, Lundquist et al. 2017), as well our understanding of regional diversification (Neffke et al. 2011, Boschma 2017).

References

Aarstad, J., O. A. Kvitastein and S.-E. Jakobsen (2016). "Related and unrelated variety as regional drivers of enterprise productivity and innovation: A multilevel study." Research Policy **45**(4): 844-856.

Acs, Z. J., L. Anselin and A. Varga (2002). "Patents and innovation counts as measures of regional production of new knowledge." Research Policy **31**(7): 1069-1085.

Acs, Z. J. and D. B. Audretsch (1990). Innovation and small firms. Cambridge, MA, The MIT Press.

Adamic, L. A. and E. Adar (2003). "Friends and neighbors on the Web." Social Networks **25**(3): 211-230.

Alegre-Vidal, J. n., R. Lapiedra-Alcamı́ and R. Chiva-Gómez (2004). "Linking operations strategy and product innovation: an empirical study of Spanish ceramic tile producers." Research Policy **33**(5): 829-839.

Almeida, P. and B. Kogut (1999). "Localization of Knowledge and the Mobility of Engineers in Regional Networks." Management Science **45**(7): 905-917.

Andersson, M. and C. Karlsson (2007). "Knowledge in Regional Economic Growth—The Role of Knowledge Accessibility." Industry and Innovation **14**(2): 129-149.

Antonietti, R. and G. Cainelli (2011). "The role of spatial agglomeration in a structural model of innovation, productivity and export: a firm-level analysis." The Annals of Regional Science **46**(3): 577-600.

Arthur, W. B. (2007). "The structure of invention." Research Policy **36**(2): 274-287.

Beaudry, C. and A. Schiffauerova (2009). "Who's right, Marshall or Jacobs? The localization versus urbanization debate." Research Policy **38**(2): 318-337.

Boschma, R. (2005). "Proximity and Innovation: A Critical Assessment." Regional Studies **39**(1): 61-74.

Boschma, R. (2017). "Relatedness as driver of regional diversification: a research agenda." Regional Studies **51**(3): 351-364.

Boschma, R., R. Eriksson and U. Lindgren (2009). "How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity." Journal of Economic Geography **9**(2): 169-190.

Breschi, S. and F. Lissoni (2009). "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." Journal of Economic Geography **9**(4): 439-468.

Castaldi, C., K. Frenken and B. Los (2015). "Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting." Regional Studies **49**(5): 767-781.

Content, J. and K. Frenken (2016). "Related variety and economic development: a literature review." European Planning Studies **24**(12): 2097-2112.

de Groot, H. L. F., J. Poot and M. J. Smit (2016). "Which Agglomeration Externalities Matter Most and Why?" Journal of Economic Surveys **30**(4): 756-782.

Duranton, G. and D. Puga (2001). "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products." The American Economic Review **91**(5): 1454-1477.

Duranton, G. and D. Puga (2004). Chapter 48 - Micro-Foundations of Urban Agglomeration Economies. Handbook of Regional and Urban Economics. J. V. Henderson and J.-F. Thisse, Elsevier. **4:** 2063-2117.

Elliott, R. J. and J. Lindley (2006). "Skill Specificity and Labour Mobility: Occupational and Sectoral Dimensions." The Manchester School **74**(3): 389-413.

Firgo, M. and P. Mayerhofer (2017). "(Un)related variety and employment growth at the sub-regional level." Papers in Regional Science.

Fitjar, R. D. and B. Timmermans (2017). "Regional skill relatedness: towards a new measure of regional related diversification." European Planning Studies **25**(3): 516-538.

Frenken, K., F. Van Oort and T. Verburg (2007). "Related Variety, Unrelated Variety and Regional Economic Growth." Regional Studies **41**(5): 685-697.

Fujita, M., P. Krugman and A. Venables (1999). The Spatial Economy: Cities, Regions and International Trade. Cambridge, MA, The MIT Press.

Gössling, T. and R. Rutten (2007). "Innovation in Regions." European Planning Studies **15**(2): 253-270.

Hassink, R., C. Klaerding and P. Marques (2014). "Advancing Evolutionary Economic Geography by Engaged Pluralism." Regional Studies **48**(7): 1295-1307.

Henning, M., K.-J. Lundquist and L.-O. Olander (2016). Regional analysis and the process of economic development: changes in growth, employment and income. Structural Analysis and the Process of Economic Development. J. Ljungberg. Oxon and New York, Routledge**:** 149-173.

Iammarino, S. (2005). "An evolutionary integrated view of Regional Systems of Innovation: Concepts, measures and historical perspectives." European Planning Studies **13**(4): 497-519.

Jacobs, J. (1969). The Economy of Cities. New York, Vintage Books.

Janssen, M. J. (2015). Service innovation in an evolutionary perspective. PhD Thesis, Eindhoven Technical University.

Kander, A., J. Taalbi, J. Oksanen, K. Sjöö and N. Rilla (2019). "Innovation trends and industrial renewal in Finland and Sweden 1970–2013." Scandinavian Economic History Review **67**(1): 47-70.

Kleinknecht, A. and D. Bain, Eds. (1993). New Concepts in Innovation Output Measurement. London, Macmillan.

Kuusk, K. and M. Martynovich (2018). What kind of related variety for long-term regional growth? Papers in Evolutionary Economic Geography (PEEG) 1834. Utrecht, Utrecht University.

Kuusk, K. and M. Martynovich (2018). What kind of related variety for long-term regional growth? Papers in Evolutionary Economic Geography #18.34. Utrecht, the Netherlands, Utrecht University.

Lundquist, K.-J., L.-O. Olander and M. Martynovich (2017). The technology shift thesis: Understanding long term growth and transformation in a regional system. Research Reports in Human Geography 2017:1. Lund, Department of Human Geography, Lund University.

Makkonen, T. and R. P. van der Have (2013). "Benchmarking regional innovative performance: composite measures and direct innovation counts." Scientometrics **94**(1): 247-262.

March, J. G. (1991). "Exploration and Exploitation in Organizational Learning." Organization Science **2**(1): 71-87.

Martynovich, M. (2016). General purpose technology diffusion and labour market dynamics: A spatio-temporal perspective. Lund, Media-Tryck.

Miguelez, E. and R. Moreno (2018). "Relatedness, external linkages and regional innovation in Europe." Regional Studies **52**(5): 688-701.

Miozzo, M. and D. Grimshaw (2005). Does EDS Add Value? The Expansion of IT Outsourcing and the Nature and Role of Computer Services Firms. DRUID Tenth Anniversary Summer Conference 2005. Copenhagen, Denmark.

Neffke, F. and M. Henning (2013). "Skill relatedness and firm diversification." Strategic Management Journal **34**(3): 297-316.

Neffke, F., M. Henning and R. Boschma (2011). "How do regions diversify over time? Industry relatedness and the development of new growth paths in regions." Economic Geography **87**(3): 237-265.

Neffke, F. and L. Nedelkoska (2010). Human capital mismatches along the career path. Jena economic research papers No. 2010,051. Jena, Leibniz Information Centre for Economics.

Neffke, F. M. H., A. Otto and A. Weyh (2017). "Inter-industry labor flows." Journal of Economic Behavior & Organization **142**: 275-292.

Noteboom, B. (2000). Learning and innovation in organizations and economies. Oxford, Oxford University Press.

Noteboom, B., W. Van Haverbeke, G. Duysters, V. Gilsing and A. van den Oord (2007). "Optimal cognitive distance and absorptive capacity." Research Policy **36**(7): 1016-1034.

Porter, M. E. (1987). "From Competitive Advantage to Corporate Strategy." Harvard Business Review **65**(3): 43-59.

Rodan, S. and C. Galunic (2004). "More than Network Structure: How Knowledge Heterogeneity Influences Managerial Performance and Innovativeness." Strategic Management Journal **25**(6): 541-562.

Saarinen, J. (2005). Innovations and industrial performance in Finland, 1945–1998, Lund University.

Santarelli, E. and R. Piergiovanni (1996). "Analyzing literature-based innovation output indicators: the Italian experience." Research Policy **25**(5): 689-711.

SCB (2016). Longitudinell integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990-2013. Arbetsmarknad och Utbildning Bakgrundsfakta 2016:1. Örebro, SCB.

Schumpeter, J. A. (1934). The Theory of Economic Development. Cambridge, MA, Harvard University Press.

Sjöö, K. (2014). Innovation and transformation in the Swedish manufacturing sector, 1970–2007, Lund University.

Sjöö, K., J. Taalbi, A. Kander and J. Ljungberg (2014). SWINNO: A Database of Swedish Innovations, 1970-2007. Lund Papers in Economic History. General Issues; No. 133. Lund, Department of Economic History, Lund University.

Spender, J.-C. and R. M. Grant (1996). "Knowledge and the firm: Overview." Strategic Management Journal **17**(S2): 5-9.

Storper, M. and A. J. Venables (2004). "Buzz: face-to-face contact and the urban economy." Journal of Economic Geography **4**(4): 351-370.

Taalbi, J. (2014). Innovation as creative response: Determinants of innovation in the Swedish manufacturing industry, 1970–2007, Lund Univeristy.

Taalbi, J. (2017). "Development blocks in innovation networks." Journal of Evolutionary Economics **27**(3): 461-501.

Taalbi, J. (2017). "What drives innovation? Evidence from economic history." Research Policy **46**(8): 1437-1453.

Taalbi, J. (2018). "Origins and pathways of innovation in the third industrial revolution1." Industrial and Corporate Change **28**(5): 1125-1148.

Taalbi, J. (2020). Regional innovation and patenting in the long run. A note on Sweden, 1970-2013. Mimeo.

Tavassoli, S. and N. Carbonara (2014). "The role of knowledge variety and intensity for regional innovation." Small Business Economics **43**(2): 493-509.

van Oort, F., S. de Geus and T. Dogaru (2015). "Related Variety and Regional Economic Growth in a Cross-Section of European Urban Regions." European Planning Studies **23**(6): 1110-1127.

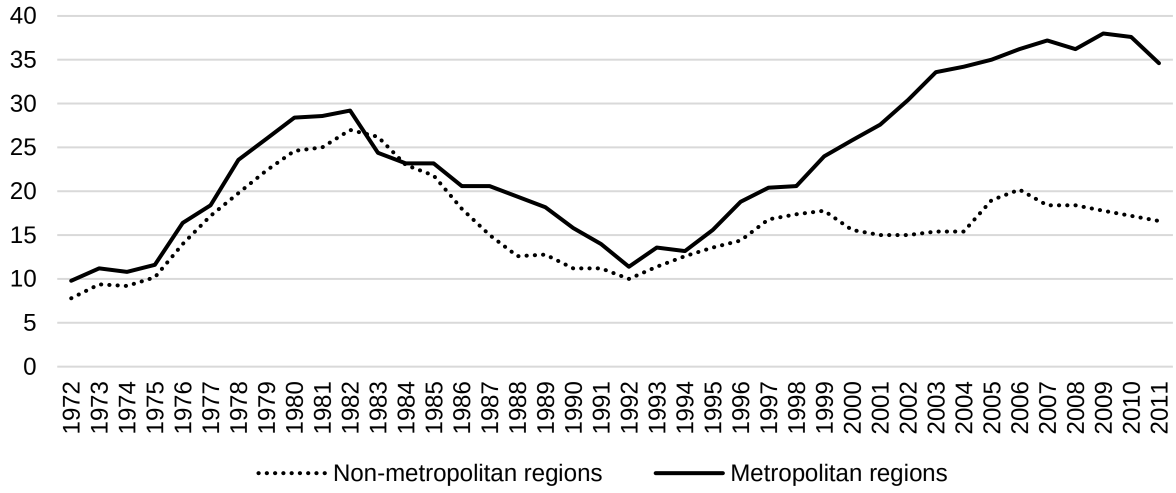
Weitzman, M. L. (1998). "Recombinant Growth." The Quarterly Journal of Economics **113**(2): 331-360.

Villar, C., J. Pla-Barber and J. Alegre (2012). "Unravelling the moderating effects of size and experience on product innovations and exports: a study in a medium knowledge-intensive industry." Technology Analysis & Strategic Management **24**(2): 219-238.

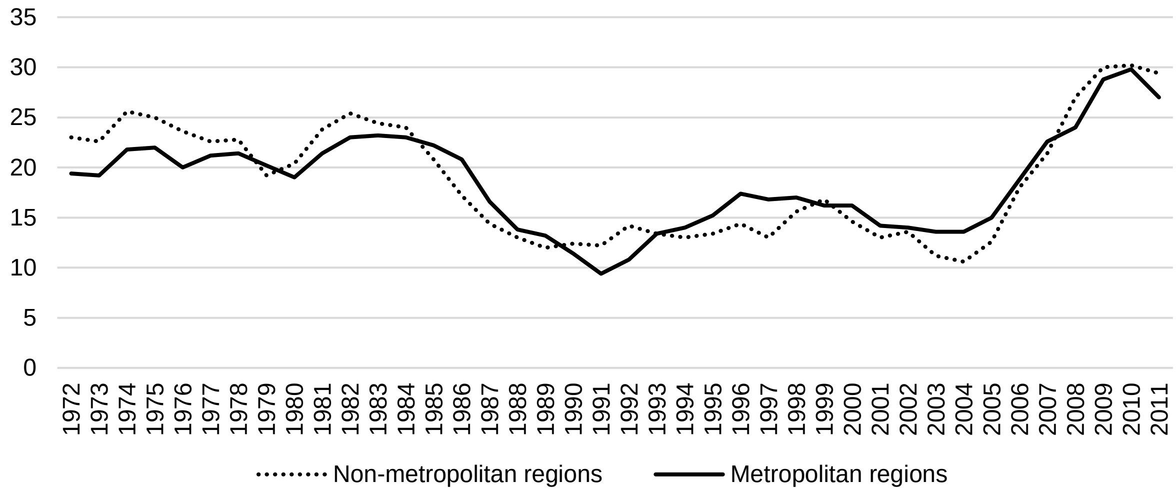
Wixe, S. (2018). "Neighbourhood related diversity, human capital and firm innovation." Papers in Regional Science **97**(2): 217-252.

Appendix 1. Innovation counts by innovation novelty and region type

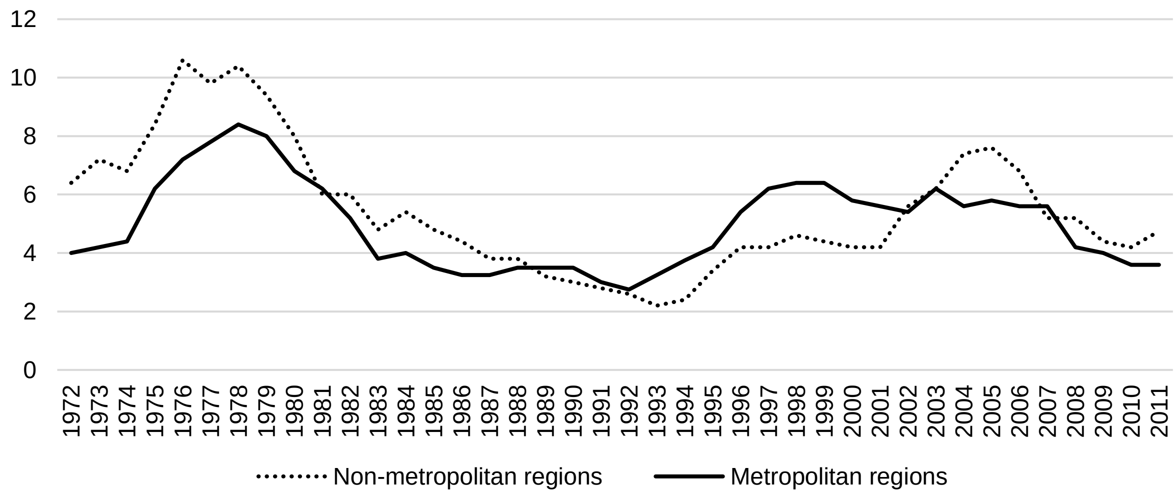
A1.1. Totally new innovations, 1970-2013, 5-year moving average



A1.2. Major improvements, 1970-2013, 5-year moving average



A1.3. Incremental improvements, 1970-2013, 5-year moving average



Appendix 2. Estimating skill-relatedness

This section describes the procedure of estimating the presence of related ties between four-digit industries. For illustration purposes, we exemplify the procedure for the first considered sub-period (1991-1994).

Data and definitions

The original data contains information on all individuals registered in Sweden for each year between 1991 and 1994. We define an individual as a worker if she (1) is in a working age (16-64) according to the pre-2007 Statistics Sweden definition, (2) has a non-zero income from employment, and (3) is affiliated with an establishment with a registered industry code. Establishments are assigned to four-digit industries according to the classification scheme explained in Section 4 of this paper. Industries that employ fewer than 250 persons on average per year are excluded from the analysis as they are too small to generate or absorb significant labour flows.

Inter-industry labour flows consist of the sum total of individual labour market moves across industries. We register a change in an industry of employment if an employee moves to another establishment at another firm in another industry from one year to the next. By requiring that an employee changes a firm and establishment of employment, we avoid a possible situation when an establishment is reassigned to a different industry.

As discussed in Section 2 of the paper, we can estimate related ties between industries more accurately by limiting our analysis to individuals who are likely to possess industry-specific skills. We therefore disregard all flows involving individuals who earn wages lower than the median wage in the respective industry. The main idea here is that firms provide higher wages to employees who possess skills that confer competitive advantage to the firm. Individuals with few skills deemed critical in the industry will earn wages that are low relative to that industry's overall wage level. This does not necessarily imply that individuals with low wages do not have any industry-specific skills. Rather, we do so to reduce the noise in our relatedness estimates.

Estimating related ties between industries

It is logical to expect that labour flows between industries depend not only on whether industries are related or not, but also on certain general characteristics of the industries involved. In other words, some industries may exhibit substantial in- and outflows of labour regardless of their relatedness to other industries. Therefore, it is necessary to develop a measure of expected labour flows which would incorporate those additional factors into the analysis. We choose three variables: size of industries, employment growth in industries, and average wages in industries involved in estimation.

Given that labour flows constitute an overdispersed count variable with majority of observations being zero (there are no labour flows between most industries), it is appropriate to use a zero-inflated negative binomial (ZINB) model. The ZINB regression equation has two components: a regime selection equation and a count data component. The regime selection equation determines whether there will be any flow at all. Next, the count data component estimates the size of the flows, assuming that a nonzero regime is selected.

We pool all data by summing labour flows and employment data across 1991-1994 to raise the efficiency of the estimates. Following (Neffke and Henning 2013), we estimate a model that uses variables in levels for the regime selection equation and log-transformed variables for the count data equation:

$$E(F_{ij}|v_i, w_j, \varepsilon_{ij}) = [1 - \pi_0(\gamma + \delta_i emp_{i,1991-1993} + \delta_j emp_{j,1992-1994})] \cdot$$

$$f(\alpha + \beta_{1i} \log(emp_{i,1991-1993}) + \beta_{2i} log(wage_{i,1991-1993}) + \beta_{3i} growth_i +$$

$$+\beta_{1j} \log\left(emp_{j,1992-1994}\right) + \beta_{2j} log\left(wage_{j,1992-1994}\right) + \beta_{3j} growth_j\right)$$

with i the industry of origin of a flow and j the industry of its destination, π_0 – the probability that a flow can, in principle, take place, $emp_{k,t}$ – the sum of employment in industry k across years t, $wage_{k,t}$ – the average wage in industry, and $growth_k$ – employment growth in industry k across the observed years.

Using the point estimates of the parameters in the equation above, we calculate the expected labour flows ($\widehat{F_{ij}}$) for all pairwise industry combinations. Comparing those to the observed labour flows (F_{ij}) for the same industry combinations, we obtain the measure of relatedness between industries:

$$SR_{ij} = \frac{F_{ij}}{\widehat{F_{ij}}}.$$

Here, values over 1 indicate the presence of an observed related tie between industries

Determining the significance levels of skill-relatedness estimates

As noted above, there are no labour flows between the vast majority of industries. Importantly, in many such cases, predicted labour flows are negligible as well. What is more, whenever $\hat{F}_{ij}$ is only a fraction of one, an increase in the labour flow from zero to one individual will lead to large changes in the skill-relatedness index. Thus, skill relatedness is not estimated with equal precision for all industry combinations. To quantify the precision of our estimates, we construct confidence intervals.

To do so, we assume that all employees in an industry have the option of switching to a new job in a new industry. If *N* denotes the number of industries present in the national economy, each individual faces *N* independent choices: one is staying in the current industry, and the other *N-1* choices represent moves into each of the remaining industries. The choice to switch jobs can now be modelled as a Bernoulli experiment with a probability of success equal to p_{ij} and the resulting aggregate labour flow from i to j, F_{ij}, is the outcome of a binomial experiment $BIN(n, p)$ where n is equal to the employment in industry i and p is equal to p_{ij}:

$$F_{ij} \sim BIN\left(emp_i, p_{ij}\right).$$

The question of how informative a specific labour flow is now translated into the question of how likely it is to observe F_{ij}^{obs} merely by chance. Let $\hat{p}_{ij}$ be the expected counterpart of p_{ij}:

$$\hat{p}_{ij} = \frac{\hat{F}_{ij}}{emp_i}.$$

If we take $\hat{p}_{ij}$ as a benchmark, the question above corresponds to a statistical test of whether F_{ij}^{obs} is exceptional, assuming that $\hat{p}_{ij}$ represents the real probability that an individual will move from industry i to industry j. If $SR_{ij} > 1$ then the p-value of the corresponding one-sided test can be calculated as follows:

$$P\left(x \geq F_{ij}^{obs} \middle| p_{ij} = \hat{p}_{ij}\right) = 1 - \sum_{r=0}^{F_{ij}^{obs}-1} \left[\hat{p}_{ij}{}^{r} \cdot \left(1 - \hat{p}_{ij}\right)^{emp_i - r} \binom{emp_i}{r}\right].$$

Based on a p-value of five percent, between 1991-1994 SR_{ij} is significantly larger than 1 in 6167 industry combinations. Given that 500 industries were present in the Swedish economy during those years and 27549 pairwise industry combinations contained non-zero observed labour flows, related ties correspond to 2.5 per cent of possible ties and 22.4 per cent of observed ties.