When regional development is path-independent

The emergence of new economic activities

Holm, Jacob Rubæk; Østergaard, Christian Richter

Publication date:
2016

Document Version
Early version, also known as pre-print

When regional development is path-independent: The emergence of new economic activities

Jacob Rubæk Holm (corresponding author),

Department of Business and Management, Aalborg University

Fibigerstræde 11, 9220 Aalborg, Denmark

jrh@business.aau.dk, +4599408247

Christian Richter Østergaard

Department of Business and Management, Aalborg University

Fibigerstræde 11, 9220 Aalborg, Denmark

cro@business.aau.dk

Paper prepared for the track on "regional diversification and growth" at the 5th GCW Conference Valencia, Spain, 22-24 June 2016

Abstract

In this paper we explore the limits of the currently dominating model of structural change, which heavily emphasizes path-dependence. Using a novel methodology we demonstrate how the role of path-dependence varies from dominating to irrelevant across regions in Denmark. We specifically focus on the role of human capital intensity and entrepreneurship in mediating the path-dependence and show that the mediating role may take both positive and negative directions. We conclude that, while path-dependence is important in regional structural change, the truly interesting cases only emerge when additional factors are taken into account and the cases of path-independent structural change are uncovered.

1. Introduction

The process of structural change at the regional level has received a revival of interest in recent years with a focus on the path-dependence in the emergence of new activities in economic systems (Hidalgo et al., 2007; Neffke et al., 2011; Essletzbichler, 2015). These studies find that the process of structural change consists of diversification into activities and industries, which are related to the current activities and industries in the region. The main argument is that new industries draw on and benefit from the existing portfolio of resources in the region. As a result the evolution of regional economies becomes a rather path-dependent process, where new industries arise from related activities in the region, but also bring in new variety to the regional economy. In addition, Neffke et al. (2011) shows that not only are industries related to the existing structure more likely to enter in a region, but unrelated industries are more likely to exit. Relatedness can be defined in a number of ways, such as co-occurrence of products in countries' exports (Hidalgo et al., 2007), co-occurrence of products produced in same plants (Neffke et al., 2011), input-output table based flows of goods and services (Essletzbichler, 2015), labour flows between industries (Neffke et al., 2013), and through firms' formal collaboration patterns (Isaksen, 2015). Despite the differences in the measurement of relatedness these studies generally provides a compelling explanation and description of the path-dependent aspects of structural change.

However, the recent studies do not focus greatly on breaks in this path of development. The recent studies are inherently limited to considering the process of catching up for less developed regions since they must focus on the diffusion of already existing activities from frontier regions to the laggards (Hidalgo et al., 2007; Neffke et al., 2011; Hausmann and Hidalgo, 2011; Essletzbichler, 2015). Another line of research rooted in evolutionary economics has developed a conceptual toolbox for studying structural change to handle the emergence of genuinely new activities (Andersen, 2004; Saviotti and Frenken, 2008) but has not made much empirical progress. In order to identify the emergence of new activities it is necessary to find the conditions under which structural change appears to move into unrelated activities. This is not because new activities are independent of previous activities. The models by Andersen (2004) and Saviotti and Frenken (2008) imply that genuinely new activities arise from dividing up old activities. This gradual process has been in focus when studying structural change since Marshall (1920) and Young (1928) with later developments by, inter alia, Yang and Ng (1993). Empirical studies of structural change that include entirely new economic activities are very difficult to handle (Andersen, 2003) not least since genuinely new activities by definition do not fit into existing industry nomenclatures. Since new activities do not have their own industry codes, relatedness is not defined between the new and old activities, and therefore the emergence of new industries will appear as a break from the previous path when studying the path-dependence of structural change. In economic geography this process is known as path creation (Martin and Sunley, 2006). Path creation happens when regions manage to break away from path dependent development processes, such as lock-in, path extension or path renewal. However, a region's possibilities for path creation depend on whether the region is a core region with many resources or a peripheral region (Isaksen, 2015).

The purpose of this paper is to analyse factors that mediate the effect of relatedness in the heavily path-dependent evolution of regional economies, in order to understand the factors that make the region evolve in new directions and thus creating new paths.

Some studies have attempted to analyse factors that mediate the effect of relatedness in path-dependent evolution from a more qualitative perspective (e.g., Neffke et al. 2011; Holm et al. forthcoming B). However, there is still a lack of studies that develop and apply a more general, quantitative approach.

This paper builds on the quantitative method applied in Essletzbichler (2015) that starts from the method of Neffke et al. (2011) and adds a partitioning analysis. This entails decomposing the evolution of relatedness into five components with primary interest in the entry, exit and selection effects. However, Essletzbichler (2015) draws on a traditional decomposition method used in studies of productivity growth. This paper applies a multivariate decomposition approach developed by Holm et al. (forthcoming A) that allows for the study of specific factors mediating the role in structural change of selection based on relatedness. A secondary contribution of the paper is thus to the development of decomposition methods by constituting the first empirical application of this new method. The paper draws on data from the Danish database on integrated labour market research that holds population census data on all firms and their workforce in Denmark from 1980 onwards. The results focus on the mediating effects of human capital intensity and entrepreneurship. It is shown that while structural change in some regions follows the path created by the existing portfolio of activities, the existing activities in other regions play a minor role and are irrelevant compared to the impact of entrepreneurship and human capital intensity.

The paper is structured as follows. Section 2 presents an overview of previous studies of the path-dependence in the evolution of regional industry structure and develops our framework. Section 3 presents the decomposition method and in section 4 the method is applied. We discuss the results in section 5.

2. Regional path-independent economic evolution

There is a growing literature in evolutionary economic geography that argues that regional economic evolution to a large extent depends on the existing set of industrial activities in the region (Boschma and Frenken, 2011; Neffke et al., 2011; Essletzbichler, 2015). In this literature, economic activities (i.e. industries) are ordered in a landscape according to their mutual relatedness. Regions inhabit a portion of this landscape and as regions evolve they move in the landscape; i.e. expanding into related activities. It is important to note the distinction that the industry landscape does not determine an equilibrium industry structure that regions move toward but instead provides an order that conditions evolution (Metcalfe, 2001). Moving in some parts of the landscape may require specific regional characteristics, and the landscape itself changes over time as technology evolves.

Boschma and Frenken (2011, p.64) claims that “to the extent that new industries emerge from existing industries, the sectoral composition of a regional economy at one moment in time provides and constrains (though does not determine) diversification opportunities of regions in the future”. Boschma and Frenken (2011) argues that new industries often emerge in a region as a process of regional branching, where the new industry grows out of an existing industry or as a recombination of existing competences coming from different industries in the region. Furthermore, Neffke et al. (2011) shows that industries unrelated to the existing set of activities in the regions are more likely to exit and less likely to enter than industries that are related to existing activities in the region. Thus, the regional portfolio of related activities affects the possibilities for new activities. This does not mean that no new variety is added to the regional economy. The regional branching process also includes the establishment of new spinoff firms that based on existing related activities move into new activities (Boschma and Frenken, 2011).

However, the evolutionary process of regional economic change also includes the destruction of industries and the emergence of new industries that cannot be explained by the regional portfolio of related activities. The emergence of new economic activities in a region could be a chance event, but there is contested evidence and a debate in the literature whether the "chance event" is a pure chance event or if it is affected by regional conditions (see e.g., Martin and Sunley, 2006). There are several potential factors in a region that can lead to the emergence of new industries in regions. Martin and Sunley (2006) argues that the sources of new industrial activities in a region could be indigenous creation, where a new industry emerges from within the region without no immediate antecedents in the region; or as transplantation of new activities from other regions; or as diversification of existing industries into related activities. Isaksen (2015) analyse industrial development in thin regions compared to core regions. He argues that the rise of totally new industries in a regions stems from factors such as, the establishment of new firms or spinoff firms, commercialisation of new knowledge stemming from universities or R&D institutions, or as transplantation by the inflow of investments and knowledge from the outside. For core regions all three sources are likely to be present, but for thin regions the two first sources are less likely. Furthermore, core regions have large pools of highly educated labour that increases the absorptive capacity of the region (Cohen and Levinthal, 1990).

Therefore, the determinants of growth of existing industries in a region depends on the portfolio of related activities in the region, while the emergence of new activities in a region depends on the level of entrepreneurial activity in a region and the presence of a university or the share of highly educated people in the regional workforce.

2.1. Degrees of path-dependence

Isaksen (2015) develops a useful taxonomy of path-dependent evolution that we will adapt to the current paper. The method applied in this paper allows us to distinguish between path-dependent evolution in the sense of increasing average relatedness in a region, and whether this is actually caused by a positive relationship between relatedness and which activities emerge.

The reference case is *path extension*. This is where new activities, whether in existing firms or new firms, are within the same industry code and hence the existing activities in the region continue to expand. This means that industries privileged with the presence of relatively many related industries will expand, while other industries contract. In regions experiencing path extension relatedness will be associated with growth and average relatedness will be increasing. In time, path extension is generally considered to run into path exhaustion or develop into path renewal. *Path renewal* means that relatedness is still positive for growth but average relatedness in a region will not increase since new activities emerging from Marshall externalities, Jacobs externalities and knowledge creation within firms are also emerging. *Path exhaustion* is the case where current activities are no longer viable and relatedness has no relationship to industry growth or the relationship may even be negative. The industries that do grow in relative terms are the industries experiencing the least decline in absolute terms. The final possible case is the most interesting but probably the rarest: the case of path creation. *Path creation* is when novelty created through entrepreneurship, through firms' innovation efforts and in the knowledge infrastructure, including universities, leads the regional industry to branch out into unrelated directions or even into entirely new activities that have no industry code yet. Path creation entails that industries experiencing high relatedness

do not necessarily expand and average regional relatedness may move in any direction. However there will be a strong effect of knowledge creation and entrepreneurship on the evolution of industry structure.

In the analysis presented in the current paper it has been possible to include data on the human capital intensity of industries and on the propensity for new plant creation within industries, which allows for the classification of regions experiencing path creation. However it has not been possible to include data on Marshall or Jacobs externalities making it more difficult to identify path renewal. Thus path renewal will be distinguished from path creation by exhibiting a significant role for relatedness in growth in addition to the indirect effects of the included variables.

3. Methods and data

In order to study the change in regional industry structures we use census data for all Danish workplaces and employees. Regions are defined by municipalities except that the different municipalities in Copenhagen are aggregated into the administrative region of "Hovedstaden" excluding the island municipality of Bornholm. This means that there are 71 regions. We use Statistics Denmark's "111 standard classification for publication purposes" for industries. This means that we have only 111 industries covering the entire economy but it gives us additional flexibility since the nomenclature is available from 1985 to 2007. The main aim of the paper is to study the degree to which structural change is path-dependent and in particular what mediates or even cancels out the path-dependence. Path-dependence means that structural change depends on the relatedness of industries currently in the region and this is elaborated below. Based on Isaksen (2015) we have chosen three variables that are expected to mediate the relationship between relatedness and structural change. 1) human capital intensity, 2) entrepreneurship 3) knowledge infrastructure. The human capital intensity of an industry is measured by the share of workers with tertiary education (SHE). Entrepreneurship in an industry is measured by the share of workers at establishments that are less than 5 years old (ENT). Knowledge infrastructure is a regional dummy taking the value 1 for the four main university cities in Denmark: Aalborg, Aarhus, Odense and Hovedstaden. It thus measures the share of workers in the university cities (UNI).

3.1. Relatedness and path-dependence

Essletzbichler (2015) divides studies of relatedness into three groups: those that measure relatedness by distance in the tree-structure of industry nomenclatures; those that measure relatedness through co-occurrence and those that measure relatedness through reliance on similar inputs. He argues that all three approaches have drawbacks: 1) relatedness based on industry nomenclatures (Frenken et al., 2007) is inherently as subjective and arbitrary as the original ordering of industries in the nomenclature. It is easy to show that the ordering of industries in industry nomenclatures has little in common with the ordering that results from more advanced approaches to measuring relatedness (Andersen, 2003). 2) Relatedness based on co-occurrence is widely used (Hidalgo et al., 2007; Hausmann and Hidalgo, 2011; Neffke et al., 2011) but suffers from the drawback that it only captures revealed relatedness. It does not contain information on in what sense the activities are related. 3) Measuring relatedness from input use is also popular (Andersen, 2003; Neffke et al., 2013; Essletzbichler, 2015), but suffers from the drawback that activities to varying degree rely on special and generic inputs and hence the choice of input will determine which activities are found to be related.

We use the skill relatedness measure of Neffke et al. (2013) and hence measure relatedness based on labour flows between industries. If flows exceed expected flows from industry j to industry i then i is relatively related to j. The fact that labour flows more often between the industries than should be expected arguably reflects that the skills acquired in industry j are applicable in industry i and hence i is skill related to j but in principle the opposite need not be true. The skill relatedness index has values from 1 to -1 where 0 means that expected flows and observed flows are equal and positive values mean that observed flows are higher than expected flows. The index used here is computed in exactly the same way as the index used in Holm et al. (forthcoming B). This entails that we compute separate skill relatedness indices for each year. The index is not symmetric in the sense that the skill relatedness of i to j generally differs from the skill relatedness of j to i. In order to emphasise that we do not impose the restrictions of constant or symmetric relatedness we refer to the variable as the skill inflow relatedness of i to j at t, SIR_{ijt}.

If the conditions for performing an economic activity are more favourable when skill related activities are already being performed in the vicinity then new activities in a region are likely to be related to the existing activities. This means that the industries that are most likely to emerge in a region are the industries that are skill related to the existing industries. However organisations (establishments, often firms), not industries, perform activities. The activities of a new firm necessarily differ to some degree from existing firms and when this difference is sufficiently large the new firm will get a different industry classification. This entails that the deterministic model of regional industrial evolution not only predicts that emerging industries will be related to existing industries but also that existing industries will expand or contract depending on the extent to which related industries are present in the region.

To study the extent of path dependence in the growth of industry *i* we need to define a measure of how related industry *i* in region *r* is to the industry structure of *r*. Neffke et al. (2011) and Essletzbichler (2015) defined this by "standardised closeness":

$$SC_{irt} = \sum_{j \in RPF_{rt}} I(SIR_{ijt} > p90_t)/N_{rt} \tag{1}$$

The standardised closeness of industry *i* in region *r* in year *t* is the share of industries in region *r* that are related to *i*. More formally RPF_{rt} is the "regional portfolio" – the set of industries that are present in *r* and N_{rt} is the number of elements in RPF_{rt}. $I(\cdot)$ is an indicator function that takes the value 1 if the expression is true and 0 otherwise. $p90_t$ is the threshold distinguishing related industries from other industries. This is defined as the 90th percentile of the cumulative distribution of SIR_{ijt}. Cases where $SIR_{ijt} = 1$ (i.e. cases where $i = j$) were excluded when determining the 90th percentile.

The path dependent model of regional industrial evolution predicts that the correlation between SC_{irt} and the growth of industry *i* in region *r*, w_{irt}, will be positive. Growth is here defined as:

$$w_{irt}(k) = \frac{x_{irt+k}}{x_{irt}} \tag{2}$$

x_{irt} is the full time equivalent employment of industry *i* in region *r* in year *t*. *k* is an integer for the number of years over which to measure growth. The number we are looking for is the correlation in a given year between the standardised closeness of industries in a region and the growth over the ensuing *k* years of the industries. This correlation has been computed for the 71 regions in the data for $k = 1, \ldots, 7$ and

$t = 1985, \ldots, 2000$. The average correlation across regions for each combination of t ("start") and k ("length") is shown in a smoothed plot in the below Figure 1, which is repeated in Figure 2 from a different. There is a distinct wave-like pattern to the correlations: they are generally high in the mid-1980s, lower around 1990, high again in the mid-1990s and then decline towards 2000. When looking at individual regions (not shown) the pattern is more or less pronounced, or even inverted, but the waves persist. This may be caused by business cycle effects or by technological change, however investigating this observation in more detail goes beyond the scope of the current paper.

Figure 1: Path-dependence - the correlation between standardised closeness and growth

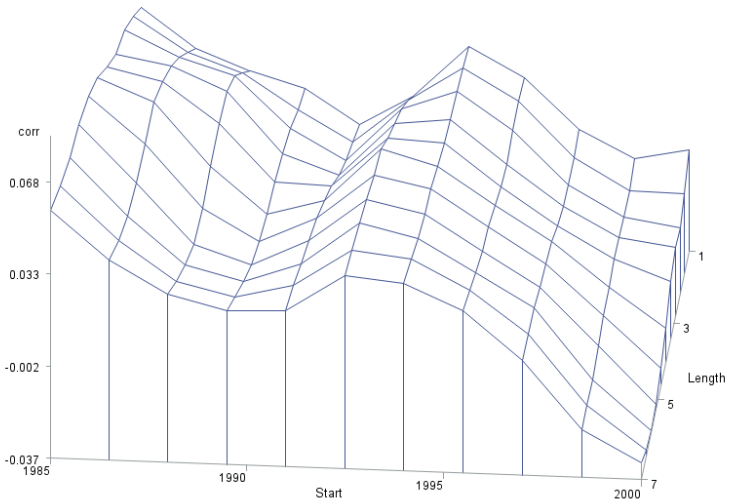
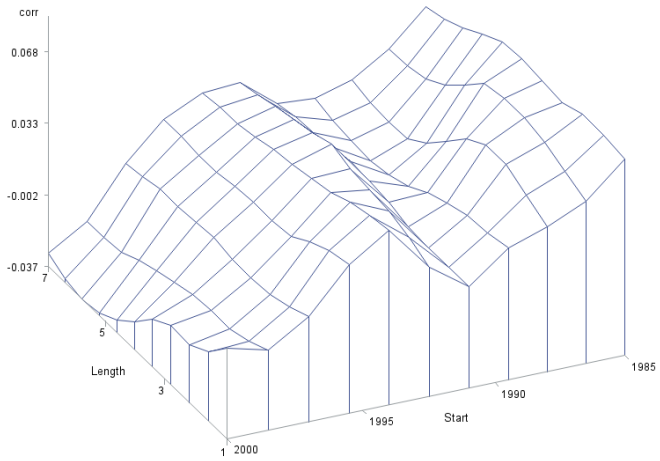


Figure 2: Path-dependence - the correlation between standardised closeness and growth



When looking at the variation in the correlation over k ("length") it appears that the correlation is often stronger for lower values of k. This means that a high value for standardised closeness is associated with growth in the short run but not in the longer run. This is consistent with a pattern of structural change

where path extension turns into path exhaustion. It is curious that this pattern is not observed in the mid-1990s where the correlations are generally high but this observation will not be studied further.

Our aim is to explain the constituents of the observed correlation; to explain why the correlation between growth and standardised closeness is higher at some points in time than in others. Since a low correlation can indicate both path exhaustion and path creation while a high correlation may indicate both path extension and path renewal this aim will help us determine the limits of the path dependent model of regional industrial evolution and uncover the conditions that lead to path creation. This aim does not entail a regressions analysis where firm growth is regressed on the SC_{ir} and a number of controls, since SC_{ir} may well be strongly associated with firm growth without correlating at the population level, because other variables correlate with SC_{ir} and confound the correlation. Instead we need to decompose the population level processes that create the correlation in a mathematical-tautological sense.

Figures 1 and 2 help us determine the interesting point in time to study. We will focus on a single period since this provides ample results for identifying factors that mediate the path dependence of structural change, and we will choose a relatively short period of 3 years in order to choose a period where the correlation is relatively high. Finally, concerns about the quality of some of the older data in the end leads us to choose a three year period starting in 1998.

3.2. Partitioning method

The evolution of industry structure is to be analysed using a partitioning technique that allows for quantifying different contributions to the average change. A similar method has been applied earlier in evolutionary economic geography by Essletzbichler (2015) but this methodology was copied directly from a tradition in productivity studies and only allows to assess the relative importance of correlation between SC_{ir} and growth for the change in average SC_{ir}, i.e. the importance of "selection". We apply a novel and extended methodology (Holm et al. *forthcoming A*) that allows us to include additional variables that explain what causes selection to be stronger or weaker. In addition, we add some further generalisation to the partitioning technique so that we can take into account that beneath the aggregate pattern there are many different regional patterns.

The tools for decomposing a change into its components are necessarily mathematical tautologies and hence the term "explains" is to be understood as such, and not as causality. An interesting partitioning technique is a method that decomposes a change into elements that are readily interpretable and adds insight into the population dynamics underlying the aggregate change. In the following the subscripts t and $t+k$ are omitted since in the ensuing analysis they will take on fixed values of $t = 1998$ and $t+k = 2001$. Values in the end year will be distinguished by a prime, i.e. x_r' is the size in employment of region r in 2001.

A common starting point for the method used by Essletzbichler (2015) and in this paper is equation 3. It is here presented as a national level equation where the population members are the regions of the national economy.

$$\Delta Z = Cov(\omega_r, z_r) + E(\omega_r \Delta z_r) \tag{3}$$

ΔZ is the change from 1998 to 2001 in average population characteristics. In Essletzbichler (2015) this is simply ΔSC while here it is a vector of change in the population average of four variables: standardised closeness, human capital intensity, entrepreneurship and university cities. ω_r is the relative growth of region r ($\omega_r = s'_r/s_r$, and s denotes a population share, $s_r = x_r/\sum_r x_r$). z_r is the characteristics of region r. In the terms of Essletzbichler (2015) $Cov(\omega_r, z_r)$ is the "selection" effect while $E(\omega_r \Delta z_r)$ is the sum of the "portfolio" and "covariance" effects. The two effects will here be referred to as the regional selection and regional adaptation effects.

The regional selection effect will reflect the tendency of regions where industries on average have a high SC to expand more than other regions. I.e. if regions with an industry structure, where the different industries complement each other, grow more than other regions (if economic activity becomes more concentrated in such regions over time) then the regional selection effect will be positive.

The above statement is true in a simple model where SC is the only variable in the Z vector. But other regional characteristics may confound the relationship between average SC and regional growth. Thus, we add the additional variables to Z and the selection term will be generalised as:

$$Cov(\omega_r, z_r) = P\beta \tag{4}$$

Where P is the variance-covariance matrix of the variables in Z. The beta is a vector of selection gradients that are mathematically identical to slope estimates from a WLS regression of regional growth on the variables in Z. Thus the selection effect on the left of equation 4 shows the role of differential regional growth in changing the average national characteristics while the beta-vector of selection gradients show whether the characteristics are actually associated with the growth of regions.

The regional adaptation effect is not divided up in the same manner as in Essletzbichler (2015). Equation 5 shows that the regional adaptation effect can be written as a weighted sum of regional changes, which may themselves be decomposed by an equation corresponding to equation 3.

$$E(\omega_r \Delta z_r) = \sum_r s'_r \Delta z_r \tag{5}$$

By re-specifying equation 3 at the regional level and substituting it into equation 5 we get

$$E(\omega_r \Delta z_r) = \sum_r s'_r \left(Cov(\omega_{ir}, z_{ir})_r + E(\omega_{ir}\Delta z_{ir})_r\right) = \sum_r s'_r \left(P_r\beta_r + \sum_i s'_{ir}\, \Delta z_{ir}\right) \tag{6}$$

However we need to take entry and exit into account at this level. The aggregate population is a population of regions and the number of regions is fixed. But each region is a population of regional industries and industries enter and exit the regional industry portfolio.

$$E(\omega_r \Delta z_r) = \sum_r s'_r \left(P_r\beta_r + \sum_{i\in C_r} s'_{ir}\, \Delta z_{ir} - \sum_{i\in X_r} s_{ir}(z_{ir} - Z_r) + \sum_{i\in E_r} s'_{ir}\,(z'_{ir} - Z_r)\right) \tag{7}$$

The second term in the brackets is the exit effect while the third term is the entry effect. The sets C_r, X_r and E_r are defined as follows: $C_r = RPF_r \cap RPF'_r, X_r = RPF_r \backslash RPF'_r, E_r = RPF'_r \backslash RPF_r$.

Thus, the change in national average characteristics can be decomposed as the sum of five effects:

$$\Delta Z = PB + \sum_r s'_r\, P_r\beta_r + (-1)\sum_r s'_r \sum_{i\in X_r} s_{ir}(z_{ir} - Z_r)$$

$$+\sum_r s'_r \sum_{i\in E_r} s'_{ir}\,(z'_{ir} - Z_r) + \sum_r s'_r \sum_{i\in C_r} s'_{ir}\,\Delta z_{ir} \tag{8}$$

1. The regional selection effect – the effect of the differential growth of regions
2. The industry selection effect – the effect of differential growth of industries within regions
3. The exit effect – the effect of industries disappearing in regions
4. The entry effect – the effect of industries emerging in regions
5. The industry adaptation effect – changes within continuing industries. For SC this is the effect of changes in the relatedness matrix over time and changes in the SC of continuing industries caused by entry and exit of other industries.

4. Analysis

In this section we will analyse the structural change of the Danish economy at the regional level using the decomposition technique presented above. In such decompositions it is typical to find that entry and exit have relatively low effects since the entering and exiting industries are small and thus have low weights, cf. equation 8. This is because the effects of entry and exit are more appropriately studied on a longer time scale. In order to show that entry and exit do have important roles in structural change we present a separate graphical description of entry and exit before analysing the components of change.

4.1. Entry and exit

Figure 3 shows the structural change in Danish regions from 1985 to 2007. The green line shows the share of industries that was a part of the regional portfolio in 1985 as a share of the total number of industries in each year, while the brown line shows the share of industries in 2007 that were a part of the regional portfolio in each year. About 10 percent of the industries that were present in 1985 have disappeared by 2007 (green), and roughly 10 percent of the industries that are present in 2007 were not there in 1985 (brown). The numbers are averages across 71 Danish municipalities. The green line, the share of industries in a year that were also present in 1985, has a positive slope in 1995-1997 meaning that a number of industries re-enter, however industries continuously exit and re-enter, and if we only look at industries that are *continuously* present from 1985 onwards then the relevant line is the blue line (and the red line for industries present in 2007). The red and blue lines indicate that slightly less than 20 percent of industries in 1985 are not continuously present until 2007, while more than 20 percent of industries in 2007 have not been continuously present since 1985.

Figure 3: Structural change in Danish regions 1985-2007

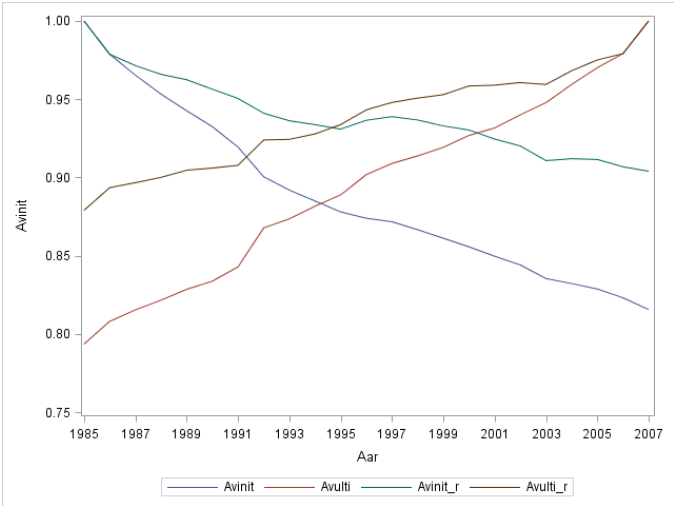


Figure 4 shows the structural change as in Figure 3 but with only the private sector. If we remove the non-private sector (Public, health, education) then the curves shift down on the vertical axis by a few percentage points (Figure 4). If we consider only manufacturing we get a much more dramatic picture (notice the change in the vertical axis), see Figure 5.

Compared to and Neffke et al. (2011) our graphs seem modest as the authors find a greater change in regional industry portfolio, even when taking into account the difference in the number of years being studied. However Neffke et al. (2011) only focus on a subset of manufacturing industries ("174 six-digit industries in the Swedish manufacturing sector" (p. 245)). On the other hand, Essletzbichler (2015) also only studies manufacturing ("… a consistent set of 362 [US] manufacturing industries", p. 756) and finds results that are relatively similar to ours, despite the relatively aggregate industries used.

Figure 4: Structural change in the private sector for Danish regions 1985-2007

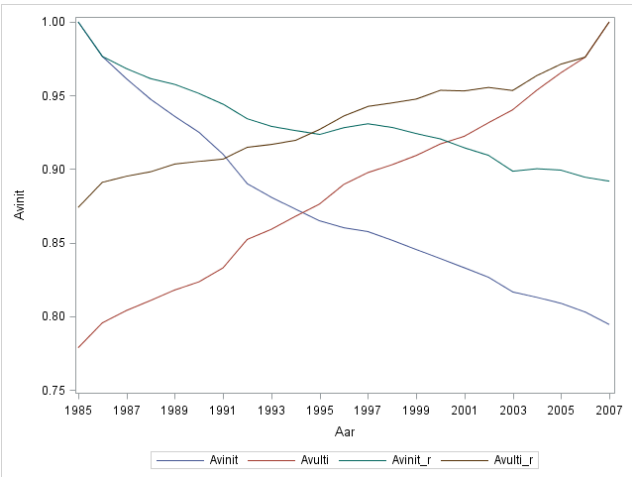
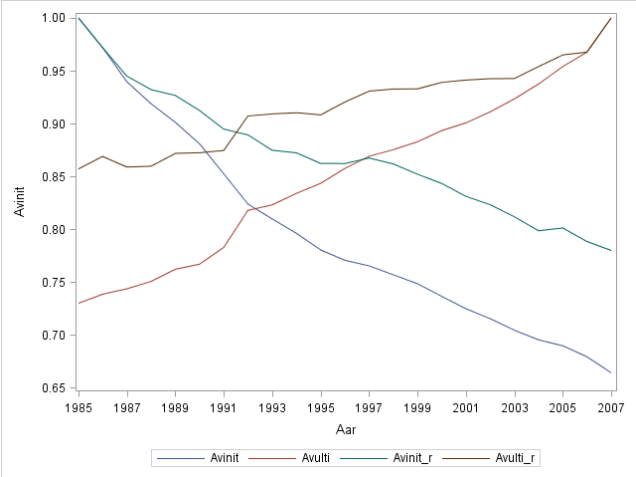


Figure 5: Structural change in manufacturing for Danish regions 1985-2007



Our analysis aims at studying the entire Danish economy and the partitioning method present in Section 3 is well suited for that aim, however, the background for the analysis is the path-dependent model of structural change discussed in Section 2 and this model is obviously more applicable to some sectors of the economy than others. These sectors consist of industries that are subject to selection in the labour market in the sense that the performance of firms in these industries depends heavily on the local availability of specific skills. The performance of firms in other industries, and hence the presence of these industries, is explained by other factors such as social policy (e.g. hospitals), infrastructure policy (e.g. utilities), natural resources (e.g. agriculture) or demand (e.g. supermarkets). These industries are, however, important to include in the analysis since their presence can condition the evolution of other industries. The presence of hospitals, for example, may entail an increase in the local availability of the skills needed in the pharmaceutical industry and hence improve the conditions for growth in the pharmaceutical industry in the region.

We have used information on all workplaces when computing the SIR index and SC values for industries but when computing average characteristics of regions and relative growth of industries we have excluded the non-private sectors, as in Figure 4.

4.2. Detailed result of decomposing the change in average standardised closeness from 1998 to 2001

Based on the graphs in figures 1 and 2 it seems prudent to choose a relatively short time span for studying the change in standardised closeness, as in the later periods (the late 1990s) the correlation between standardised closeness and the growth of a regional industry decreases when growth is measured over longer periods. Experimenting with different starting years and different time spans shows that many of the results are more or less the same, while others differ dramatically. Studying these variations in results is of course important but beyond the scope of the current paper. The following results pertain to decomposing the change from 1998 to 2001.

4.2.1. Evolution in a population of regions: the national level

In 1998 the national average standardised closeness was 0.113, which is not very informative. 11.4% of employment had tertiary education, 17.5% of employment was at plants established in 1994 or later, and 47% of employment was in the four university cities.

$$Z_{1998} = \begin{bmatrix} SC_{1998} \\ SHE_{1998} \\ ENT_{1998} \\ UNI_{1998} \end{bmatrix} = \begin{bmatrix} 0.113 \\ 0.114 \\ 0.175 \\ 0.470 \end{bmatrix} \tag{9}$$

Over the next three years, until 2001, the national average standardised closeness decreased by 0.0033. The shares of employment with higher education, in new establishments and in the four university cities increased by 1.3, 1.2 and 1.0 percentage points respectively.

$$\Delta Z_{1998-2001} = \begin{bmatrix} \Delta SC_{1998-2001} \\ \Delta SHE_{1998-2001} \\ \Delta ENT_{1998-2001} \\ \Delta UNI_{1998-2001} \end{bmatrix} = \begin{bmatrix} -0.0033 \\ 0.0127 \\ 0.0119 \\ 0.0104 \end{bmatrix} \tag{10}$$

The national level change can be decomposed into the effect of some regions growing relatively faster than others and changes at the regional level; regional selection and regional adaptation. The share of employment in university cities only changes by differential growth since regions are either university cities or they are not– there is no adaptation. In order to facilitate interpretation the effects are reported as shares of the total change in equation 11. In order to emphasise that the change in SC is negative the regional selection and adaptation effects are shown as negative.

$$\frac{\Delta Z}{\Delta Z} = \frac{ReSelect}{\Delta Z} + \frac{ReAdapt}{\Delta Z} = \begin{bmatrix} -0.039 \\ 0.098 \\ 0.070 \\ 1.000 \end{bmatrix} + \begin{bmatrix} -0.961 \\ 0.902 \\ 0.930 \\ 0 \end{bmatrix} \tag{11}$$

Equation 11 shows that 4 percent of the change in national average SC can be attributed to regions with low average SC growing more than others, and the remaining 96 percent are attributed to changes within regions. The evolution of SHE and ENT at the national level have a slightly larger role for selection but they are also dominated by adaptation. The regional adaptation effect will be elaborated below. The negative regional selection effect for SC is unexpected, since SC should be positively related to the growth of regional industries (e.g. Essletzbichler, 2015) and hence regions with high average SC should grow more than other regions. According to equation 2 the regional selection effect can be decomposed into the product of the variance-covariance matrix of the characteristics and a vector of selection gradients that illustrate the strength of selection on each variable. The selection gradients are shown in equation 12. The gradients are all positive indicating that regions with higher average SC, regions with more educated employment, regions with a larger share of employment in new establishments and university cities grow more than other regions, ceteris paribus.

$$ReSelec = P\beta = P \begin{bmatrix} \beta_{SC} \\ \beta_{SHE} \\ \beta_{ENT} \\ \beta_{UNI} \end{bmatrix} = P \begin{bmatrix} 0.4074 \\ 0.3276 \\ 0.2434 \\ 0.0003 \end{bmatrix} \tag{12}$$

The result in equation 11 already casts doubt on the path-dependent model of regional change since it shows that regions with higher SC do not grow. The result presented in equation 12 rescues the path dependent model by showing that the selection gradient for SC is positive which must mean that the negative relationship between regional growth and average SC must be caused by covariance with other variables. The selection gradients are not scale free and hence equation 12 does not say anything about the relative importance of selection on the four different variables. However, when multiplying with the matrix P it is possible to show the share of the regional selection effect that is caused by direct selection on the variable in question, and the shares caused by covariance with other variables. This is illustrated in Figure 6 below.

Figure 6: Sources of regional selection

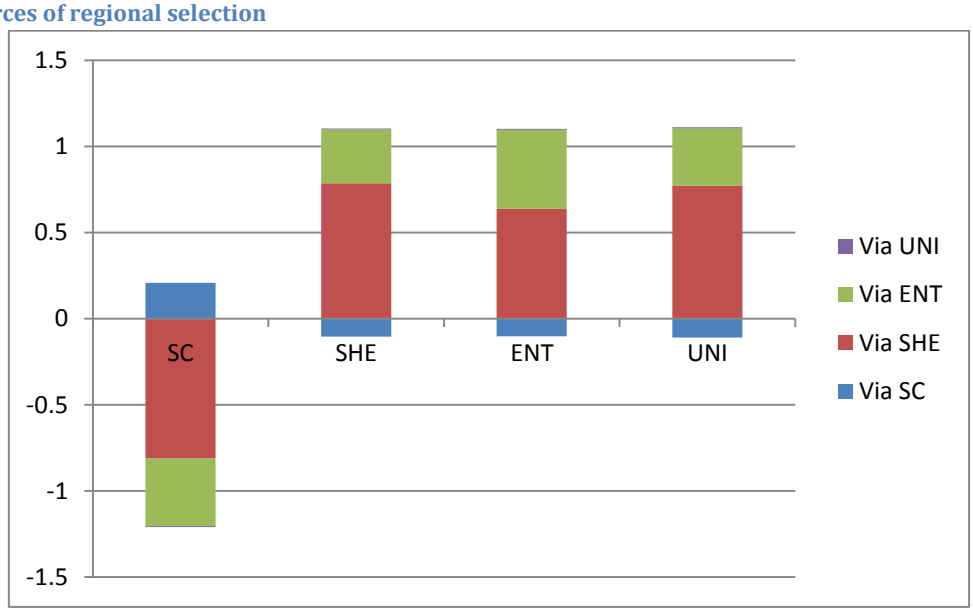


Figure 6 shows that the positive selection on SC at the regional level (blue) is more than cancelled out by selection on SHE and ENT (red and green respectively). Selection on these two variables is also positive, cf. equation 12, but they have negative covariance with SC and hence create negative selection on SC. This negative covariance is also why SC contributes negatively to selection on SHE and ENT. There is hardly any selection at all on UNI and thus it is hardly visible in Figure 6 (purple).

4.2.2. Evolution in populations of industries: the regional level

Equation 11 showed that the lion's share of evolution at the national level is driven by regional adaptation, and the adaptation of any one region can be decomposed into the effects of selection on industries, adaptation of industries, entry of industries and exit of industries, cf. equation 8.

Regional adaptation can be decomposed into average industry selection – the effect of some industries expanding more than others within a region; averaged over the 71 regions of Denmark, industry adaptation, industry entry and industry exit. The result of this decomposition is shown in equation 13. As before, the components of the change in SC are shown as negative so that they sum to negative 1.

$$\frac{ReAdapt}{ReAdapt} = \frac{InSelect}{ReAdapt} + \frac{InAdapt}{ReAdapt} + \frac{InEntry}{ReAdapt} + \frac{-1*InExit}{ReAdapt} = \begin{bmatrix} -0.043 \\ 0.257 \\ 0.202 \\ 0 \end{bmatrix} + \begin{bmatrix} -0.949 \\ 0.730 \\ 0.740 \\ 0 \end{bmatrix} + \begin{bmatrix} -0.011 \\ 0.016 \\ 0.044 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.003 \\ -0.003 \\ 0.014 \\ 0 \end{bmatrix} \quad (13)$$

Here, again, selection on SC is negative meaning that industries with low average closeness grow more than other industries. The primary source of change in average SC at the regional level is industry adaptation. This means changes in the relatedness matrix and changes caused indirectly by the entry or exit of other industries. The (direct) contribution of entry is negative and the contribution of exit is positive meaning that both industries that enter a region and industries that exit a region tend to have below average SC. This is consistent with the results of earlier studies, such as Neffke et al. (2011) and Essletzbichler (2015).

Selection on the education level of the workforce and on the propensity to establish new plants are positive for the regional level and contribute a relatively large share (26 and 20 percent respectively) to the evolution of these two variables, though adaptation at the industry level is still the dominating factor. This means that the regional averages for the shares of the workforce with tertiary education and the share working in new plants primarily change because industries become more human capital intensive and employment in new establishments increase. There are also positive effects of entry meaning that entering industries tend to be human capital intensive and include new plants. The latter is hardly surprising. Finally, the exit effect is small and negative for SHE and positive for ENT meaning that exiting industries tend to have above average human capital intensity and a below average share of workers in new plants.

The decomposition presented in equation 13 is an average over the 71 regions in the data. The distribution of the regional level effects are illustrate in Figure 7 to show the variations in these effects across regions.

Figure 7: Regional variations in industry level effects

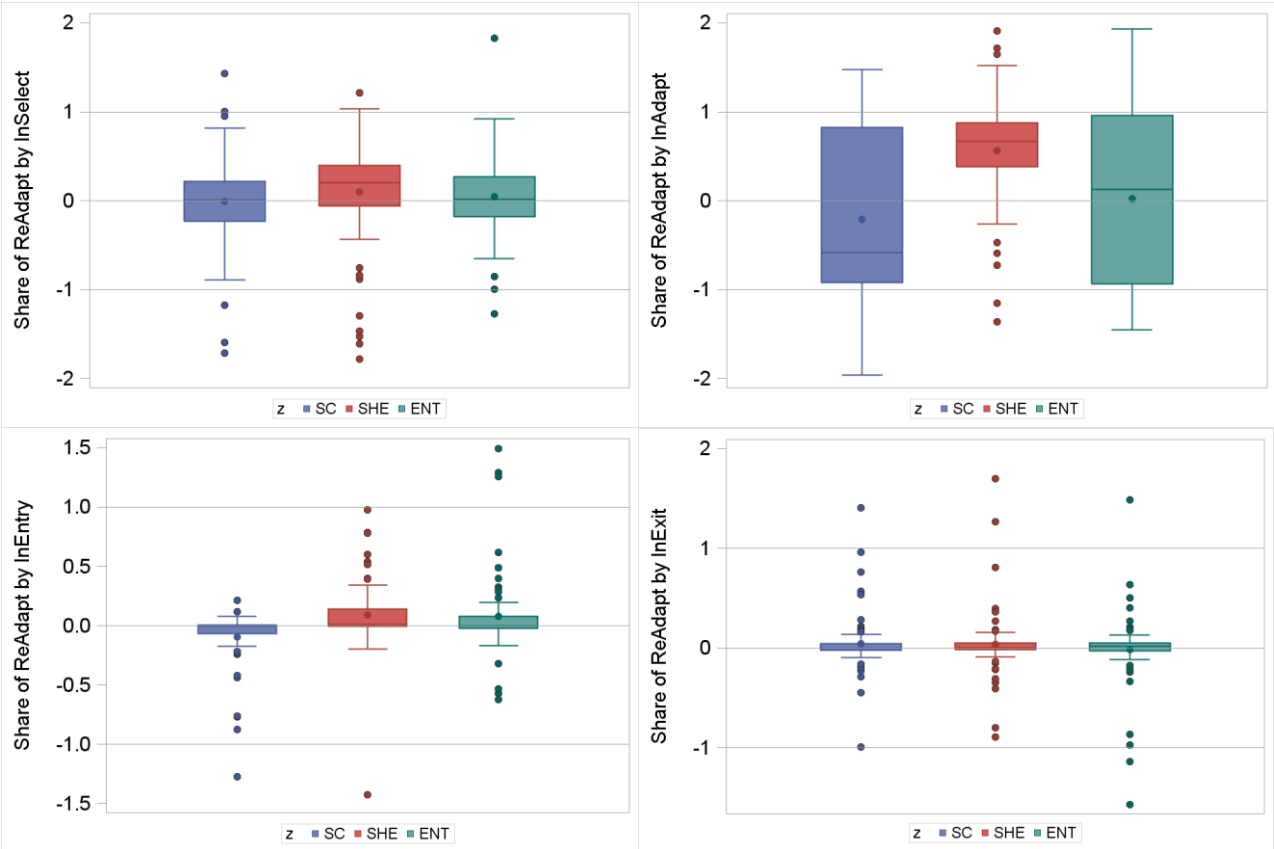


Figure 7 contain boxplots where the boxes indicate the inter-quartile range. The mean (circle) and the median (line) are indicated within the box and the whiskers illustrate 1.5 times the inter-quartile range (or the min/max if there are no outliers).

The smallest regions (in terms of employment) tend to produce very large values for the different effects that are not easy to interpret when the regional evolution is decomposed. Since these are the smallest regions, it does not have a significant impact on the average decomposition in equation 13 but it means that effects above 2 or below -2 have been omitted from Figure 7. Since the entry and exit effects generally constitute only a small share of total change, the plots in Figure 7 have more scope to illustrate outliers for these effects. Thus, it seems from Figure 7 that the entry and exit effects have a greater number of extreme cases than the selection and adaptation effects, but this is not the case.

The top left plot in Figure 7 shows that the industry selection effects for SC and ENT are generally a minor share of total evolution at the regional level though the whiskers extend almost to 1 indicating that there is a number of cases where the selection effect is dominating. Selection is more consistently positive for SHE. Adaptation is generally the dominating component and can be either positive or negative for SC and ENT while it is generally positive for SHE. Taken together this shows that in most regions the human capital intensive industries expand and industries generally become more human capital intensive. However there is much variety regarding the role of new establishment creation and the presence of skill related industries (SC) across regions.

The final step in the partitioning methodology is to decompose the industry selection effect into the average of the product of the regional variance-covariance matrices and the average selection gradients.

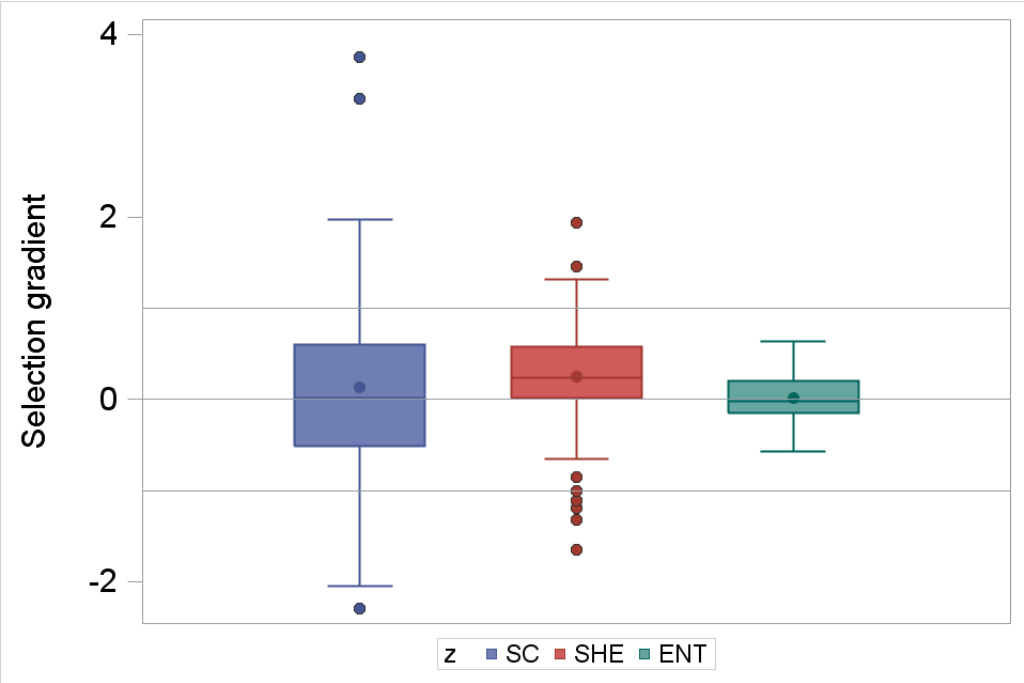
$$InSelec = \sum_r s_r' P_r \beta_r \tag{14}$$

This is not easy to report concisely but the average selection gradients are reported in equation 15, though they cannot be used directly to compute the industry selection effects.

$$\sum_r s_r' \beta_r = \begin{bmatrix} -0.0936 \\ 0.3206 \\ 0.1901 \\ 0 \end{bmatrix} \tag{15}$$

Surprisingly the average regional selection gradient for SC is negative meaning that industries with below average standardised closeness grow more than others. In other words, for a given industry the presence of a greater number of related industries in a region leads to less growth for the industry in question. In order to illustrate the variation in selection gradients across regions the distributions of the regional selection gradients are shown in Figure 8.

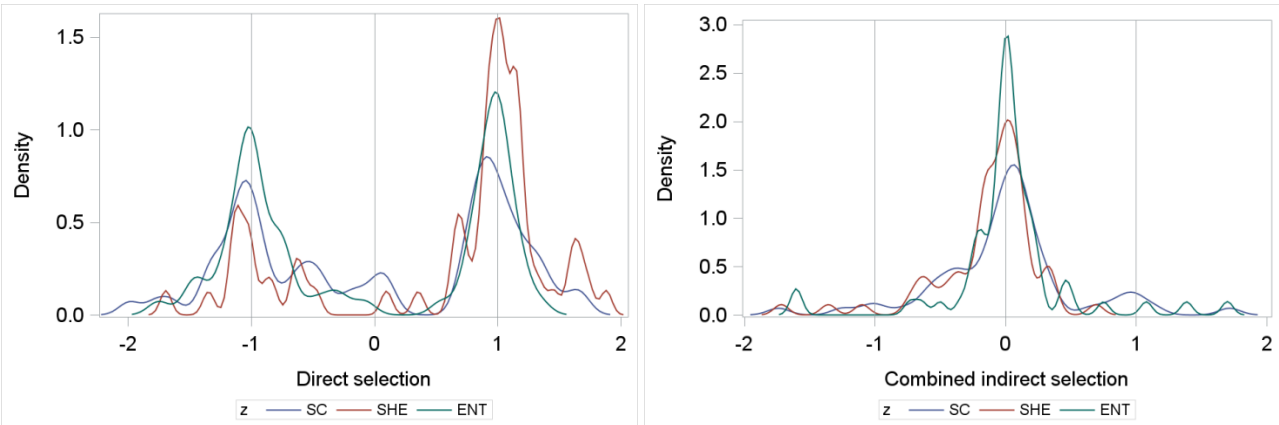
Figure 8: Regional variations in selection gradients



As mentioned earlier the selection gradients are not scale free and hence cannot be compared directly for different variables. Hence what Figure 8 shows is that there is a lot of regional variation in selection on SC. The inter-quartile range (i.e. the box) stretches almost half way from zero to the reference lines at positive and negative one. Interestingly the *average* selection gradient reported in the plot, the circle in the blue box, is slightly positive while the weighted average reported in equation 15 was negative. This leads to a suspicion that selection on SC is more often negative in large regions.

In Figure 6 it was show that, at the aggregate level, covariance with SHE and to some degree with ENT domines selection on all four variables. At the regional level each variable generally account for the full selection effect itself, and the combined indirect selection of the two remaining variables is close to zero. This is illustrated with the two sets of kernel density graphs in Figure 9.

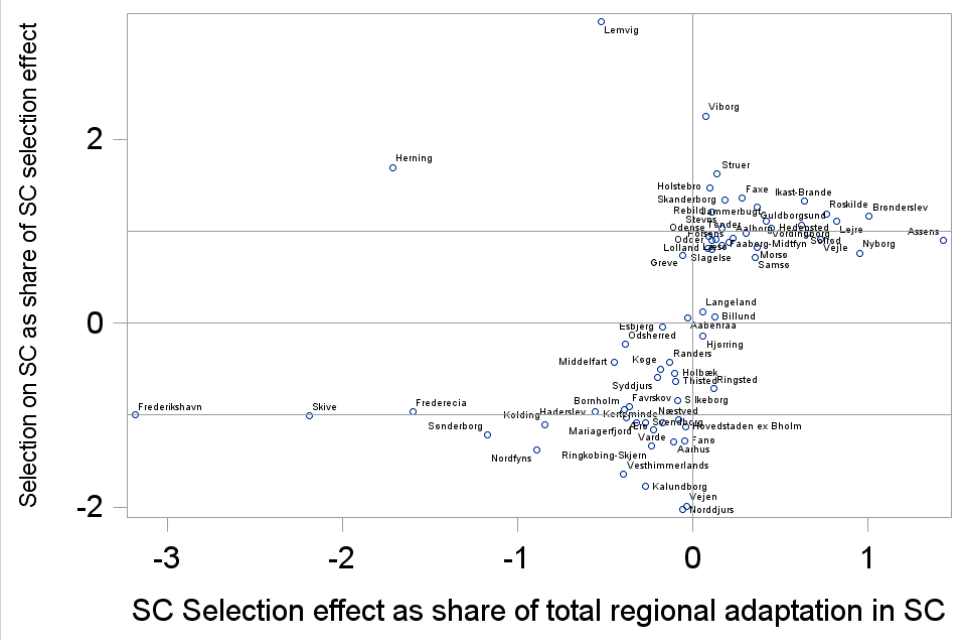
Figure 9: the balance between direct and indirect selection at the regional level



4.2.3. Changes in activities at the regional level

Selection among industries – or the differential growth of industries – only accounts for a small share of the change in average SC in most cases cf. Figure 7. This can also be seen in Figure 10 where most of the regions cluster near the vertical reference line at zero indicating that the industry selection effect is a small share of the total regional adaptation. However, the industry selection effect is not necessarily uninteresting simply because it constitutes a small share of total regional adaptation. In fact, cases where the selection effect is a small share of the change in average SC may be the most interesting cases since other factors than the presence of related industries dominate the change in activities. In this paper we specifically study entrepreneurship and human capital intensity as these "other factors". Figure 10 additionally shows that there is wide variation in the degree to which the industry selection effect for SC is actually caused by selection on SC or caused by covariance with entrepreneurship and human capital intensity creating indirect selection for SC.

Figure 10: SC selection effect vs. selection on SC



For regions with negative selection on SC that are located near the reference line at -1, e.g. "Skive", the decrease in average SC caused by differential growth of industries in Skive is attributed completely to negative selection on SC: industries with high SC grow less than others. Thus, the evolution of Skive exhibits path exhaustion as there is no more growth for related industries in the region. The direct opposite is the regions of "Assens", where the selection effect for SC is positive, accounts for a large share of total change in average SC and is caused by positive selection on SC: industries with high SC grow more than others. Thus, the evolution of the activities undertaken in Assens exhibits path extension.[1]

[1] The fact that direct selection on SC accounts for the entire evolution of average SC in Skive and Assens does not rule out that there are indirect selection on SC too. It only entails that any indirect selection must necessarily cancel out.

Figure 11: Indirect effect on the SC selection effect

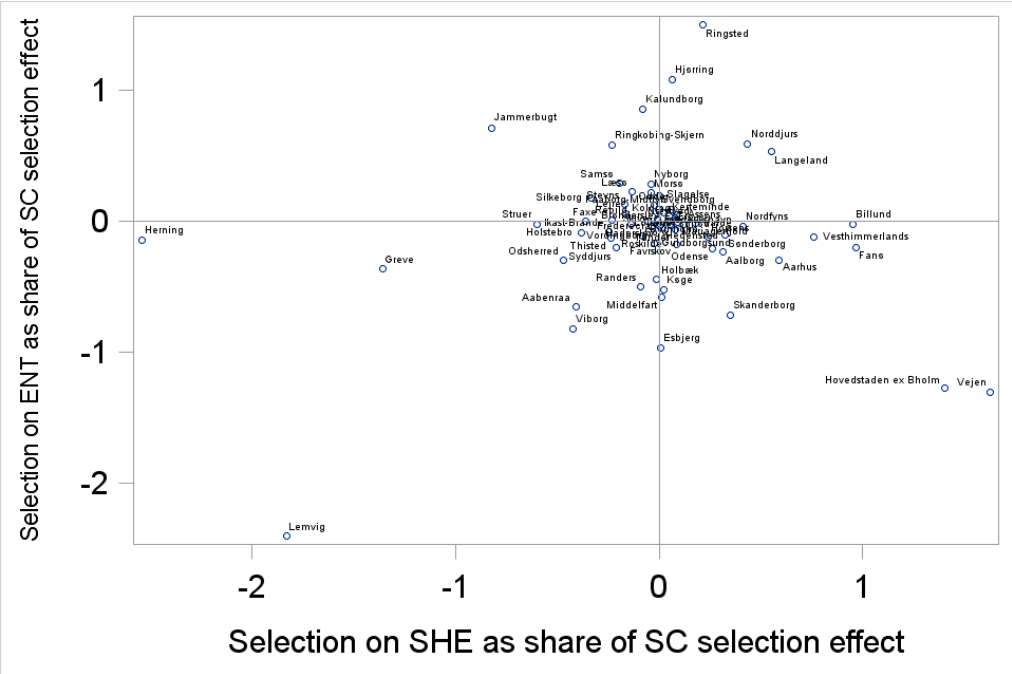


Figure 11 shows the indirect components in the selection effect for SC. Most regions are clustered near (0,0) indicating that there are no indirect effects, as was also shown in Figure 9. In these cases the selection effect for SC is directly caused by selection on SC. In the top right corner of Figure 11 are regions where selection on SHE and ENT has increased the selection effect for SC. That is, in these regions both human capital intensity and entrepreneurship are associated with growth, and covariance between these characteristics and standardised closeness leads differential growth to increase standardised closeness. In the bottom left the opposite is the case and this explains why the regions Herning and Lemvig were shown in Figure 10 to have a negative relationship between SC and industry growth (a negative selection effect), while at the same time industries with higher SC actually grow more (a positive direct contribution from SC). The reason is that selection on the characteristics SHE and ENT coupled with covariance between these characteristics and SC more than cancelled out the effect of direct selection on SC. In the final two quadrants of Figure 11 are the cases where the indirect effects cancel each other out, e.g. the peripheral Jammerbugt and the capital Hovedstaden regions. These two regions are difficult to spot in Figure 10 since they have low selection effects which are completely accounted for by direct selection on SC. However, underneath this we find large but opposite components from indirect selection.

Figure 10 shows what regions are likely to be experiencing path exhaustion and path extension while Figure 11 adds clarification and certainty by illustrating the relative magnitudes of the indirect components. However, Figure 11 does not say anything about the direction of selection on the variables creating the indirect components. For example, in the Jammerbugt region there is a positive indirect contribution from selection on ENT to the selection effect for SC, but it is not possible to see in Figure 11 whether this is because there is negative selection on ENT coupled with negative covariance between ENT and SC, or there is positive selection on ENT coupled with positive covariance between ENT and SC. Figure 12 therefore plots the selection gradients showing the direction of selection for SHE and ENT. Only regions where the sum of the absolute values of the two variables in Figure 11 exceeds 0.5 are included in Figure 12

since observations clustering around (0,0) in Figure 11 have low indirect components and hence the selection gradients for the indirect components are not interesting.

Figure 12: Selection gradients for SHE and ENT

Figure 12 contains some unexpected results. It was expected that entrepreneurship and human capital intensity would lead to growth in an industry, and therefore that these two factors would create breaks from path extension and path exhaustion. This is only the case for regions in the top right of Figure 12, while in Norddjurs (in the lower left corner) industries consisting of old plants and industries with low human capital intensity are expanding. Considering Norddjurs throughout the three figures 10-12 it can be seen that the region represents an interesting case. The selection effect is only a small part of the regional adaptation in SC, however, there is quite strong negative selection on SC (Figure 10). This is possible because there are positive indirect contributions from selection on SHE and ENT (Figure 11) but selection is negative on both variables indicating that the both have negative covariance with SC (Figure 12) or else their contributions could not be positive. There is definitely not path extension or renewal in Norddjurs since related activities do not grow. Norddjurs instead seems to be a prime example of path exhaustion: the activities that grow in relative terms are undertaken in old plants with low human capital intensity, and such activities tend to have high SC in Norddjurs, which thus pushes average SC back up. Thus, Norddjurs seems to be reverting to a specialisation in relatively simple activities.

A prime example of a region exhibiting path renewal is Skanderborg. The region may be hard to spot in Figure 10 since the selection effect for SC is only a small share of the reginal adaptation of SC (near 0 on the horizontal axis) but selection on SC is strong indicating that industries with high SC do expand (near 1 on the vertical axis). The reason why average SC then does not increase is that there are strong indirect components in the selection effect (Figure 11) caused by positive selection gradients for both SHE and ENT (Figure 12). Thus in Skanderborg related industries do expand but so does industries with high entrepreneurship and/or human capital intensity and correlation between these two variables and SC entails that differential growth does not have a large effect on average SC.

Finally, a prime example of path creation is Ringkobing-Skjern. Like the Skanderborg region, the Ringkobing-Skjern region has positive selection gradients for both SHE and ENT (Figure 12) resulting in indirect effects on the selection effect for SC (Figure 11) but unlike Skanderborg there is no positive direct effect of selection on SC (Figure 10). In fact, selection on SC is strong and negative leading to differential growth causing a decrease in average SC. Thus in Ringkobing-Skjern there is no expansion into related activities but instead human capital intensity and entrepreneurship are leading the region to expand into new and unrelated activities.

One of the more interesting, though not necessarily surprising results from figures 10 to 12 is that the diversity in regional change is much larger than we can capture with the four categories in the taxonomy of path extension, renewal, exhaustion and creation. Whether this can be solved with more data or the taxonomy needs additional categories is left for further studies.

5. Discussion and conclusion

In this paper the evolution from 1998 to 2001 of the four variables: average standardised closeness (SC), share of workers with tertiray education (SHE), share of workers in new plants (ENT) and share of workers working in univeristy urbanisations (UNI) has been mathematically decomposed into a list of effects. It was shown that the evolution of the first three variables at the national level has been the result of change within regions and only to a modest degree to differential growth of regions. The evolution of the fourth variable, UNI, has by definition been driven solely by differential growth. An unexpected result was that differential regional growth has led to a decrease in average SC, even when it was confirmed that the selection gradient for SC is positive. The explanation to this contradiction is that SC correlates negatively with other variables, especially SHE, and these variables also have positive selection gradients. Thus, regions with high average SC grow ceteris paribus, but other things are not equal, as such regions tend to have low values of both SHE and ENT leading to low growth in average SC and mitigating the positive effect of SC. There is no reason to expect that the model of path-dependent evolution of a regional industry structure is also applicable to the evolution of a nation consisting of a population of regions. But it is nevertheless pivotal to include this step in the analysis in order for inter-regional dynamics to be separated out before studying the intra-regional dyanmics in the ensuing steps.

While evolution at the national level is mathematically speaking the outcome of population dynamics in the population of regions, the evolution of regions are the outcome of population dynamics among the regional industries. Across the 71 regional populations of industries the average selection on SC is also negative. This can, in general, be traced to a negative selection gradient and not to covariances with other characteristcs. This is puzzling since other studies (e.g., Neffke et al., 2011) have found that the presence of skill related industries in a region is conducive for the growth of such industries, which should be reflected in a positive selection gradient. However a preliminary hypothesis for the cause of this result was observed when the distribution of the unweighted regional selection gradienst for SC was plotted, as it turned out that the unweighted average gradient is indeed positive. Thus the negative result can be driven by differences in the evolutionary processes between large (urban) regions and smaller regions. A more comprehensive study should include additional analyses of individual regions to uncover such differences in evolution.

The primary aim of this paper is to identify cases where the positive relationship between regional industry growth and the presence of related industries does not explain the change in average relatedness, and thus indentify which factors that lead to regions evolving in less path dependent ways. It turned out that the positive relationship was elusive in the first place. This is probably caused by the industry classifications being too aggregate. The relatively aggregate industry classifications were chosen so that the introductory graphical descriptions of the relationship between relatedness and growth and of structural change could cover a longer time period. However this may have created problems for the enusing decomposition analysis. Later studies should explore this possibility.

Despite this drawback, the analysis still casts light on the question of path creation. Out of the four variables studied here, human capital intensity and entrepreneurship dominate the evolution of regions in the country. These two factors are more important for growth *of regions* than standardised closeness and their negative covariance with standardised closeness makes the latter almost irrelevant. Covariances matter less in the average regional population *of industries*. Here selection is more direct and highly diverse. The results show that human capital intensity and entrepreneurship indeed do mediate the role of relatedness in structural change. Surprisingly it was shown that this mediating effect may go in any direction: relatedness, in form of the standardised closeness variable, has both positive and negative covariance with entrepreneurship and human capital intensity, and selection on the two variables may also be either positive or negative. Therefore, it is not clear that human capital or entrepreneurship are the sources of new path creation or path renewal as suggested by Isaksen (2015). The analyses show that other factors may lead to new paths dominated by existing firms or low human capital intensity.

This study has focussed on the change in economic activities within regions but we do not study the change in scale of activities. Hence it is not clear whether path creation, as identified here, is desirable in for example the sense of creating more jobs. The main contributions are instead the limits and caveats identified for the model of path-dependent structural change, and secondly the first empirical test of a new methodology. The contributions add to our understanding of structural change and, as more variables are added in later analyses, adds to the potential for uncovering the role of active policy in mediating structural change.

6. References

Andersen, E. S. (2003). “The evolving tree of industrial life: An approach to the transformation of European industry”. Paper for the second workshop on the Transformation of Europe, 31 Jan - 2 Feb, Turin, Italy.

Andersen, E. S. (2004). “Knowledges, specialisation and economic evolution: modelling the evolving division of human time”. In Foster, J. and Metcalfe, J. S., editors, Evolution and Economic Complexity, pages 108–150. Edward Elgar, Cheltenham and Northampton.

Boschma, R. and Frenken, K. (2011) “*Technological relatedness and regional branching*”. In H. Bathelt, M. P. Feldman, D. F. Kogler (eds) Dynamic Geographies of Knowledge Creation and Innovation. London: Routledge, Taylor and Francis.

Cohen, W.M., Levinthal, D.A., (1990). Absorptive capacity: a new perspective of learning and innovation. *Administrative Science Quarterly* 35 (1), 128–152.

Essletzbichler, J. (2015). “Relatedness, Industrial Branching and Technological Cohesion in US Metropolitan Areas,” Regional Studies, 49(5), 752-766.

Frenken, K., Van Oort, F., and Verburg, T. (2007). “Related variety, unrelated variety and regional economic growth.” *Regional Studies*, 41 (5), 685-697

Hausmann, R. and Hidalgo, C. A. (2011) “The network structure of economic output”. *Journal of Economic Growth* 16(4):309-342.

Hidalgo, C. A., Klinger B., Barabási A.-L., and Hausmann R. (2007): "The Product Space Conditions the Development of Nations." *Science* 317, no. 5837 : 482-487

Holm, J. R., Andersen, E. S., and Metcalfe, J. S. (Forthcoming A). “Confounded, augmented and constrained replicator dynamics: Selection in multiple traits and how to measure it. “*Journal of Evolutionary Economics*.

Holm, J. R., Østergaard, C. R., and Olesen, T. R. (Forthcoming B) “Destruction and Reallocation of Skills Following Large Company Closures”. *Journal of Regional Science.*

Isaksen, A. (2015). “Industrial development in thin regions: trapped in path extension?” *Journal of Economic Geography* 15 (3): 585-600.

Marshall, A. (1920). *Principles of Economics*. Macmillan, London, 8th edition

Martin, R. and Sunley, P. (2006). Path dependence and regional economic evolution. *Journal of Economic Geography* 6, 395-437.

Metcalfe, J. S. (2001) “Institutions and progress”. *Industrial and Corporate Change* 10(3):561-586.

Neffke, F., Henning, M. and Boschma, R. (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.; Hidalgo, C.; Otto, A. and Weyh, A. (2013). “Inter-industry labor flows.” Cambridge: Center for International Development, Harvard Kennedy School.

Saviotti, P. P. and Frenken, K. (2008). “Export variety and the economic performance of countries.” *Journal of Evolutionary Economics*, 18(2):201–218

Yang, X. and Ng, Y.-K. (1993) “*Specialization and Economic Organization”*. North-Holland, Amsterdam.

Young, A. A. (1928). “Increasing returns and economic progress.” *The Economic Journal*, 38(152):527–542.