Regional labour market structures

Regional labour market structures

David Sijmons
Jessie Bakens

ROA Technical Report

ROA-TR-2020/4

Researchcentrum voor Onderwijs en Arbeidsmarkt | ROA
Research Centre for Education and the Labour Market | ROA

Regional labour market structures

David Sijmons
Jessie Bakens

ROA-TR-2020/4
July 2020

Research Centre for Education and the Labour Market
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
T +31 43 3883647 F +31 43 3884914

secretary-roa-sbe@maastrichtuniversity.nl
www.roa.nl

ISSN: 2666-884X

Contents

Acknowledgements

This research is part of Education and the Labour Market Project (POA; https://www.roa.nl/research/research-projects/project-onderwijs-arbeidsmarkt-poa), funded by Nationaal Regieorgaan Onderwijsonderzoek (NRO; file number 405-17-900), the UWV Werkbedrijf, the Cooperation Organisation for Vocational Education, Training and the Labour Market (SBB), and employment agency Randstad Netherlands. Four Dutch ministries contribute to the NRO funding: the Ministry of Education, Culture and Science (OCW), the Ministry of Interior and Kingdom Relations (BZK), the Ministry of Social Affairs and Employment (SZW), and the Ministry of Agriculture, Nature and Food Quality (LNV). It is based on the internship work of David Sijmons as part of his Master thesis in Econometrics and Operations Research at Maastricht University School of Business and Economics.

1 Introduction

Heterogeneity in economic structures between different countries or regions has been examined by economists for a long time. Much of the literature has focused on explaining (and quantifying) interregional differences in production and income. However, little attention is focused on explaining and quantifying differences in occupational and educational compositions of the workforce across regions. This is the gap this paper tries to fill by analyzing regional occupational and educational structures in the Netherlands.

More specifically, the first aim of this paper is to examine to what extent regional sectoral occupational and educational structures currently are heterogeneous across regions in the Netherlands. In other words, we look at whether all regions currently utilize the same kind of human capital for their production in a given sector or not. The second aim of this paper is to investigate whether there are any trends in the development of this heterogeneity across regions in the Netherlands over time. This provides us with an answer to the question whether regional sectoral occupational and educational structures have been converging over time or not (in the sense of σ-convergence). It is important to note that this paper is primarily occupied with *quantifying* heterogeneity in sectoral occupational and educational structures and its development over time, but not so much with explaining *why* the observed heterogeneity occurs. Only for the observed current heterogeneity it is examined to some extent what could have generated this heterogeneity. A similar discussion for the dynamic analysis of heterogeneity is omitted. This is a task that remains for future research.

This research is part of ROA's project on labour market forecasts (POA), and focuses on a better understanding of the regional component of the labour market forecasts. In constructing the regional forecasts, ROA currently assumes that there is no heterogeneity in the sectoral occupational and educational composition of regions. More specifically, we assume that the sectoral composition of each of the 35 distinguished regions match the national sectoral composition. Answering the previously presented research questions will give an indication of the validity of this assumption, and how the validity is changing over time.

Besides providing an answer to the two main research questions at hand, at the end of this research a related extension is provided which is directly relevant to ROA's forecasting procedure. More specifically, an exploratory analysis is conducted to see whether the forecasts of regional sectoral occupational and educational structures can be improved by using data on regional structures, instead of only using data on national structures. Note that forecasting these regional structures is necessary in order for ROA to obtain the final labour market forecasts of occupation and education supply and demand.

Looking at current average deviations from the national share for individual occupations or educations across regions, it turns out that there are only a few combinations of occupation/education and sector for which there seems to be relatively much heterogeneity across regions. We also find structural divergence over time in regional educational and occupational distributions, but at a very small pace. We also find that the current regional forecast can be improved by including information on regional sectoral educational and occupational differences for those sectors for which this information is available. The fact that for some region/sector combinations information is not available or unreliable due to a small number of observations illustrates the difficulty of implementing this strategy for the full regional forecasts.

The paper starts in Section two by examining why one would observe heterogeneity in regional sectoral occupational and educational structures in the first place. Furthermore, this section discusses the already existent literature on heterogeneity in occupational and educational distributions. This literature gives some guidance as to which methodology can be used to examine the main research questions of this paper. Afterwards, Section three discusses the data which is used in this paper. Subsequently, Section four provides some remaining background information on the Dutch educational system and the Dutch economy. This information is helpful in understanding the findings presented later on in this paper. Section five discusses the methodology used to quantify heterogeneity in regional sectoral occupational and educational distributions and its development over time. The heterogeneity results for the regional sectoral occupational distributions are then presented in Section six and the heterogeneity results for the regional sectoral educational distributions are presented in Section seven. Afterwards, Section eight examines the observed current heterogeneity patterns in more detail, so as to be able to better understand what generates this heterogeneity. This will be done at the level of specific occupations and types of education. Lastly, Section nine provides an extension directly related to POA's forecasting model. This section provides an exploratory analysis of whether the forecasts of regional sectoral occupational and educational structures can be improved with the currently available data. Finally, Section ten concludes and described the implications of this results for the POA regional labour market forecasts.

2 Theoretical Background and Literature

The aim of this paper is to examine to what extent regional sectoral occupational and educational structures differ across regions in the Netherlands. To enhance our understanding of the findings presented later on in this paper, it is important to understand why such regional differences would exist in the first place. This will be the topic of the first subsection. The second subsection will discuss some literature on the specific topic of comparing occupational and educational compositions across regions. This provides us with some guidance on which methodology can be used to make such comparisons.

2.1 Rationale for the Existence of Heterogeneity in Regional Sectoral Occupational and Educational Structures

What causes regional sectoral occupational and educational structures to differ? One obvious answer to this question is within-sector regional specialisation patterns in output. To put it simply, not all regions produce the same bundle of goods and/or services within a certain sector. As a result, each region has its own specific human capital requirement in terms of which occupations and types of education are needed. Naturally, this leads one to ask the question why within-sector regional specialisation in production occurs in the first place. The answer to this question can be found in the large body of literature on spatial economics.

The spatial economics literature is predominantly occupied with explaining regional production patterns, both within and across countries. For many decades, the Ricardian model was the dominant theory in this regard (Ricardo, 1821). This model explains heterogeneity in production patterns across regions by technological disparities. One region may know how to produce excellent wine, whereas another region is not so good at producing wine but is excellent at producing clothes. Both regions are then better off by specialising in what they are most productive at, instead of producing both goods themselves. In the end they can trade wine for clothing, making both regions better off. This example illustrates the main line of reasoning of the Ricardian model. An important extension to this model was developed by Ohlin (1935). Ohlin's model became known as the Heckscher-Ohlin model of trade, and distinguishes itself from the Ricardian model by also taking into account differences in factor endowments. The main conclusion from this model is that, all else equal, regions will tend to specialize in the production of goods/services that are relatively intensive in the factors of production that region is relatively abundant in.

In time, more and more empirical evidence started to appear that indicated that these models are substantially awed. The empirical findings did not match the theoretical expectations formed by these models. Most notably, Leontief et al. (1953) showed that U.S. exports were less capital intensive than its imports, indicating that the country did not specialise in capital intensive products/services. Given that current models at the time were unable to explain specialisation patterns across regions, new theories started to be developed. Most notably, Krugman (1991) developed a model to explain regional specialisation patterns by means of external economies of scale. External economies of scale are any benefits to firms which arise from the regional clustering of firms. The idea of linking external economies to regional specialisation had already been introduced by Marshall (1961), but Krugman was the first to formalise this. Krugman's model is able to explain specialisation across regions which have similar levels/types of technology and resource endowments. Note that models in the style of Ricardo and Ohlin would predict no regional specialisation in such a case.

Rather than replacing existing trade theories all together, Krugman's model gave rise to a new strand of literature on international and interregional trade, combining aspects from different theories. This strand of literature has been termed the New Economic Geography (NEG), and is comprehensively described by Krugman himself (1998). Although not so new anymore, the NEG is still a widely used framework to evaluate and understand regional differences in production patterns. Largely following Krugman (1998), let us briefly examine how the NEG explains regional specialisation patterns. The core idea underlying the NEG, is that for any decision of where to produce, there are always both forces that tend to promote geographical concentration and forces that tend to oppose it. Krugman names these forces respectively `centripetal' and `centrifugal' forces. He discusses several of the most important of such forces. These will also be briefly outlined here. Note that the centripetal forces are mainly related to external economies of scale. External economies of scale can be defined as the benefits which arise for firms and people from locating near one another together in industrial clusters and cities (Glaeser, 2010). The existence of such external economies of scale is widely accepted in the literature, and has been extensively documented. For a comprehensive discussion of this, see e.g. Glaeser (2010).

The first centripetal force discussed by Krugman (1998), is that the concentration of economic activities creates both backward and forward linkages. Secondly, by concentrating their activities, firms enhance the creation of a so called `thick' labour market. This is a labour market in which the cost for employees to find a new job is relatively low, and the cost for organisations to find new employees is relatively low. A last centripetal force discussed by Krugman, are the external economies of scale resulting from knowledge spillovers. The clustering of economic activity naturally creates an environment relatively favourable to information sharing. As indicated, there are also forces which tend to oppose geographic concentration. The first such force discussed by Krugman, is the fact that some economic factors are by definition immobile. This can be input factors, such as land or natural resources. Such factors can only be relocated at a relatively high cost, and sometimes not even at all. However, it can also be demand side factors. Some production will have an incentive to locate close to consumers. A second centrifugal force is that the concentration of economic activity drives up land rents in the areas in which the concentration occurs, thereby making the decision to locate near an economic cluster less appealing. A last centrifugal force discussed by Krugman, is that concentration can lead to external diseconomies of scale. An example of this is the congestion which occurs in densely populated areas.

How then can such centripetal and centrifugal forces explain within-sector regional specialization in production? The answer to this question is rather simple. For each type of activity within a sector such forces are at play. The resulting balance from all these forces determines where a certain activity locates. If these forces cause certain activities within a sector to locate in different regions than other activities, within-sector specialisation occurs. Note that this paper will not dig any deeper into the literature on regional specialisation in output, as this is only indirectly related to the topic of this paper. Only when the paper considers sectoral heterogeneity across regions at the level of specific occupations and types of education, it will be looked at why the observed heterogeneity patterns occur. It is there that these theories provide helpful, as they guide us on how to explain within-sector regional specialisation patterns. In general, however, these theories on geographical specialisation in output serve merely as a background for the topic under study.

Within-sector regional specialisation is probably the most important reason why heterogeneity in regional sectoral occupational and educational structures exists. However, it is not the only reason. Even if all regions produce the same bundle of goods and services within a certain sector, it may still be possible to observe differences in occupational and educational structures across regions. In this regard it is important to realize that the execution of certain tasks is often not limited to one type of educational background. People with different educational backgrounds are often perfectly able to perform the same task. The educational structure of a certain sector in a region is therefore probably also strongly linked to the type of educational backgrounds which are supplied in that region. Another reason why we can still observe regional heterogeneity in structures even though the final production is the same across regions, is that some regions have a slightly different production process than others to get to the same end product. These regions therefore require workers with different human capital, even though the final product is the same.

2.2 Literature and Expectations

The previous subsection discusses reasons for heterogeneity in regional sectoral occupational and educational structures, the most important one being within-sector regional specialization in production. To form expectations with regard to the extent, nature and development of regional sectoral heterogeneity in structures in the Netherlands, one should consider how these discussed factors play a role. It can, for example, be looked at to what extent within-sector regional specialisation in production is expected in the Netherlands, and how this could affect regional heterogeneity in occupational and educational structures in the Netherlands. Obviously one expects some heterogeneity, however, the nature, extent and development of heterogeneity are difficult to predict. There are very few comparable previous studies examining heterogeneity in regional sectoral occupational and educational structures in the Netherlands. Even if one searches for such studies in an international context, only a few indirectly related papers can be identified. All of these papers provide comparative analyses of certain occupational structures. The most important ones are discussed below. Do note though, that as these papers focus on different geographic regions, their relevance with regard to forming heterogeneity related predictions in the Netherlands is limited. However, they do provide useful insights in that they indicate what methodology can be used to compare occupational and educational structures. Therefore, in discussing these papers, the focus lies on the methodology used in these papers. This methodology will be further built upon in the methodology section of this paper.

Firstly, Barbour and Markusen (2007) examined to what extent metropolitan occupational structures in California and the rest of the U.S. are the same. Their theory is that innovative and developmental activities will be anchored in regions of origin, while more routine production and service functions will be dispersed to lower cost and downstream consuming regions. Finding disparities in occupational structures between California and the rest of the country can provide evidence in favour of this theory, California resembling a `region of origin'. They end up finding that metropolitan structures in California in fact have very similar occupational structures as compared to metropolitan clusters in other parts of the country. More importantly, to quantify heterogeneity in occupational structures they develop an index themselves. For a certain Californian metropole, this index sums over all sectors of production the absolute value of the deviation between an occupation's expected share in that sector and the actual share in that sector in that specific metropole. The expected share is simply chosen to be the national share. This is thus a method which could also be used to answer the research questions presented in this paper.

Another comparison of occupational structures across regions has been done by Scherer and Folch (2017). In their paper they compare occupational structures between cities in Brazil and the U.S., to get an idea of to what extent a country's stage of economic development is associated with a certain occupational structure. To quantify the extent of heterogeneity in occupational structures, this paper makes use of the Krugman Dissimilarity Index, also known as the Krugman Specialisation Index. This index was developed originally by Krugman (1993). In a very similar fashion to the index used by Barbour and Markusen (2007), the Krugman index in this context sums absolute deviations of a baseline city's occupational shares from a reference city's occupational shares. This is thus another method which could potentially be employed by this paper to investigate regional sectoral heterogeneity in the Netherlands.

A third article on this topic which is worth mentioning, is the article by Cörvers and Meriküll (2007) on occupational structures within the European Union. In this paper looks into what extent E.U. countries differ in their use of skilled versus non-skilled workers, and how this has developed over time. In doing so, they split their paper roughly into a static part and a dynamic part. This also what we will do in the analysis in this paper. To investigate their research questions, Cörvers and Meriküll (2007) make use of a so-called shift-share analysis to decompose observed cross-country differences in occupational structures into within-industry and between-industry effects. They use the shift-share analysis in a similar fashion as Esteban (2000) has used it. They end up concluding that new E.U. member countries employ a lower share of skilled workers because they use fewer skills within industries and because their industry structure is biased towards less skill-intensive industries. Furthermore, it seems that the latter effect dominates. The idea to use a shift-share analysis to investigate differences in occupational structures across regions is an interesting one, but not immediately transferable to this paper. The reason for this, is that the main focus of this paper lies on within-sector differences, i.e., differences in occupational (and educational) structures across regions for a given sector. Nevertheless, it would be an interesting idea for future research to apply such a shift-share analysis to occupational and educational structures in the Netherlands.

3 Data

All data on people's occupation and education used in this research is from the Dutch Labour Force Survey (EBB), and has been obtained for the period 1996-2017. The EBB is an annual survey conducted by Statistics Netherlands of one percent of the Dutch working population, and is the only source that registers occupation and education in the Netherlands. The EBB contains respondents' highest obtained education.

To understand how the data which has been obtained is constructed, note that ROA makes use of a certain occupational and a certain educational classification to make their labour market predictions. These are also the classifications used to construct the data which is used in this paper.[1] More specifically, for the occupations the Beroepenindeling ROA CBS (2014) is used (short BRC2014). This classification is largely based on the International Standard Classification of Occupations (ISCO 2008) and it distinguishes three different levels of aggregation. From a high to low level of aggregation these are Beroepsklasse, Beroepssegment and Beroepsgroep. These levels of aggregation will be referred to respectively as Occupational class, Occupational segment and Occupational group. The classification distinguishes 13 occupational classes, 41 occupational segments and 115 occupational groups. Furthermore, for the educational classification the Opleidingsclassificatie naar Niveau en Richting (2019) is used (short ONR2019). This classification makes a distinction between different types of education based on the level and field of the education. It distinguishes four different levels of aggregation, from high to low being ONRniveau, ONRsector, ONRsubsector and ONRtype. These contain respectively 9, 33, 52 and 137 different types of education. Both the occupational and educational classifications can be found in appendix A.[2]

The EBB data contains for all different occupational and educational aggregation levels (except for the ONRniveau level) the number of workers by occupation (or education), employment sector, labour market region and year. The data also includes this information separately for the country as a whole. It is important to note that the labour market region is based on the residential location of the workers, not the working region of the workers. The EBB only contains the residential location.[3] The disadvantage of

1. More detailed information on these classifications can be found through the website of ROA (although exclusively available in Dutch): http://roa.sbe.maastrichtuniversity.nl/roanew/wp-content/uploads/2015/06/ROA_TR_2015_5.pdf (occupational classification) and http://roa.sbe.maastrichtuniversity.nl/roanew/wp-content/uploads/2019/05/ROA_TR_2019_1.pdf (educational classification).
2. These classifications are defined in Dutch. Unfortunately, no official translation of these classifications exists. As a result, it was chosen to exclusively translate those occupations and types of education which appear in the text of this paper. However, to provide the non Dutch speaking reader with at least some information on how these classifications are constructed, Appendix A includes a table with translations of the occupational classes (i.e. a translation of the highest occupational aggregation level). Furthermore, the next section of this paper provides some information on how the Dutch educational system is constructed, and how this relates to the educational classification that is used.
3. Datasets with work location are available, however these are less reliable and not available for all workers.

the use of residential location as a proxy for the labour market region is that residential locations are probably less specialized as many people with many different educational and occupational backgrounds live together in residential areas while firms in comparable fields and sectors tend to cluster more. This would mean that using residential locations as a proxy for the labour market region probably underestimates the regional differences in sectoral occupational and educational composition.

Furthermore, if no people where observed for a certain combination of all of these variables, that combination has been left out of the dataset. The 22 different employment sectors used by Statistics Netherlands can be found in Appendix B, together with an English translation. It is chosen to drop those people from the analysis for which the sector they work in is labeled as `unknown'. This leaves us with 21 employment sectors in total. Furthermore, for the regional classification the Dutch Arbeidsmarktregio (amr) classification is used. This is the classification is not only used by ROA, but also by UWV, and splits the Netherlands into 35 labour market regions. A map displaying these labour market regions can be found in Appendix C.

A very important aspect of the dataset used for this study, is that it contains many missing observations. This is a result of Statistics Netherlands' confidentiality policy. If for a certain year, a certain combination of occupation/education, sector and region was observed less than ten times, this cell has been left blank in the dataset. The reason behind this, is that with such small numbers it might be possible to trace back certain observations to specific individuals. Ultimately, it appears that this is not so much of a problem as it seems at first sight because those types of occupation and education which account for a relatively large fraction of the total working population in a certain sector and region, are also more likely to be observed more than nine times. Therefore, we still observe the majority of the working population in the Netherlands. Moreover, one also wonders how reliable observations smaller than ten are in the first place. Such observations are prone to a lot of sampling variability, and therefore would not be very valuable to work with. Still, to mitigate the number of missing cells, it was chosen to pair observations per two years. This increases the fraction of cells with values larger or equal to ten. For the period 1996-2017 this then results in eleven different periods, which will be denoted by the first year the respective period contains. So if for example the period 1996 is mentioned, what is actually meant is the years 1996 and 1997 together.

To quantify the extent to which the data is incomplete, the concept of *coverage* is introduced. Since the main focus of this paper is on regional sectoral occupational and educational structures of the working population, what is important, is for which proportion of the working population in a certain region and sector in a given period the occupation or education is observed, depending on whether we are examining occupational or educational structures. For region r, sector s and period t, let us denote this proportion by $coverage_{rst}$. Obviously, to calculate these proportions correctly from the data as discussed would be impossible, as the partials don't sum up to the totals because of missing values. Therefore, extra data has been obtained from Statistics Netherlands, including the total number of people by region, sector and period (so not also by occupation or education). Note furthermore that this extra data also contains the number of people working in each sector nationally. Of course, this supplementary data is also subject to the rule that observations smaller than ten result in a missing values. However, since it is based on higher aggregates, there are only six missing values on the regional level (out of a total of 8470 cells: 35 regions * 22 sectors * 11 periods), and no missing values on the national level. This data can therefore be used to calculate the coverage percentages correctly for almost all combinations of region, sector and period.

For the aggregation level of occupational classes, the average coverage calculated over all different combinations of region, sector and period is well above 90%. This means that the data on this aggregation level covers almost the entire Dutch working population in this period. The average coverage calculated for the occupational segments is slightly higher than 80%, so these data are still fairly representative of the whole. However, it is found that for occupational types the average coverage is well below 80%, and the data on this level thus seem inappropriate to work with. It then remains to be decided whether we want to work on the level of occupational classes or occupational segments. Ultimately it was chosen to work on the level of occupational segments, as the occupational class aggregation level is deemed too generic to provide useful insights when working by sector. The agricultural sector will be dominated by agricultural occupations, the public governance sector will be dominated by public governance occupations, the ict sector will be dominated by ict occupations, etc. Furthermore, for the educational aggregations, the ONRsector aggregation level has a coverage slightly above 80%. Hence, these data are fairly representative of the whole. However, the coverage drops rapidly if one considers the two lower educational aggregation levels. Therefore, with respect to educational structures this paper works at the ONRsector level.

4 Background

This section provides some additional context for the analysis in the remainder of this paper. Firstly, some background information will be provided on the Dutch educational system. The ONR2019 educational classification that is used in this paper is of course a by-product of the way the Dutch educational system is designed, and therefore it is informative to understand how the Dutch educational system works. Secondly, we look at the sectoral educational and occupational composition as some sectors are rather homogeneous and others are more heterogeneous. Then, some general information about the Dutch economy will be provided, like e.g. its employment structure and degree of specialization, and we briefly discuss the regional classification that is used in this research.

4.1 The Dutch Educational System

Just like in most other western countries, after primary education (`basisonderwijs' in Dutch) students continue with secondary education (`voortgezet onderwijs' in Dutch). The secondary education is split roughly into four different levels, from low to high being vmbo-b/k, vmbo-g/t, havo and vwo. After secondary education students move into tertiary education. Tertiary education in the Netherlands can roughly be divided into three different levels. From low to high, these level are vocational training, bachelor education and master education. Note that vocational training in the Netherlands is called mbo, and is itself divided into four different levels. These are from low to high mbo1 up to mbo4. The different educational levels outlined above are roughly the distinctions made by ONR2019 with regard to the educational level. However, this classification not only distinguishes different levels of study, but also different fields of study. From secondary education onwards, students in the Netherlands have to choose a field of specialisation at each level of education. One can roughly distinguish four different fields of study: economics and society; agriculture and nature; technical and ict; and healthcare, education and services. Except on the level of primary education, vmbo-g/t, havo and vwo education, the ONR2019 classification distinguishes for each level of study also these different fields. Note that the different levels and fields of study outlined above are more or less the starting point for the ONRsector aggregation level of ONR2019. The other two observed lower aggregation levels are even more detailed than this, but obviously build on the same structure.

Several remaining remarks should still be made. Firstly, note that the Dutch educational system is rather loose in the sense that the rules allow people to move easily from lower to higher educational levels and vice versa. Secondly, note that children in the Netherlands are obliged to attend education until they are 16, and until they are 18 if they do not yet have obtained a certain minimum degree. Thirdly, in the context of this paper it is interesting to note that the average educational level of the Dutch population has increased rapidly over the past years. This increase seems to be equally spread among all types of education. As an example, whereas in 2003 29 percent of people with a business economical or commercial occupation had obtained a bachelor diploma, this number increased to 44 percent in 2019 (Statistics Netherlands, 2019b).

4.2 Sectoral classification

In this research, we use an aggregate sector classification of 21 sectors based on the SBI-2008 sector classification. Table 1 gives an overview of the 21 sectors used in this research and by ROA for the labour market forecasts. The second and third column in Table 1 gives the total share of the three largest occupational groups and educational types in the sector. This gives an indication of how homogeneous a sector is in terms of occupational and educational composition. We show this because some sectors are much more homogeneous by construction than other sectors, which may affect the results on regional specialisation we will find in this research. For example, in the Catering sector, 60% of the workers has one of three occupations (waiters/bartenders, cooks, kitchen helpers), while in Specialist business services 18% of the workers works in one of the three largest occupations. In the latter case, this means that the Specialist business services sector contains many small occupations that makes the sector very broad or diverse. The homogeneity of a sector can influence the regional specialization both ways. If a sector is rather homogeneous by nature, we do not expect to find many regional deviations of the occupational structure of a sector, while we can find the opposite for sectors that are very heterogeneous by nature. If a sector is very heterogeneous by construction of the classification, like for example the Specialist business services, part of the regional occupational specialisation structures may be due to the sector composition.

For the educational sectoral composition, we find a bit less deviations between de shares of the three largest educational types as is shown in Table 1. However, we see roughly the same pattern as with occupational shares. Sectors like Public governance, Specialist business services, Wholesale, Energy, and the Chemical industry have a rather heterogeneous educational structure.

Table 1: Sectoral employment and occupational composition 2017-2018.

Sector	Share of 3 largest occupational groups in sector	Share of 3 largest educational types in sector
Agriculture	54.7%	29.9%
Food and stimulants industry	27.0%	23.9%
Chemical industry	25.2%	16.2%
Metal industry	23.1%	21.6%
Other industry	18.5%	26.5%
Energy	18.1%	16.4%
Construction	31.7%	28.8%
Retail	52.3%	39.8%
Wholesale	21.2%	20.5%
Transport and storage	38.9%	26.3%
Catering	59.7%	46.4%
Information and communication	43.2%	27.2%
Financial services	28.0%	25.6%
Specialist business services	17.7%	20.3%
Rental and other business services	28.7%	27.9%
Public governance	23.5%	19.7%
Education	48.0%	31.8%
Healthcare	37.6%	27.9%
Well-being	55.8%	32.4%
Culture, sports and recreation	29.9%	31.7%
Other services	42.4%	25.3%

Source: CBS/EBB calculations by ROA

4.3 The Dutch Economy and Labour Market

The Netherlands is a relatively small country with around seventeen million inhabitants. Its Gross Domestic Product (GDP) amounted to approximately 737 billion euros in 2017. The last measurement of its GDP per capita stems from 2016 and yielded a value of 41258 euro per capita (Statistics Netherlands, 2019a). This places the country among the wealthiest countries worldwide.

To get an idea of how the Dutch economy looks like, consider the overview of sectoral employment shares in Table 2a. This table is constructed using the Statistics Netherlands data, and is based on employment counts for the period 2016. Similar to most other developed economies, it can be seen that service sectors account for a large part of total employment. The majority of this employment is located in and around large cities like Amsterdam, Utrecht and Rotterdam. Furthermore, it appears that transport and storage is relatively important in the Netherlands as compared to some other countries. This is mainly because the Netherlands has an important transit function as a result of its favourable geographical location. Through the Rotterdam harbour and the Amsterdam airport Schiphol, the Netherlands connects mainland Europe to the rest of the world. A last notable observation from this table is that industrial sectors like the metal industry, the chemical industry and the energy industry only account for a relatively small share of employment in the Netherlands. However, this is not unusual for a developed economy like the Dutch one.

Table 2a: Sectoral employment shares for 2016, ordered high to low.

Sector	**Employment Share**
Retail	10.5%
Well-being	9.0%
Specialist business services	7.1%
Healthcare	7.1%
Education	6.8%
Public governance	6.0%
Construction	5.1%
Rental and other business services	5.1%
Wholesale	4.7%
Transport and storage	4.6%
Catering	4.2%
Financial services and real estate	4.0%
Metal industry	3.4%
Information and communication	3.3%
Other industry	3.1%
Other services	2.4%
Agriculture	2.1%
Culture, sports and recreation	2.1%
Food and stimulants industry	1.7%
Chemical industry	1.6%
Energy	0.9%
Unknown	5.3%

Source: own calculations based on EBB (Statistics Netherlands).

Table 2b: Sectoral specialisation index for 2016, ordered high to low.

Sector	Specialisation index
Energy	0.76
Agriculture	0.72
Chemical industry	0.64
Metal industry	0.63
Food and stimulants industry	0.52
Culture, sports and recreation	0.50
Information and communication	0.44
Other industry	0.31
Construction	0.29
Public governance	0.27
Specialist business services	0.27
Transport and storage	0.24
Healthcare	0.23
Wholesale	0.23
Rental and other business services	0.21
Other services	0.20
Catering	0.19
Well-being	0.17
Education	0.17
Retail	0.12
Financial services	*No data*

Source: own calculations based on index retrieved from the ROA AIS: https: // roastatistics. maastrichtuniversity. nl/ .

As the main interest of this paper lies in identifying regional heterogeneity, it is also worth wile to consider some regional aspects of the Dutch labour market. More specifically, Table 2b provides an index which indicates for each sector under consideration to what extent its employment was geographically concentrated across labour market regions in 2016. This index is based on a regional specialisation index which has been obtained from ROA's online labour market information system (AIS). This regional specialisation index is defined for each combination of region and sector, and is simply equal to the regional employment share of a sector divided by the national employment share of that sector. As a result, values of the index below one indicate that in a certain region a certain sector is less important than it is nationally, and vice versa for values above one. It is then a simple step to transform this regional index into a national index for each sector. For a certain sector, one can simply calculate the standard deviation of the regional index over all 35 labour market regions. Obviously, the higher the resulting number, the more regionally concentrated a certain sector is. The resulting indices are displayed in Table 2b. Note that in ROA's information system unfortunately no data was available for the financial sector. Therefore, the index for this sector could not be calculated. From the table it can be seen that the industrial sectors are relatively strongly clustered in the Netherlands. This makes sense, as the industry generally benefits significantly from scale economies. Furthermore, also the agricultural sector seems to be highly clustered. For the moment, the results displayed in this table will not be discussed in more detail. However, this table will be useful later on this paper, as it helps us to understand which parts of the data are missing.

As stated in the Data section (Section 3), in this research we use the residential location of workers, not the work location. This is driven by data availability. Administrative data from CBS observes the residential location of workers. The work location, however, is more difficult to obtain as we observe at which firm or company someone works, but if a firm or company has multiple constituencies or locations, we do not know at which location a worker works. For the regional labour market forecasts, ROA therefore uses the residential location of workers instead of the work location.

Appendix C gives an overview of the labour market regions used in this research and in the labour market forecasts. The labour market regions are based on the implementation of regional labour market policies for employers by the unemployment agency UWV. Especially in a small country like The Netherlands with a very efficient infrastructure both for public transport as well as the road network, people can easily commute and live and work several labour market regions apart.[4] There are several consequences of this commuting pattern combined with the fact we use residential location. One is that we assume that we observe less sectoral occupational or educational specialisation based on residential locations than we would have observed based on working location. This is because firms tend to cluster more at the same location or city than residents do.

A second implication is that we will sometimes observe specialisation in a location that has a more residential economic function than a productive (work) economic function. This will be the case in regions with a clustering of people with the same educational or occupational background (this would be very pronounced in regions that are known to be very expensive residential locations like region 12 (Gooi- en Vechtstreek) and region 34 (Zuid-Holland Centraal). Generally, the larger cities like Amsterdam, Rotterdam, Utrecht and the Hague have both a strong productive as well as residential function. Many surrounding areas have a more residential function, while the areas outside of the Randstad generally have a construction with a dominant city and its depending area. It is good to bare in mind the construction of regional labour market areas and the fact we use residential locations when interpreting the results presented at the regional level.

4. See, for example, De Groot (2015). Arbeids- en Woningmarktdynamiek, Platform31 Essay 6, and Verkade and Bakens (2020), Commuter Flow Predictions in POA: Evaluation Study, ROA Technical Report 005.

5 Methodology

To be able to answer the two main research questions of this paper, it should be defined how heterogeneity in regional sectoral occupational and educational structures at a given point in time can be measured. This will be the main topic of the first subsection. Secondly, once we are able to measure regional heterogeneity at a given point in time, we also want to know how this heterogeneity has developed over time. The methodology needed to do so, is developed in the second subsection.

5.1 Measuring Heterogeneity in Regional Sectoral Occupational and Educational Structures

Evaluation of Different Indices

To be able to answer the main research questions of this paper, ideally we would have a single summary measure for each sector and period, capturing the extent to which regional occupational and educational structures are different across regions for this respective sector and period. In this way sector specific heterogeneity conclusions can be drawn. The papers which have been discussed in the literature section have given us some idea of what such a measure could look like. An important thing to realize in this regard, is that in comparing occupational and educational distributions, one is not simply measuring one-dimensional heterogeneity. This is the case because a certain region's sectoral occupational or educational structure cannot be summarized with a single number, as such structures consist of multiple different occupations and types of education. As a consequence, standard one-dimensional measures of inequality are inadequate for this analysis. This excludes for example the use of the widely applied Gini-index, but also the use of a simple standard deviation.

The answer to measuring regional sectoral heterogeneity lies in the literature on economic specialisation. This literature is concerned with measuring absolute- and relative specialisation of, for example, production structures, employment structures, and export structures. It is the relative specialisation measures this paper is interested in, as these can be transferred neatly to the context of the current paper. A measure of relative specialisation is meant to quantify to which extent a certain region's (country's) structure for some kind of multidimensional object deviates from a baseline structure for that object, the object being for example a region's (country's) employment structure. It should be noted that all elements of such an object should sum up to 1, as is usually the case for economic structures. In the context of this paper, the object would be either the occupational or educational structure for a specific combination of region, sector and period, and the baseline region's structure would be the occupational or educational structure for all regions in that sector and that period. The baseline region's structure could then be chosen to be either the averaged structure over all regions or directly the country structure. For this research it was chosen to use the latter one, as taking the regional average structure is not very representative in many cases. This is because there are quite some regional observations missing in the data. As an example, consider comparing the fraction of managers in the metal industry in region A with the fraction of managers in the metal industry in the rest of the country. Also suppose that the fraction of managers in the metal industry is only observed for 12 out of 35 regions. It is obvious that taking the average fraction over these 12 regions is not very representative of the fraction of managers in the metal industry in all regions, and this would thus not be an appropriate reference for region A's share. On the contrary, using the national proportion of managers in the metal industry will be much more representative of the state of all regions together.

In conclusion, a measure of relative specialisation can provide us with a summary measure of heterogeneity for every combination of region, sector and period. Palan (2010) discusses the appropriateness of a set of commonly used absolute and relative specialisation measures based on their axiomatic properties. The axiomatic properties under consideration are anonymity, progressive transfers, bounds, decomposability, insensitivity to splitting/merging groups and independence of irrelevant alternatives. These properties have closely connected to the extensive body of literature on income inequality. For more details about these properties, see Palan (2010). Palan (2010) concludes that with regard to measures of relative specialisation, it is best to either use the Krugman Specialisation Index (KSI) (Krugman, 1993) or the Theil Index (Theil, 1967), depending on the specific context under consideration. One important property of the KSI which the Theil Index does not possess, is the property of insensitivity to changes in occupational or educational classifications. This implies that for a certain combination of region, sector and period, the value of the KSI will not change if a certain occupational group is assigned to a different occupational segment (a similar definition applies of course to other aggregation levels and to measuring the KSI for educational structures). This property is mainly important if one examines heterogeneity over time, as classifications often vary over time. As this research concerns heterogeneity over time, it is better to use the KSI. However, perhaps a more important reason to choose the KSI, is its appealing intuitive interpretation. The main point is that on the basis of the Theil Index one can merely conclude whether there is much or little heterogeneity, but one cannot attach any further meaning to this index. This is mainly a result of taking the natural logarithm of employment shares in calculating this index. The KSI is calculated in a much simpler way, making the interpretation of the index more appealing.[5]

Krugman Specialisation Index

For ease of understanding, the construction of the KSI will only be discussed for measuring regional heterogeneity in sectoral occupational structures. However, this framework can neatly be transferred to the measurement of heterogeneity in educational structures. The index is outlined for an arbitrary choice of occupational/educational classification.

5. For examples of how the KSI is used in the context of measuring economic specialisation, see e.g. Midelfart-Knarvik and Overman (2002), Marelli (2007), Höhenberger and Schmiedeberg (2008),Belke and Heine (2006), Combes and Overman (2004) and Suedekum (2006).

For a certain region r, sector s and period t, we define the total number of observed occupations (i.e. those occupations which are observed at least ten times), by K_{rst}. We define the proportion of people with a certain occupation k in region r in sector s in period t out of the total number of people working in region r, sector s and period t, by X_{krst}. Furthermore, the proportion of people with a certain occupation k in sector s in period t relative to the total number of people working in sector s in period t nationally is defined as Q_{kst}. Then, for each combination of occupation, region, sector and period, one can calculate how much the regional proportion differs from the national proportion. The absolute value of this percentage point deviation is defined as $Z_{kst} = abs(X_{krst} - Q_{kst})$. To get a summary measure for the deviation of a region's sectoral occupational structure from the national sectoral structure, it then makes sense to sum the deviations over all observed occupations to obtain the total deviation.[6] This is what is called the KSI:

$$KSI_{rst} \equiv \sum_{k=1}^{K_{rst}} Z_{krst} = \sum_{k=1}^{K_{rst}} abs(X_{krst} - Q_{kst}) \qquad (1)$$

This index thus measures for a certain combination of region, sector and period, the total percentage point deviation of the region's occupational composition from the national occupational composition. It is clear that the higher this index, the more heterogeneity there is between that region and the rest of the country. The intuitive interpretation of a total percentage point deviation is a big advantage of using the KSI. This intuition can even be extended further. Consider dividing the KSI by two. The resulting number is the total fraction of the workforce (in a certain region, sector and period) one would have to reallocate to a new occupation to obtain the national composition in that sector in that period (Dixon et al., 2009). In the literature, this number is often called the Coefficient of Regional Specialization (CRS) (see e.g. Hoover (1948)). Lastly, Krugman (1993) has shown that the KSI is bounded between 0 and $\frac{2(K-1)}{K}$, where K denotes the total number of different occupations. This implies that the upper bound for the CRS is equal to $\frac{K-1}{K}$. Since K_{rst} is rarely equal to K in the dataset at hand, the theoretical upper-bounds for the KSI and CRS are in fact even lower for almost all combinations of region, sector and period.

The KSI thus provides us with a way to measure heterogeneity in the sectoral occupational structure for a single region in a given period. However, as indicated before, it is preferred to have a single measure of heterogeneity for every combination of sector and period. Such a measure should tell us how much heterogeneity there is across the regions on average. To be able to obtain such a measure, it makes sense to simply average the KSI over the regions for each combination of sector and period. This regional average can be defined as:

$$\overline{KSI_{.st}} \equiv \frac{\sum_{r=1}^{35} KSI_{rst}}{35} \qquad (2)$$

6. This then differs from the specialisation computed in ROA's AIS, and depicted in Tables 2a and 2b.

The regional average of the KSI provides us with a basic measure for the extent of heterogeneity in regional sectoral occupational and educational structures. However, as there are quite some missing observations in the data, the index is slightly flawed. It is not calculated over all different occupations and types of education, but only over the observed ones. The calculated KSI_{rst} are thus in fact lower bounds for the actual unobserved total percentage point deviation KSI^*_{rst}, and the $\overline{KSI_{.st}}$ is thus a lower bound for $\overline{KSI^*_{.st}}$. How close these lower bounds are to the actual values, depends of course on the extent to which data is missing. To get an idea of how representative each $\overline{KSI_{.st}}$ is of $\overline{KSI^*_{.st}}$, it is best to simultaneously examine the regional average coverage (as defined in Section 2) for that sector and period $\overline{coverage_{.st}}$. Therefore, when examining the KSI values, the coverage will also be reported.

In general, quantifying the extent of heterogeneity in regional sectoral occupational and educational structures for a specific combination of sector and period by first quantifying heterogeneity for a given region, sector and period combination and then averaging this over the regions is just one possibility. Alternatively, one could also choose to first obtain a heterogeneity measure for each combination of occupation/education, sector and period, and subsequently averaging this number over all occupations/types of education within a sector for a given period. For each combination of occupation/education, sector and period, one could e.g. calculate such a specialisation index in a similar way as the indices reported in Table 2. Subsequently, the resulting index could then simply be averaged over all occupations/types of education, resulting in an alternative heterogeneity index for each combination of sector and period. The main reason why in this research the KSI approach is preferred over averaging ROA-like specialisation indices, is that indices resulting from the latter approach do not possess a similar convenient interpretation. The average standard deviation of the ratio of the regional share relative to the national share is not as straightforward to interpret as the average total percentage point deviation. Even more importantly, averaging ROA-like specialization indices causes us to lose the convenient interpretation of a `lower-bound' on heterogeneity. With a complete dataset, the resulting indices could become lower or higher, we simply do not know.

The potential downside of solely using the $\overline{KSI_{.st}}$ to quantify regional heterogeneity, is that these values tell us nothing about how large the deviations are for individual occupations or types of education. To see why this matters, consider the following example. Suppose we observe for the occupational structure in a certain sector in region A a K_{rst} equal to 20 percentage points. Now suppose in region B (same sector) we also observe a K_{rst} equal to 20 percentage points. Also suppose both region/sector combinations have a similar $coverage_{rst}$, which is well above 80%. Purely based on the KSI, one would conclude that both regions are more or less heterogeneous to the same extent. However, these values of 20 can be due to entirely different deviation patterns. It can for example be due to ten observed occupations each having an absolute deviation equal to two percentage points, but it can also be due to only two observed occupations each having an absolute deviation of ten percentage points. If one wants to measure heterogeneity as objectively as possible, these two deviation patterns should be regarded more or less equally (as would be the result from purely focusing on the K_{rst}). However, if one makes labour market forecasts based on these two different structures (as ROA would do), assuming that the national occupational structure in a sector reflects the regional structure in that sector may lead to more harmful forecasting biases when the average observed deviation is larger. In the example given above, assuming the national structure for the regional structure with ten occupations deviating two percentage when forecasting the demand for these occupations will yield forecasts which are slightly off for all occupations. Such small discrepancies are probably not so difficult to adjust for by economic decision makers basing their actions on these forecasts. However, assuming the national structure for the regional structure with two occupations deviating ten percentage points when forecasting the demand for these occupations will yield forecasts which are much less accurate. For economic decision makers basing their decisions on this information, two such largely inaccurate forecasts are probably much more disruptive than ten forecasts which are all slightly wrong. They key message underlying this example is that for economic decision makers it does matter how large the average observed deviations are. Therefore, in quantifying regional sectoral heterogeneity, it is informative to also consider this aspect. More specifically, it was chosen to also examine the how much occupations/types of education deviate on average (in absolute value). For a certain combination of region, sector and period, this average deviation is defined as:

$$AD_{rst} \equiv \frac{\sum_{k=1}^{K_{rst}} z_{krst}}{K_{rst}} = \frac{KSI_{rst}}{K_{rst}} \tag{3}$$

So for a certain combination of region, sector, and period, this measure indicates how much on average the observed occupation/education shares differ from their corresponding national shares. Just like for the KSI, it makes sense to average this index over the regions, as it is convenient to have a single measure for the entire country for each sector and period. One then obtains the `average average deviation'. For a certain combination of sector and period this index is defined as:

$$\overline{AD_{.st}} = \frac{\sum_{r=1}^{35} AD_{rst}}{35} \tag{4}$$

This index will give us some idea of the magnitude of the deviations which drive the observed $\overline{KSI_{.st}}$. Be sure not to interpret this $\overline{AD_{.st}}$ as a proxy for the actual unobserved $\overline{AD^*_{.st}}$. For most combinations of region, sector and period, around ten to fifteen different occupations and types of education are observed (out of a total of 41 different occupations and 31 different types of education). The unobserved shares are by construction relatively small, and are therefore mostly shares with a deviation from the national share close to zero. Think e.g. of doctors in the agricultural sector or farmers in the health care sector. As a result, the actual unobserved $\overline{AD^*_{.st}}$ is likely to be much lower than the $\overline{AD_{.st}}$, and will in fact be close to zero. Even if the full dataset would be observed, it would be much more interesting to examine $\overline{AD_{.st}}$ rather than $\overline{AD^*_{.st}}$, as the former captures how large on average the largest deviations are. In other words, $\overline{AD_{.st}}$ captures the extremes of the distribution of deviations. The example in the previous paragraph has illustrated that more extreme deviation patterns are potentially more harmful to ROA's models, even if they are on average similar to other patterns.

Therefore, in what follows, the $\overline{AD_{.st}}$ should be interpreted as approximate average deviations at the high end of the distributions.

Measuring Current Heterogeneity in Regional Sectoral Occupational and Educational Structures

Now that a framework has been developed for measuring regional sectoral heterogeneity in occupational and educational structures at a given point in time, let's see how this can be used to answer the main research questions of this paper. Firstly, it is of interest how much regional heterogeneity in sectoral structures there is currently. To examine this, it would make sense to examine the three measures defined in Section 5.1.2 for the last available period, i.e.$t = 2016$. However, in the end it was chosen to report for each of the measures the average of its value over $t = 2014$. and $t = 2016$. The reason to do so, is to eliminate some of the sampling variability inherent to the survey data. A careful examination of the data has pointed out that such variability is indeed present. Also, it is not expected that occupational and educational structures of the labour force shift severely within three or four years, given the fairly rigid nature of such economic structures. Hence, averaging over these two periods (i.e. four years) should still yield a representative image of current heterogeneity. As indicated before, the indices for the occupational structures will be calculated at the occupational segment level and for the educational structures at the ONRsector level. Still, if one works at these aggregation levels, there are certain sectors which have an average $\overline{coverage_{.st}}$ over the periods 2014 and 2016 which is well below 80%. This can be seen from Table 3, which displays for all sectors the average coverage and standard deviation of the coverage over the regions, averaged over 2014 and 2016. For both occupational structures and educational structures, it concerns the sectors agriculture, food and stimulants industry, chemical industry, other industry, energy, culture, sports and recreation, and other services. There seem to be two plausible explanations why these seven sector have a substantially lower average coverage than the other sectors. Firstly, as can be seen from Table 1, these are the seven sectors with the smallest employment shares. Obviously, the regional structures of these sectors are therefore more likely to be less fully observed, as the chance of a certain occupation/education having less than ten observations is much higher. However, this does not seem to be the main reason why these sectors have a lower average coverage. The employment shares are indeed lower, but not drastically so. What seems to play a more important role, is the extent to which these sectors are regionally clustered. Table 2 shows that, expect for the other services sector, these sectors are among the top most clustered sectors. If a sector is strongly regionally clustered, this implies that there are certain regions with little to no activity in this sector in terms of employment. Hence, for those regions the chance to observe incomplete occupational and educational structures is much higher, as the chances of having less than ten people with a certain occupation or education are much higher. This argument is strengthened by observing in Table 3 the high standard deviation of the $\overline{coverage_{rst}}$ across regions for these sectors relative to the standard deviation for the other sectors. This indicates that there are certain regions with very low coverage and certain regions with very high coverage.

The observation that the sectors with a relatively low coverage are strongly clustered sectors, indicates that it is probably possible to exclude those region/sector combinations with a relatively low coverage without losing too much of the Dutch working population. Moreover, in light of constructing labour market forecasts, it is obviously more important to be properly informed about region/sector combinations with a larger workforce as compared to those combinations with a relatively small workforce. The idea to solve the `coverage issue' by excluding those region/sector combinations with a small coverage, thus seems to be reasonable. In the worst case, the calculated indices for sectors for which regions are dropped will merely be proxies for what is happening in the country as a whole.

Ultimately, it was chosen for these seven sectors for both occupational and educational structure to omit those regions for which the averaged coverage over 2014 and 2016 is lower than 80%. Appendix D provides an overview of the regions which are included for these sectors. Note that this criterion causes the energy sector to disappear from the analysis all together. Even though potentially some relevant information is lost by eliminating a part of the sample, the problem of having a low coverage was deemed to be more important. Moreover, as indicated, the omission of these sector/region combinations is probably less severe than it looks at first sight, since these concern mostly highly clustered sectors. To illustrate this, consider the fact that the sector/region combinations which have been dropped only accounted for around 8% of the Dutch working population in the period 2016.

There is another rationale to exclude combinations of sector and region with a low coverage. Naturally, those combinations with a relatively low coverage will be combinations for which even for the proportion wise more important occupations and types of education, relatively little people are observed for a given period. This implies that for such occupations or types of education small changes in the number of observed people can lead to relatively large changes in the estimated corresponding regional sectoral share. In statistical terms, the distribution of such shares has a large standard deviation. As a simple test of this statement, consider the fact that the correlation between AD_{rst} and $coverage_{rst}$ is equal to -0.5 for the occupations and equal to -0.6 for educations. This implies that, in general, the smaller the coverage for a certain combination of region, sector and period, the higher that region's observed average deviation from the national sectoral structure is in that same period. There is no reason to believe that region/sector/period combinations with a relatively low coverage systematically deviate more from the corresponding national sectoral structure than region/sector/period combinations with a relatively large coverage. Just because relatively little people work in a certain combination of sector and region, doesn't mean its occupational or educational composition should differ relatively more. The sampling variation argument seems to be the only viable explanation for this observation. This then leads us to conclude that the observed occupational and educational structure for region/sector combinations with a low coverage are relatively less representative of the true underlying structures than is the case for the combinations with a higher coverage.

As a result, one can argue that it is better to eliminate all sector/region combinations with a relatively low coverage, and not just those for which the average coverage across the region is too low (those which have been discussed earlier). Indeed, when initially calculating the indices for each sector, it appeared that those sectors which were strongly clustered displayed the most heterogeneity. In light of the previous arguments, these results make one doubt their reliability. However, in solving this issue and thus potentially omitting region/sector combinations with a low coverage, we would also not want to unnecessarily throw away relevant information. Therefore, as a middle ground, it was chosen to also omit regions with an average coverage over $t = 2014$ and $t = 2016$ lower than 80% for the metal industry sector and the information and communication sector. Both these sectors are among the highest clustered ones, and both have a relatively high standard deviation of the coverage (for both occupational and education structures) as can be seen from Table 3. In combination with an average coverage close to 80% for these sectors, these observations imply that for these sectors there are many regions with an average coverage lower than 80%. The included regions for these two sectors are also displayed in Appendix D. Note that this argument also applies to a lesser extent to the rental and other business services sector. However, Table 1 indicates that this sector account for a rather substantial employment share. Therefore, omitting regions for this sector will likely result in dropping a large part of the Dutch working population. As a result, it was chosen to not apply a similar selection criterion to this sector.

Table 3: Sectoral average coverage and standard deviation over the regions (x100), averaged over 2014 and 2016. Standard deviation in percentage points.

	Coverage (Occupation)		**Coverage (Education)**	
Sector	*mean*	*sdev*	*mean*	*sdev*
Agriculture	75%	16	65%	28
Food and stimulants industry	66%	21	57%	21
Chemical industry	59%	23	57%	24
Metal industry	81%	18	82%	14
Other industry	73%	17	75%	16
Energy	53%	25	43%	25
Construction	88%	5	87%	6
Retail	93%	3	95%	4
Wholesale	86%	8	87%	9
Transport and storage	87%	5	86%	8
Catering	89%	5	83%	9
Information and communication	80%	15	80%	12
Financial services	86%	7	85%	9
Specialist business services	88%	8	90%	8
Rental and other business services	83%	10	84%	14
Public governance	89%	7	84%	9
Education	91%	6	90%	7
Healthcare	93%	4	93%	4
Well-being	96%	2	95%	4
Culture, sports and recreation	66%	18	67%	19
Other services	70%	12	70%	16
Average	81%	11	80%	12

Source: own calculations based on EBB (Statistics Netherlands).

5.2 Measuring the Development of Heterogeneity in Regional Sectoral Occupational and Educational Structures over Time

Besides measuring current regional sectoral heterogeneity in occupational and educational structures, it is also interesting to look at how this heterogeneity has developed over time. A natural question to ask in this regard, is whether regional sectoral occupational and educational distributions are becoming more similar or less similar over time. In other words, are these regional structures converging or diverging over time? To answer this question, in the simplest form one could simply look at how the KSI_{rst} have developed over time. However, merely looking at whether these values have increased or decreased is not necessarily very interesting. It is more interesting to look at whether there are patterns in convergence or divergence. Moreover, it is also likely that such an approach would lead to erroneous conclusions, as a result of the incomplete dataset. A combination of region, sector and period which has a low coverage is namely also more likely to have a low KSI. This would not be a huge problem if the coverage in the dataset would remain constant over time. However, in general the data show that the coverage increases over time. This implies that KSI_{rst} values are generally increasing in the data, and would therefore most likely lead us to conclude that heterogeneity has steadily been increasing over time. Therefore, to prevent the results from being biased by developments in the coverage, one needs to control in some way for the increasing coverage. This naturally leads us to consider a regression model in which we can control for the coverage. More specifically, to investigate convergence patterns in the development of heterogeneity over time, the following model (model 1) can be estimated:[7]

$$KSI_{rst} = \alpha + \gamma t + \delta coverage_{rst} + \epsilon_{rst}, \qquad (5)$$

where α is an intercept and t is a deterministic time trend. If the resulting estimate for γ is significantly positive it can be concluded that regional heterogeneity has steadily increased, and thus that there has been divergence. Similarly, if the estimate for γ is significantly negative, it can be concluded that there has been convergence on average. Note that if the coverage was omitted from the model, this would most likely result in a positive bias on the estimate for γ. Note furthermore that a similar convergence analysis has been performed by Höhenberger and Schmiedeberg (2008), who examine structural convergence of European economies.

An important thing to realize in the context of convergence models, is that the literature generally distinguishes two different types of convergence. These are σ-convergence and β-convergence. σ-Convergence occurs when the dispersion of the distribution of the variable of interest across regions becomes lower over time. β-Convergence occurs when the partial correlation between growth in the variable of interest and the variable's initial level is negative. Well-known papers investigating β -convergence in

7. Note that it was chosen not to allow for fixed effects in MODEL 1, as this would greatly complicate the relationship between MODEL 1 and MODEL 2 (which is introduced shortly), thereby losing the intuitive connection between the two models.

income across countries are by Barro and Sala-i Martin (1992) and Mankiw et al. (1992). Even though most convergence literature focuses on β-convergence, economists have acknowledged that it is not a sufficient condition for σ-convergence, but merely a necessary condition. This is illustrated by Young et al. (2008). In this spirit, both Friedman (1992) and Barro and Sala-i-Martin (1993) argue that it is more interesting to examine σ-convergence, as this speaks directly to what happens to the distribution of interest.

So what kind of convergence does MODEL 1 examine? In principle, one could interpret MODEL 1 as a σ-convergence model, as it simply looks at whether on average the dispersion between regional sectoral occupational and educational structures (as measured by the KSI_{rst}) has systematically increased or decreased over time. It should be said though, that this approach measures trends in σ-convergence only indirectly, as the KSI_{rst} is a summary measure of what happens to the dispersion of the distributions for all types of occupations and education for a certain combination of region, sector and period. Furthermore, note that this paper will limit itself to an examination of trends in σ-convergence, and will thus not also examine trends in β-convergence. Examining trends in σ-convergence should provide us with a basic idea of whether there are any trends in how heterogeneity has developed over time, and this is deemed sufficient for the moment.

In itself, the resulting estimate for γ in MODEL 1 thus gives us a basic answer to the question whether regional sectoral occupational or educational structures have systematically been converging/diverging over time or not. However, this model restricts all possible combinations of region and sector to have the same trend and intercept. It is very unlikely that this in fact represents reality. Therefore, an interesting extension to MODEL 1 is to allow for region and sector specific intercepts and time trends. The model (model 2) to be estimated is then as follows:

$$KSi_{rst} = \alpha^r + \alpha^s + \gamma^r t + \gamma^s t + \delta coverage_{rst} + \epsilon_{rst}, \quad (6)$$

where the superscript r denotes a region specific coefficient, and the superscript s denotes a sector specific coefficient. The interpretation of γ remains largely the same, the only difference being that it is now a region or sector specific coefficient. This is the second main convergence model which will be estimated for both occupational and educational structures. It is now thus possible to draw a more nuanced conclusion with regard to convergence/divergence, as it is possible to do this for each region and sector separately. In this way one can see how individual regions and sectors are driving the estimation results for MODEL 1.

Note that in order to draw such conclusions, it needs to be tested whether the obtained estimates for $\gamma^r$and γ^s are significantly different from zero. To be able to test this, ideally one would estimate a regression model with dummy variables for all sectors, dummy variables for all regions, and interactions of all those variables with a time trend. However, multicollinearity issues arise in such a case. Instead, at least one region dummy or one sector dummy needs to be dropped from the model. If one drops a region dummy, the significance tests on the estimated γ^r are no longer with respect to zero, but with respect to γ^r for the region which is dropped. However, the significance tests for the estimated γ^s will actually be with respect to zero in such a model. Hence, in order to perform all required tests, for both occupational and educational structures MODEL 2 is estimated twice, once by dropping a region dummy and once by dropping a sector dummy.

Lastly, it should be noted that the models presented in this subsection will only be estimated for those combinations of region and sector which have a sufficiently high coverage over the entire period. If the coverage is very low for a certain combination, the development of its KSI is not reliable. For uniformity purposes, it was chosen for both occupational and educational structures to include the same region/sector combinations as will be included for the static analysis (i.e. the combinations displayed in Appendix D). It turns out that, after reducing the sample, each of these nine sectors has an average $coverage_{rst}$ of at least 75% over the entire period. This is deemed sufficient to draw sensible conclusions with regard to convergence/divergence trends in heterogeneity. However, when interpreting the results, one should be aware of the fact that they are based on an incomplete dataset. Also be aware that the sample selection that is performed implies that the energy sector is again dropped all together from the analysis.

6 Results Part 1: Heterogeneity in Occupational Structures

This section provides the results for the occupational structures. The results for the educational structures will be looked at in the next section. The first subsection lays out the current heterogeneity patterns in the Netherlands. The second subsection looks at how the heterogeneity has evolved over time.

6.1 Current Heterogeneity in Regional Sectoral Structures

This subsection aims to give an answer to the question to what extent current regional sectoral occupational structures differ across regions. To this purpose, Table 4 displays for every sector the $\overline{KSI_{.st}}$, the $\overline{AD_{.st}}$ and the $\overline{coverage_{.st}}$, averaged over $t = 2014$ and $t = 2016$. Besides the regional mean of the three main measures, the table also reports the regional standard deviation of each of the three measures in parentheses behind the mean. Note that all sectors for which certain regions have been dropped from the analysis as a result of the selection based on the coverage are marked with a †. This notation will be maintained throughout the entire paper.

Table 4: Regional average and standard deviation of KSI, AD and coverage per sector (x100), averaged over 2014 and 2016 - occupational structures.

Sector	$\overline{\text{KSI}}$		$\overline{\text{AD}}$		$\overline{\text{Coverage}}$	
Agriculture†	22.1	(8.4)	3.9	(1.7)	85%	(4)
Food and stimulants industry†	27.4	(4.9)	2.5	(0.8)	84%	(5)
Chemical industry†	26.0	(7.5)	1.9	(0.7)	86%	(4)
Metal industry†	26.9	(6.6)	1.9	(0.6)	89%	(6)
Other industry†	25.8	(4.3)	1.4	(0.3)	87%	(5)
Construction	18.6	(6.7)	2.0	(1.0)	88%	(5)
Retail	14.0	(4.5)	1.1	(0.6)	93%	(3)
Wholesale	24.9	(5.2)	1.7	(0.6)	86%	(8)
Transport and storage	20.3	(7.4)	3.0	(2.0)	87%	(5)
Catering	15.1	(6.2)	2.8	(1.7)	89%	(5)
Information and communication†	23.3	(7.2)	2.6	(1.2)	85%	(9)
Financial services and real estate	22.9	(6.7)	2.4	(1.2)	86%	(7)
Specialist business services	23.8	(5.2)	1.5	(0.8)	88%	(8)
Rental and other business services	23.9	(6.3)	2.7	(0.8)	83%	(10)
Public governance	23.2	(5.9)	1.7	(1.1)	89%	(7)
Education	13.5	(4.5)	1.4	(0.9)	91%	(6)
Healthcare	20.6	(8.1)	1.9	(1.1)	93%	(4)
Well-being	19.2	(5.9)	1.5	(0.7)	96%	(2)
Culture, sports and recreation†	30.3	(10.3)	2.4	(1.0)	87%	(7)
Other services†	31.8	(6.4)	2.7	(1.0)	86%	(4)
Average of all sectors	22.7	(6.5)	2.2	(1.0)	88%	(6)

Source: own calculations based on EBB (Statistics Netherlands).

Let us first examine the values of $\overline{KSI_{.st}}$, as these provide us with a basic idea of how much heterogeneity there is in occupational structures. Remember that these are in fact lower bounds for the actual unobserved total percentage point deviations $\overline{KSI^*_{.st}}$..

However, as the coverage is on average 88% and nowhere lower than 83%, these $\overline{KSI_{.st}}$ values are likely to be the close to the actual unobserved total percentage point deviations. From the bottom line of the table, it can be seen that the average total percentage point deviation of a region's sectoral occupational structure from the national sectoral occupational structure across all sectors is currently at least around 23 percentage points. Dividing this number by two and thereby calculating the Coefficient of Regional Specialisation (CRS), yields an average CRS of 11.5 percentage points. This means that on average, for a certain combination of sector and region, one would currently have to reallocate at least approximately 11.5% of the workforce to a different occupational segment to obtain the national occupational structure of that sector. To be able to better understand the magnitude of this deviation, consider the fact that on average around 11000 people are working in a certain sector in a certain region. This implies that on average for each combination of region and sector at least roughly 1250 workers would have to be reallocated to a different occupational segment to obtain the national structure. Furthermore, looking at sector specific values of $\overline{KSI_{.st}}$, several conclusions can be drawn. Firstly, it can be seen that all values lie somewhere between 13 and 32 percentage points. There thus seems so be quite some variety in the index values across sectors. More specifically, it can be seen that the sectors retail, catering and education all have relatively low values of the index (approximately 14 percentage points), indicating that these sectors are probably relatively the most similar across regions. These are indeed sectors for which one would not expect too much heterogeneity in occupational structures across regions. At the upper end of the $\overline{KSI_{.st}}$ distribution, one can identify the sectors culture, sports and recreation, and other services to have relatively large values for the index (approximately 31 percentage points). The total percentage point deviation for these sectors is thus approximately twice as large as the deviations for the sectors at the lower end of the distribution. The heterogeneity for the remaining sectors seems to be more or less equal, most index values being close to the average of 22.7 percentage points.

As indicated in the methodology section, it is important to also examine how much the observed occupations deviate on average (in absolute value). This gives us some idea of how large deviations are at the extremes of the distributions. The sectoral regional average of this number is $\overline{AD_{.st}}$, and is also reported in Table 4. Looking at the bottom line of the table, it can be seen that the average of this index across all sectors is equal to 2.2. This means that, on average, the observed occupations' shares deviate 2.2 percentage points from the corresponding national shares. Dividing the average of the $\overline{KSI_{.st}}$ across all sectors by 2.2, it can be seen that a region and sector's total deviation as measured by the $\overline{KSI_{.st}}$ is on average constructed out of around ten different occupations. Furthermore, it can be seen that for most sectors the $\overline{AD_{.st}}$ does not vary much around the cross-sector average. Therefore, it can be concluded that in most sectors the deviations at the extremes of the distribution are equal to around 2.2 percentage points. It can also be concluded from this that, even though some sectors have substantially higher $\overline{KSI_{.st}}$ than others, this seems to be mainly driven by the fact that such sectors have more heterogeneous occupational structures in terms of number of relevant occupations, and not so much by individual occupations deviating more. Lastly, one sector specific value of $\overline{AD_{.st}}$ which is worth pointing out, is the relatively high value for the agricultural sector. This implies that observed regional shares for the agricultural sector tend to deviate relatively much from the national shares. Note that later on in this paper occupation specific developments will be discussed, shedding some more light on this finding for the agricultural sector.

6.2 Heterogeneity in Regional Sectoral Structures over Time

This subsection aims to answer the question whether regional heterogeneity in sectoral occupational structures has steadily decreased or increased over time, i.e. whether there are any convergence or divergence patterns. To this purpose, let us first consider the results from estimating MODEL 1 for the occupational structures using the Pooled Ordinary Least Squares (POLS) estimator. The estimation results are displayed in Table 5. Below the estimated coefficients and their significance, also some regression diagnostics are reported. Here, $\hat{\sigma}^2_\epsilon$ is the estimated variance of the error term ϵ. Note furthermore, that this model is estimated using 5830 observations for KSI_{rst}. To see this, consider the fact that twelve `complete' sectors are included in the estimation (i.e. twelve sectors for which all 35 regions are included) and that there are nine sector for which only a subset of the regions is included (the subset displayed in the first table in Appendix D). This subset contains in total 110 unique sector/region combinations. As a result, the total number of observations is indeed equal to 5830 (calculated as 110 * 11 (periods) + 35 * 12 (sectors) * 11 (periods)).

Table 5: Estimation results MODEL 1 using POLS estimator - occupational structures.

	MODEL 1
Parameter estimates	
Trend	0.0023**
	(0.0002)
Coverage	0.0825**
	(0.0093)
Regression diagnostics	
R-squared	0.065
$\hat{\sigma}_\epsilon$	0.080
N	5830
Number of groups	530

Clustered standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$

Source: own calculations based on EBB (Statistics Netherlands).

Looking at the estimated coefficients in Table 5, it can be seen that a positive and highly significant trend is estimated. The estimate implies that, on average, the KSI_{rst}. is estimated to increase with 0.23 percentage points per year.[8] Hence, these findings

8. Note that the deterministic trend variable is constructed in such a way that the estimated coefficients can be interpreted as yearly effects, even though the data is based on two year periods.

provide evidence in favour of structural σ-divergence of regional sectoral occupational structures over time, be it very slowly. The increase of 0.23 percentage points every year implies an increase in KSI_{rst}. measured heterogeneity of 2.3 percentage points every ten year. This is not a very large number in the context of the average KSI values displayed in Table 4. Another important result from this table, is the highly significant estimate for the coverage. It thus seems that by controlling for the coverage, indeed an omitted variable bias has been prevented. Furthermore, the coverage also has the expected positive coefficient. This implies that generally the higher the coverage is, the higher the value of the will be.

To get a more nuanced view of convergence/divergence patterns in heterogeneity in occupational structures, it is now looked at the estimation results for MODEL 2. This model allows both all regions and all sectors to display different trends at the same time. The results therefore give some idea as to how certain regions or sectors are driving the estimation results for the trend in MODEL 1. Table 6 displays all estimated coefficients for the γ^r and γ^s which are significant at the 5% level. Moreover, also the coefficient for the coverage is reported. These estimation results are obtained using the simple Ordinary Least Squares (OLS) estimator.

One thing that can be noticed immediately from Table 6, is that all reported significant trend coefficients for both regions and sectors have a positive sign. Moreover, their magnitudes are more or less equal, ranging between 0.0021 and 0.0032. On a ten year basis, this range implies a KSI_{rst}. increase ranging between 2.1 and 3.2 percentage points. In the context of the $\overline{KSI_{.st}}$ values found in the previous subsection, this is not a very large increase. The only exception is the agricultural sector, with its coefficient being equal to 0.0093. This amounts to a 9.3 percentage points increase in the index in ten years. Given the $\overline{KSI_{.st}}$ for the agricultural sector of 22.1, a 9.3 percentage point increase in this index is quite substantial. Apparently there have been some developments in the agricultural sector causing heterogeneity in occupational structures across regions to increase over time. It can also be noted from the table that the estimated coefficient for the coverage is again highly significant (and positive). In general, the finding that there are certain regions and sectors which have a significant trend and others that do not have a significant trend, indicates that there is likely to be heterogeneity between regions and sectors in heterogeneity developments over time. Lastly, note that as indicated in the introduction, this paper will not dig any deeper into the underlying reasons for observing these convergence/divergence results. This is a task that remains for future research.

Table 6: Estimation results MODEL 2 using OLS estimator - occupational structures.

		MODEL 2
Parameter estimates		
Region Trend	Stedendriehoek	0.0028*
		(0.0013)
	Midden-Gelderland	0.0025*
		(0.0014)
	Flevoland	0.0029*
		(0.0014)
	Midden-Utrecht	0.0027*
		(0.0013)
	Noord-Holland Noord	0.0028*
		(0.0013)
	Gorinchem	0.0029*
		(0.0014)
Sector Trend	Agriculture†	0.0093**
		(0.0011)
	Metal industry†	0.0026*
		(0.0010)
	Other industry†	0.0026*
		(0.0011)
	Financial services	0.0021*
		(0.0010)
	Culture, sports and recreation†	0.0032*
		(0.0016)
	Other services†	0.0029*
		(0.0014)
Coverage		0.3080**
		(0.0089)
Regression diagnostics		
R-squared		0.915
$\hat{\sigma}_\epsilon$		0.060
N		5830
Number of groups		530

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$

Source: own calculations based on EBB (Statistics Netherlands).

7 Results Part 2: Heterogeneity in Educational Structures

Now that the regional heterogeneity in occupational structures has been examined, this section provides the results for the educational structures. Similar to the previous section, the first subsection lays out the current heterogeneity patterns in the Netherlands. The second subsection then looks at how the heterogeneity has evolved over time.

7.1 Current Heterogeneity in Regional Sectoral Structures

This subsection aims to give an answer to the question to what extent current regional sectoral educational structures differ across regions. To this purpose, Table 7 displays for every sector the $\overline{KSI_{.st}}$, the $\overline{AD_{.st}}$ and the $\overline{coverage_{.st}}$, averaged over $t = 2014$ and $t = 2016$. Note that besides the regional mean of the three main measures, the table also reports the regional standard deviation of each of the three measures in parentheses behind the mean.

Table 7: Regional average and standard deviation of KSI, AD and coverage per sector (x100), averaged over 2014 and 2016 - educational structures.

Sector	$\overline{\text{KSI}}$		$\overline{\text{AD}}$		$\overline{\text{Coverage}}$	
Agriculture†	28.9	(7.3)	2.0	(0.5)	85%	(5)
Food and stimulants industry†	28.1	(7.3)	2.0	(0.5)	85%	(5)
Chemical industry†	31.2	(8.0)	2.2	(0.7)	87%	(4)
Metal industry†	28.6	(7.1)	2.1	(0.8)	88%	(5)
Other industry†	28.2	(6.3)	1.7	(0.4)	89%	(4)
Construction	23.1	(6.3)	1.9	(0.8)	87%	(6)
Retail	22.0	(5.5)	1.1	(0.4)	95%	(4)
Wholesale	27.4	(5.9)	1.8	(0.7)	87%	(9)
Transport and storage	27.5	(5.9)	1.9	(0.8)	86%	(8)
Catering	22.8	(7.4)	2.4	(1.3)	83%	(9)
Information and communication†	27.0	(6.8)	2.7	(0.9)	86%	(6)
Financial services and real estate	27.0	(7.1)	2.7	(1.4)	85%	(9)
Specialist business services	27.3	(7.5)	2.0	(1.0)	90%	(8)
Rental and other business services	26.6	(7.8)	1.7	(0.7)	84%	(14)
Public governance	27.5	(6.8)	1.6	(0.8)	84%	(9)
Education	23.6	(7.5)	2.0	(1.2)	90%	(7)
Healthcare	25.7	(9.5)	1.9	(1.0)	93%	(4)
Well-being	21.7	(6.5)	1.4	(0.6)	95%	(4)
Culture, sports and recreation†	30.2	(9.2)	2.8	(1.0)	84%	(5)
Other services†	31.5	(7.8)	2.3	(0.5)	86%	(4)
Average of all sectors	26.8	(7.0)	2.0	(0.8)	87%	(12)

Source: own calculations based on EBB (Statistics Netherlands).

Let us first examine the values of $\overline{KSI_{.st}}$, as these provide us with a basic idea of how much heterogeneity there currently is in educational structures. Remember that these are in fact lower bounds for the actual unobserved total percentage point deviations

$\overline{KSI^*_{.st}}$,. However, as the coverage is on average 87% and nowhere lower than 84%, these $\overline{KSI_{.st}}$, values are likely to be the close to the actual unobserved total percentage point deviations. From the bottom line of the table, it can be seen that the average total percentage point deviation of a region's sectoral educational structure from the national sectoral occupational structure across all sectors currently is at least around percentage points. Dividing this number by two and thereby calculating the CRS, yields an average CRS of 13.5 percentage points. This means that on average, for a certain combination of sector and region, one currently would have to reallocate at least approximately 13.5% of the workforce to a different type of education to obtain the national educational structure for that sector. Based on an average of 11000 workers per combination of sector and region, this amounts to approximately 1500 workers. Furthermore, looking at sector specific values of $\overline{KSI_{.st}}$, it can be seen that all values lie roughly between 21 and 32 percentage points. From this it can be concluded that heterogeneity in educational structures is more uniformly distributed across sectors than heterogeneity in occupational structures. Furthermore, the three sectors with the most heterogeneity are the chemical industry, the culture, sports and recreation sector and the other services sector. Their index values lie around 31 percentage points. At the lower end of the $\overline{KSI_{.st}}$ distribution the sectors construction, retail, catering, and well-being can be found, with a $\overline{KSI_{.st}}$, of around 22 percentage points. The remaining sectors lie relatively close to the average value of the index.

Looking at the last line of Table 7, it can be seen that the average of $\overline{AD_{.st}}$, across sectors is equal to 2.0. This means that, on average, the current observed educational shares deviate 2.0 percentage points from their corresponding national share. Dividing the average of $\overline{KSI_{.st}}$, across sectors by this value, it can be concluded that, on average, the observed values for $\overline{KSI_{.st}}$, are the result of observed deviations for around thirteen or fourteen different types of education. Furthermore, from the sector specific values of $\overline{AD_{.st}}$, it can be seen these values generally don't vary too much around 2.0. Therefore, it can be concluded that in most sectors the deviations at the extremes of the distribution are equal to around 2.2 percentage points. Another conclusion which can be drawn from this, is that even though some sectors have higher $\overline{KSI_{.st}}$, values than others, this seems to be mainly driven by the fact that such sectors have more heterogeneous educational structures in terms of number of relevant types of education. Note that this conclusion has also been drawn for the occupational segments.

7.2 Heterogeneity in Regional Sectoral Structures over Time

This subsection aims to answer the question whether heterogeneity in regional sectoral educational structures has steadily decreased or increased over time, i.e. whether there are any convergence or divergence patterns. To this purpose, let us first consider the results from estimating MODEL 1 for the educational structures using the POLS estimator. The estimation results are displayed in Table 8. Note that this model is estimated using 5742 observations for KSI_{rst}. To see this, consider the fact that twelve `complete' sectors are included in the estimation (i.e. twelve sectors for which all 35 regions are included) and that there are nine sector for which only a subset of the regions is included (the subset displayed in the second table in Appendix D). This subset contains in total 102 unique sector/region combinations. As a result, the total number of observations is indeed equal to 5742 (calculated as 102 * 11 (periods) + 35 * 12 (sectors) * 11 (periods)).

Table 8 shows that the estimated coefficient for the trend is positive and highly significant. The estimate implies that, on average, the KSI_{rst} for educational structures is estimated to increase with 0.26 percentage points every year. Hence, these findings provide evidence in favour of σ-divergence of sectoral educational structures over time, be it very slowly. The increase of 0.26 percentage points every year implies an increase in KSI_{rst} measured heterogeneity of 2.6 percentage points every ten year. This is a rather small increase in the context of the values found in the previous subsection. Note also that just as for the occupational structures, the coverage is highly significant.

Table 8: Estimation results MODEL 1 using POLS estimator - educational structures.

	MODEL 1
Parameter estimates	
Trend	0.0026**
	(0.0002)
Coverage	0.1675**
	(0.0069)
Regression diagnostics	
R-squared	0.206
$\hat{\sigma}_{\epsilon}$	0.073
N	5742
Number of groups	522

Clustered standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$

Source: own calculations based on EBB (Statistics Netherlands).

To get a more nuanced view of convergence/divergence patterns in heterogeneity in educational structures, it is now looked at the estimation results for MODEL 2. This model allows both all regions and all sectors to display different trends at the same time. The results therefore give some idea as to how certain regions or sectors are driving the estimation results for the trend in MODEL 1. Table 9 displays all estimated coefficients for the γ^r and γ^s which are significant at the 5% level. Moreover, also the coefficient for the coverage is reported. Note that these estimation results are again obtained using the simple Ordinary Least Squares (OLS) estimator.

Perhaps the most striking observation from these estimation results, is that almost all regions and sectors are found to have a significant positive trend in their KSI_{rst}. More specifically, the table includes 28 out of 35 regions and 20 out of 20 sectors, and all of them have an estimated coefficient which is positive. Apparently σ-divergence patterns in regional sectoral educational structures are widespread across different regions and sectors. Nevertheless, the differences in estimates across regions and sectors indicate that there is still some heterogeneity across regions and sectors in the development of heterogeneity. More specifically, the magnitude of the significant coefficients ranges between 0.0019 and 0.0067. On a ten year basis, this range amounts to a $KSI_{rst}.$ increase ranging between 1.9 and 6.7 percentage points. In the context of the $\overline{KSI_{rst}}.$ values found in the previous subsection, significant trends at the higher end of this range seem to be rather substantial. The regions and sectors with a trend above 0.006 are Drenthe, Rivierenland, Gorinchem, chemical industry, other industry, culture, sports and recreation and other services. Why these region and sector specific σ-divergence patterns are observed, is a question that remains for future research.

Table 9: Estimation results MODEL 2 using OLS estimator - educational structures.

		MODEL 2
Parameter estimates		
Region Trend	Friesland	0.0040**
	Drenthe	0.0060**
	Regio Zwolle	0.0037**
	Twente	0.0030**
	Stedendriehoek	0.0029*
	Midden-Gelderland	0.0032**
	Rijk van Nijmegen	0.0040**
	Achterhoek	0.0055**
	Rivierenland	0.0060**
	Flevoland	0.0044**

	Midden-Utrecht	0.0034**
	Noord-Holland Noord	0.0027*
	Zuid-Kennemerland	0.0028*
	Groot Amsterdam	0.0040**
	Midden-Holland	0.0044*
	Haaglanden	0.0027*
	Drechtsteden	0.0054**
	Zeeland	0.0042**
	West-Brabant	0.0027*
	Midden-Brabant	0.0027*
	Noordoost-Brabant	0.0024*
	Zuidoost-Brabant	0.0040**
	Noord-Limburg	0.0051**
	Zuid-Limburg	0.0034**
	Food-Valley	0.0027*
	Helmond-De Peel	0.0051**
	Midden-Limburg	0.0041**
	Gorinchem	0.0062**
Sector Trend	Agriculture†	0.0049**
	Food and stimulants industry†	0.0053**
	Chemical industry†	0.0061**
	Metal industry†	0.0057**
	Other industry†	0.0065**
	Construction	0.0047**
	Retail	0.0023*
	Wholesale	0.0047**
	Transport and storage	0.0044**
	Catering	0.0019*
	Information and communication†	0.0043**
	Financial services	0.0047**
	Specialist business services	0.0031**
	Rental and other business services	0.0036**
	Public governance	0.0037**
	Education	0.0041**
	Healthcare	0.0043**
	Well-being	0.0020*
	Culture, sports and recreation†	0.0067**

Other services†	0.0061**
Coverage	0.2832**
	(0.0072)
Regression diagnostics	
R-squared	0.941
$\hat{\sigma}_\epsilon$	0.059
N	5742
Number of groups	522

All standard errors (robust) for region trends are between 0.0011 and 0.0012, all standard errors (robust) for sector trends are between 0.0009 and 0.0015.

* $p < 0.05$, ** $p < 0.01$

Source: own calculations based on EBB (Statistics Netherlands).

8 Current Heterogeneity in Regional Sectoral Occupational and Educational Structures at the Level of Specific Occupations and Types of Education

The previous two sections have given a general idea to what extent sectoral occupational and educational structures differ across regions currently, and how heterogeneity in these structures has developed over time. To enhance our understanding of where the observed current heterogeneity comes from (as observed in Sections 6.1 and 7.1), this section examines for each sector which types of occupation and education currently show the most heterogeneity across regions. Not only will this analysis help us to understand how the KSI values are generated, it is also able to point out certain combinations of sector and occupation/education for which assuming the national composition is relatively much more harmful than doing so for other combinations. To this purpose, the first subsection examines occupation specific heterogeneity for each sector and the second subsection examines education specific heterogeneity for each sector. Note that to ensure coherence with the previous sections, the analysis in this section again excludes the same sector/region combinations as before (i.e. the combinations displayed in Appendix D).

8.1 Occupation Specific Heterogeneity

This subsection aims to identify for each sector for which occupational segments the sectoral employment share varies the most across the country. In order to do so, a straightforward idea is to look at the distributions of the Z_{krst} across the regions, for a give occupation, sector and period (as defined in the methodology section). As the interest of this subsection is to identify current heterogeneity patterns, it is chosen to average regional sectoral occupation shares over the last two periods of the data. This implies that the Z_{krst} are obtained in a slightly adjusted way. For example, if the share of doctors in the healthcare sector in Groningen was 30% in 2014 and 40% in 2016, the share of doctors in Groningen is taken to be 35%. Similarly, for each occupation it is also chosen to average its share in the national workforce over 2014 and 2016. Subsequently, using these two obtained numbers, a certain occupation's regional sectoral share's deviation from the national share can be calculated. This is thus a slightly adjusted version of Z_{krst}. The reason to average the shares over two periods is to reduce the noise which is present in the data. This is the same reason as why in the static heterogeneity analysis the calculated KSI were averaged over the periods 2014 and 2016. Note that, if for a certain regional share its value is missing in one of the two periods, its value is simply set equal to the single share that is available. The same is applied to calculating the national shares.

Ultimately, it was chosen to report for each sector the three occupational segments with the highest regional average of the slightly adjusted Z_{krst}, as these are the most heterogeneous occupations across regions for a given sector. Let's define this regional average as $\overline{Z_{k.st}}$. Note that by doing so, one is probably not picking up peculiarities which only occur in a few regions. If there are two regions for which the sectoral share of a certain occupation deviates a lot from the national share and for all the other regions that regional share is equal to the national share, this occupation will probably be not among the occupations with the highest $\overline{Z_{k.st}}$. Such deviations are obviously still interesting. However, there is quite a substantial number of such deviations, and it is nearly impossible to present all of them in a structured and readable way. Moreover, it is generally more interesting to find regularities in heterogeneity than to find peculiarities in heterogeneity. Therefore, it was chosen to confine the analysis to those observed occupations which vary most on average across all regions. Furthermore, it was chosen to only report those combinations of occupational segment and sector for which there is an observation for at least 20 out of 35 regions, to ensure that the observed $\overline{Z_{k.st}}$ is reasonably representative of what happens in that sector in the country as a whole. Note that this requirement can obviously not be applied to the nine sectors for which only a subset of the subset of the regions is selected. For those sectors, it was chosen to require an observation for at least all but one of the regions. Lastly, for each reported occupational segment also the value of its sectoral share in the national occupational distribution is given.

Table 10 displays the results for all 20 sectors. It can be seen that most average absolute deviations are not too far from the average $\overline{AD_{.st}}$. across sectors, obtained in Section 6.1. The average deviation found there was equal to 2.2 percentage points. From this it can be concluded that even those occupations which show most heterogeneity, generally don't show much more heterogeneity than the average. However, this is not true for all combinations of sector and occupation, as there a few which seem be characterized by relatively more heterogeneity. These are worth wile elaborating a bit more on. It concerns the combinations gardeners, arable farmers and cattle breeders for the agricultural sector, vehicle drivers and mobile machine operators for the transport and storage sector, writers and artists for the culture, sports and recreation sector, and personal services employees for the other services sector. All these combinations display relatively much heterogeneity across regions, with a value for $\overline{Z_{k.st}}$ of at least 6.9. To get a better idea of how each of these occupations' shares are distributed across the country and the underlying reasons for this, it is useful to look for each of these combinations at a graph of the Netherlands with the regional values of the shares' deviation from the National share. Note that these graphs do not display the absolute values of these deviations, but the actual deviations (i.e. not the Z_{krst}. but simply the regional sectoral share minus the corresponding national sectoral share). Note also that the legend corresponding to these figures can be found in Figure 1. If a region is coloured blue, this means that its sectoral share for a certain occupation is higher than the national sectoral share for that occupation (and vice versa for orange). Regions which are labeled `no data', are either regions which are omitted as a result of the selection corresponding to Appendix D, or regions which don't have an observation for that specific share.

Table 10: Top 3 highest $\overline{Z_{k.st}}$ per sector (ordered high to low) and corresponding national shares (x100), averaged over 2014 and 2016 - occupations.

Sector	Occupational Segment	$\overline{Z}$	Share
Agriculture†	91. Gardeners, arable farmers and cattle breeders	9.0	56%
	92. Agricultural auxiliary workers	4.1	10%
	121. Vehicle drivers and mobile machine operators	2.0	6%
Food and stimulants industry†	33. Salesman	4.0	12%
	77. Production-machine operators and assembly workers	3.4	11%
	72. Nature and technical specialists	2.6	7%
Chemical industry†	72. Nature and technical subject specialists	5.2	16%
	77. Production-machine operators and assembly workers	2.3	14%
	122. Transport and logistics auxiliary workers	2.3	3%
Metal industry†	74. Metal workers, machine operators	4.4	21%
	71. Engineers math, nature and technical sciences	2.5	9%
	43. Administrative personnel	2.0	10%
Other industry†	75. Food processing occupations and other crafts	2.0	12%
	43. Administrative personnel	1.5	11%
	73. Construction workers	1.5	5%
Construction	73. Construction workers	3.6	43%
	76. Electricians	1.9	8%
	43. Administrative personnel	1.8	9%
Retail	33. Salesman	2.7	48%
	122. Transport and logistics auxiliary workers	1.5	15%
	74. Metal workers, machine operators	1.2	6%
Wholesale	33. Salesman	2.6	15%
	43. Administrative personnel	1.9	15%
	122. Transport and logistics auxiliary workers	1.7	6%
Transport and storage	121. Vehicle drivers and mobile machine operators	6.9	39%
	43. Administrative personnel	3.3	24%
	111. Personal services employees	2.1	4%
Catering	111. Personal services employees	3.5	49%

Sector	Occupational Segment	$\overline{Z}$	Share
	112. Cleaning staff and kitchen aids	3.2	17%
	33. Salesman	2.1	10%
Information and communication†	81. Ict specialists	5.0	37%
	21. Writers and artists	3.0	9%
	43. Administrative personnel	2.2	8%
Financial services	43. Administrative personnel	3.5	17%
	41. Business management and administrative specialists	2.4	19%
	42. Business management and administrative subject specialists	2.2	16%
Specialist business services	72. Nature and technical subject specialists	5.2	16%
	77. Production-machine operators and assembly workers	2.3	14%
	122. Transport and logistics auxiliary workers	2.3	3%
Rental and other business services	112. Cleaning staff and kitchen aids	2.7	19%
	111. Personal services employees	2.0	6%
	91. Gardeners, arable farmers and cattle breeders	1.8	5%
Public governance	41. Business management and administrative specialists	2.6	15%
	63. Security workers	2.6	15%
	61. Government officials	2.5	14%
Education	11. Teachers	3.4	60%
	41. Business management and administrative specialists	1.3	4%
	43. Administrative personnel	1.0	6%
Healthcare	101. Doctors and therapists	4.0	35%
	103. Healthcare subject specialists	3.3	21%
	105. Nurses	2.5	4%
Well-being	105. Nurses	3.0	27%
	104. Social workers	2.3	19%
	13. Day-care employees and teaching assistants	2.2	10%
Culture, sports and recreation†	21. Writers and artists	7.5	25%
	111. Personal services employees	4.8	9%
	43. Administrative personnel	2.8	8%
Other services†	111. Personal services employees	10.0	40%
	112. Cleaning staff and kitchen aids	2.4	4%
	42. Business management and administrative subject specialists	2.2	4%

Source: own calculations based on EBB (Statistics Netherlands).

Firstly, consider Figure 2 displaying for the agricultural sector whether a certain region's share of gardeners, arable farmers and cattle breeders is higher or lower than the national share, and how much higher or lower. Firstly, note that the regions which are omitted as a result of the selection based on the coverage (i.e. the regions which are coloured black), seem to be the most rural areas of the country. This is a sensible observation, as such regions typically have less space for agricultural activities. Furthermore, based on this graph it seems that region's in which this type of occupation is relatively more important in this sector are located more in the east of the country, and region's in which this occupation is relatively less important are located in the west of the country. It is not immediately clear why this would be the case. A potential reason could be that the west of the country is located near the sea, and that these regions are therefore relatively more specialised in fishing (note that fishing is also included in the agricultural sector). This would be an example of withinsector regional specialisation, and could explain the relatively low presence of standard agricultural occupations captured by the segment gardeners, arable farmers and cattle breeders. Furthermore, having less people working in one category means having more people working in other categories. In this regard, it has been examined whether there is a certain other occupational segment which is generally more present in these western regions as compared to the eastern regions. It is found that this is roughly the case for the occupational segment agricultural auxiliary workers. This can be seen from Figure 3. Perhaps this is a result of western regions being relatively more specialized in fishing. However, this is rather speculative.

Figure 1: Legend corresponding to Figures 2, 3, 4, 5, 6, and 7.

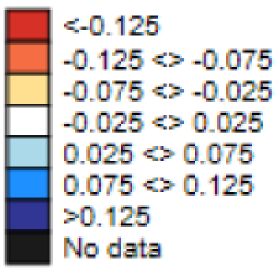
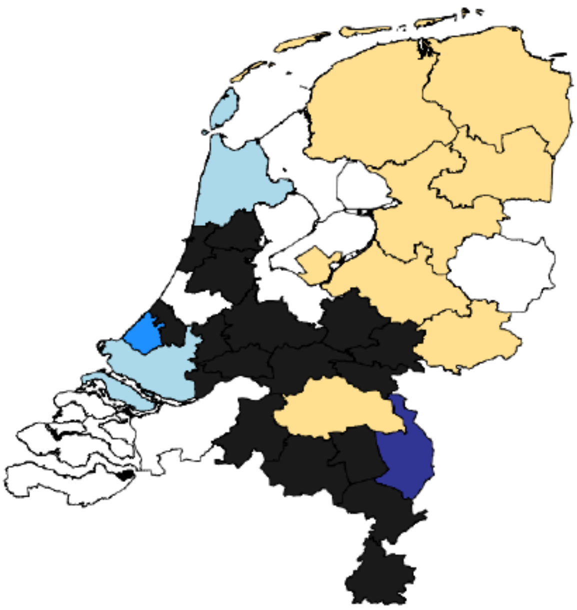


Figure 2: Deviation of regional sectoral share of gardeners, arable farmers and cattle breeders from its corresponding national share - agricultural sector.

Source: own calculations based on EBB (Statistics Netherlands).

Figure 3: Deviation of regional sectoral share of agricultural auxiliary workers from its corresponding national share - agricultural sector.



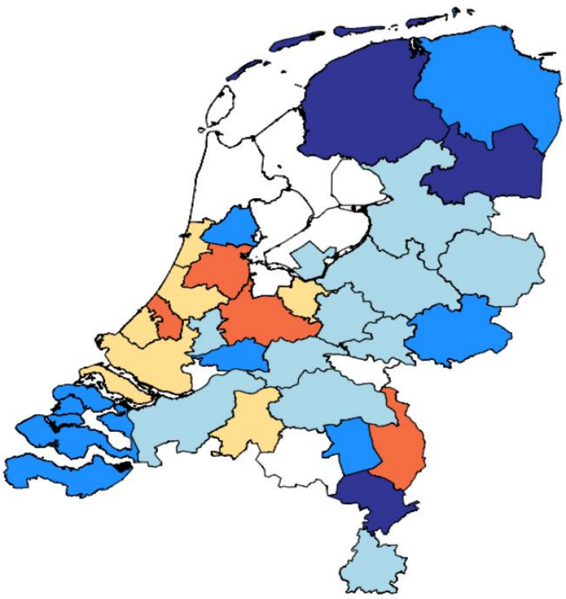
Source: own calculations based on EBB (Statistics Netherlands).

Next, let us consider the distribution of vehicle drivers and mobile machine operators shares across the Netherlands for the transport and storage sector, displayed in Figure 4. It can clearly be seen that the transport and storage sector in the mid-west of the country employs relatively less vehicle drivers and mobile machine operators than this sector does in other parts of the country. As indicated in the background section, an important distinguishing feature for this part of the country with regard to the transport and storage sector, is that it contains the large Rotterdam harbour and the large Amsterdam airport Schiphol. This is yet another example of within-sector specialisation, and could explain why this region would have relatively less vehicle drivers and mobile machine operators, as such large transit hubs obviously require different staffing than more small-scale transport and storage operations. One may wonder how this pattern for the mid-west of the country can be attributed to two transit hubs which are exclusively part of only two of these regions. In this regard it should be noted that in the data a worker's region is the region in which he or she lives. The region of working and living will often be the same, but this is not a necessity. Especially in the densely populated mid-west of the country, one can expect people to live a bit further from their place of work.

What type of occupations might then replace the vehicle drivers and mobile machine operators in the mid-west of the country? Some investigation of the data has pointed out that up to a certain extent these region roughly have more personal services employees as compared to other parts of the country. This can be seen from Figure 5. Especially for the airport region, it makes sense that personal services are relatively more important. This is confirmed by observing the blue area in the north-west of the country. Just as for the previous combination of occupation and sector that was examined, it remains rather unclear why these patterns are observed. However, for the transport and storage sector, the Rotterdam harbour and Amsterdam airport do seem to play an important role.

Figure 6 displays the deviation from the national share for writers and artists in the culture, sports and recreation sector. From the graph it can be seen that this is a rather small sector (in terms of employment) in most labour market regions in the Netherlands, given the large number of black coloured regions. Furthermore, note that the share of writers and artists is substantially above average for the regions Groot Amsterdam and Midden-Utrecht. As implicit in their names, these regions contain respectively the two large cities Amsterdam and Utrecht. The observation that these regions employ relatively more writers and artists in this sector is logical, as these are the regions which are specialised in the `culture' part of this sector. Cities have always been a hub for culture and creativity related businesses, and this doesn't seem to be any different in the Netherlands. Furthermore, note that the graph also shows that the region Midden-Brabant has a fraction of writers and artists in this sector which is substantially less than the national average (the orange region). A sensible explanation for this, is the presence of the large theme park `De Efteling' in this region. It is indeed found that this region has a relatively larger share of services related personnel. It can thus be said that this region seems to be relatively more specialised in the `recreation' part of the sector.

Figure 4: Deviation of regional sectoral share of vehicle drivers and mobile machine operators from its corresponding national share - transport and storage sector.



Source: own calculations based on EBB (StatisticsNetherlands).

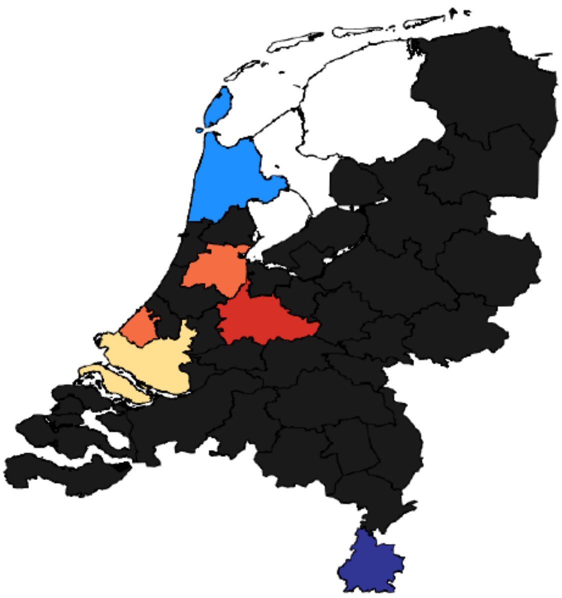
Figure 5: Deviation of regional sectoral share of personal services employees from its corresponding national share - culture, sports and recreation sector.

Source: own calculations based on EBB (Statistics Netherlands).

Figure 6: Deviation of regional sectoral share of writers and artists from its corresponding national share -transport and storage sector.

Source: own calculations based on EBB (Statistics Netherlands).

Figure 7: Deviation of regional sectoral share of personal services employees from its corresponding national share - other services sector.



Source: own calculations based on EBB (Statistics Netherlands).

Lastly, consider the deviation of the share of personal services employees in the other services sector, displayed in Figure 7. With regard to the observed deviation pattern for this specific combination of occupation and sector, I could not come up with a sensible specific explanation. However, note that for this sector it is not very unexpected to encounter a lot of heterogeneity in occupational structures across regions in general, given that it is defined as `other' services. This means that it captures plenty of different types of economic activities which cannot be captured by any of the other sector categories.

8.2 Education Specific Heterogeneity

After having examined current regional heterogeneity in occupational structures at the level of specific occupations, this subsection will focus on current regional heterogeneity in educational structures at the level of specific types of education. This will provide us with some information on how the current KSI values related to educational structures have been generated. Moreover, it can point out certain combinations of education and sector for which it is relatively much more harmful to assume the national composition than doing so for others. To this purpose, this subsection proceeds in exactly the same way as the previous subsection. This implies that the main focus lies on the slightly adjusted version of Z_{krst} and its regionally averaged version $\overline{Z_{k.st}}$. Again, for each sector the three types of education with the highest $\overline{Z_{k.st}}$ are reported, ordered high to low. For those sectors for which no prior sample selection has been performed, it was again chosen to only report those combinations of education and sector for which there is an observation for at least 20 out of 35 regions. Also the requirement of all but one region having an observation for a certain type of education for the remaining sectors (i.e. the sectors from Appendix D) is maintained.

Table 11 displays the results for all 20 sectors. It can be seen that most average absolute deviations are not too far from the average $\overline{AD_{.st}}$ across sectors, obtained in Section 7.1. The average deviation found there was equal to 2.0 percentage points. From this it can be concluded that even those observed types of education which show the most heterogeneity, generally don't show much more heterogeneity than the average. However, this is not true for all combinations of sector and education, as there a few which seem be characterized by relatively much heterogeneity. These are worth wile elaborating a bit more on. It concerns the combinations mbo4 - technical and ict for the chemical industry, master - economics and society for the financial services sector, master - economics and society for the specialist business services sector, and bachelor - economics and society for the culture, sports and recreation sector. Each of these combinations has an average absolute deviation of at least 6.3 percentage points. To examine these combinations in more detail, it is useful to again look for each of these combinations at a graph of the Netherlands with the regional values of the shares' deviation from the National share. The same legend is used as before, and can be found in Figure 8.

Table 11: Top 3 highest $\overline{Z_{k.st}}$. per sector (ordered high to low) and corresponding national shares (x100), averaged over 2014 and 2016 - education.

Sector	ONRsector	$\overline{Z}$	Share
Agriculture†	242. Mbo4 - agriculture and nature	3.9	15%
	122. Vmbo-b/k, mbo1 - agriculture and nature	3.1	8%
	110. Primary education	2.9	8%
Food and stimulants industry†	210. Havo, vwo	3.2	9%
	130. Vmbo-g/t, havo-, vwo-onderbouw	2.6	9%
	223. Mbo2 - technical and ict	2.4	2%
Chemical industry†	243. Mbo4 - technical and ict	6.3	12%
	223. Mbo2 - technical and ict	3.0	5%
	110. Primary education	2.6	5%
Metal industry†	123. Vmbo-b/k, mbo1 - technical and ict	3.2	11%
	233. Mbo3 - technical and ict	3.1	9%
	243. Mbo4 - technical and ict	3.0	13%
Other industry†	110. Primary education	2.0	12%
	233. Mbo3 - technical and ict	2.2	7%
	243. Mbo4 - technical and ict	2.4	9%
Construction	233. Mbo3 - technical and ict	3.1	13%
	123. Vmbo-b/k, mbo1 - technical and ict	2.8	16%
	243. Mbo4 - technical and ict	2.5	14%
Retail	210. Havo, vwo	2.4	14%
	130. Vmbo-g/t, havo-, vwo-onderbouw	1.9	17%

Sector	ONRsector	$\overline{Z}$	Share
	241. Mbo4 - economics and society	1.7	7%
Wholesale	311. Bachelor - economics and society	2.7	13%
	241. Mbo4 - economics and society	2.7	9%
	321. Master - economics and society	2.4	6%
Transport and storage	110. Primary education	2.7	8%
	210. Havo, vwo	2.7	10%
	123. Vmbo-b/k, mbo1 - technical and ict	2.5	7%
Catering	210. Havo, vwo	4.8	20%
	130. Vmbo-g/t, havo-, vwo-onderbouw	3.2	17%
	110. Primary education	2.8	10%
Information and communication†	321. Master - economics and society	4.6	14%
	313. Bachelor - technical and ict	4.3	14%
	311. Bachelor - economics and society	3.8	17%
Financial services	321. Master - economics and society	6.5	18%
	311. Bachelor - economics and society	4.0	23%
	241. Mbo4 - economics and society	4.0	10%
Specialist business services	321. Master - economics and society	6.4	22%
	311. Bachelor - economics and society	2.9	18%
	323. Master - technical and ict	2.2	6%
Rental and other business services	210. Havo, vwo	3.3	12%
	311. Bachelor - economics and society	3.2	12%
	130. Vmbo-g/t, havo-, vwo-onderbouw	2.2	8%
Public governance	321. Master - economics and society	5.1	16%
	311. Bachelor - economics and society	2.5	13%
	241. Mbo4 - economics and society	2.1	7%
Education	314. Bachelor - Healthcare, education and services	5.6	32%
	321. Master - economics and society	4.3	14%
	311. Bachelor - economics and society	2.3	11%

Sector	ONRsector	$\overline{Z}$	Share
Healthcare	324. Master - Healthcare, education and services	4.9	17%
	244. Mbo4 - Healthcare, education and services	3.1	11%
	314. Bachelor - Healthcare, education and services	3.0	23%
Well-being	244. Mbo4 - Healthcare, education and services	3.2	17%
	314. Bachelor - Healthcare, education and services	2.4	16%
	234. Mbo3 - Healthcare, education and services	2.3	15%
Culture, sports and recreation[†]	311. Bachelor - economics and society	7.1	23%
	321. Master - economics and society	5.4	11%
	130. Vmbo-g/t, havo-, vwo-onderbouw	2.8	8%
Other services[†]	321. Master - economics and society	5.6	10%
	244. Mbo4 - Healthcare, education and services	4.1	10%
	234. Mbo3 - Healthcare, education and services	4.0	13%

Source: own calculations based on EBB (Statistics Netherlands).

Firstly, let us examine the regional shares of people with mbo4 - technical and ict education in the chemical industry sector relative to the national share, as displayed in Figure 9. As a result of the selection based on the coverage, not many regional shares are left. Note that it is not entirely clear why this pattern is observed. A potential reason could be differences in the regional supply of educational programs, as I cannot see why there would be within-sector specialisation across regions in this sector. However, one would have to look into this in more detail.

Secondly, let us examine the deviations for the combinations master - economics and society and the financial services sector, and master - economics and society and the specialist business services sector. These regional deviations are displayed respectively in Figures 10 and 11. From the figures it can be seen that both combinations display a very similar pattern, which is exactly the reason why these combinations are looked at together. It can be seen that for both sectors the share of people with master - economics and society education lies above the national average in roughly the rural area in the mid-west of the country, and below average for the remaining areas.

Figure 8: Legend corresponding to Figures 9, 10, 11, 12

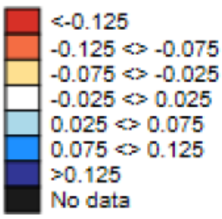


Figure 9: Deviation of regional sectoral share of people with mbo4 - technical and ict education from its corresponding national share - chemical industry sector.

Source: own calculations based on EBB (Statistics Netherlands).

This rural area in the mid-west of the country is often called Randstad, and contains large cities such as Den Haag, Amsterdam, Rotterdam and Utrecht. What could potentially explain this observed pattern? One potential answer lies in the concept of external economies of scale, given that both these sectors are service sectors. To understand why such external economies can explain the strongly clustered pattern that is observed (both within and across regions) for people with master -economics and society education in these two service sectors, one should first accept that essentially what we are looking at, is the spread of master educated people in general in these sectors. The reason for this, is that the share of master educated workers specialised in different fields of study in these sectors is negligible. Furthermore, note that these kind of services are generally knowledge intensive for parts of their production. It is not unreasonable to think that mainly those sector parts which are highly knowledge intensive (and thus require master educated workers), benefit relatively more from external economies than other sector parts. Firstly, even though the supply of master educated people in the Netherlands is relatively large as compared to other western economies[9], the demand for such highly educated people is also relatively large as a result of the Netherlands having a knowledge-intensive economy. The result is that it

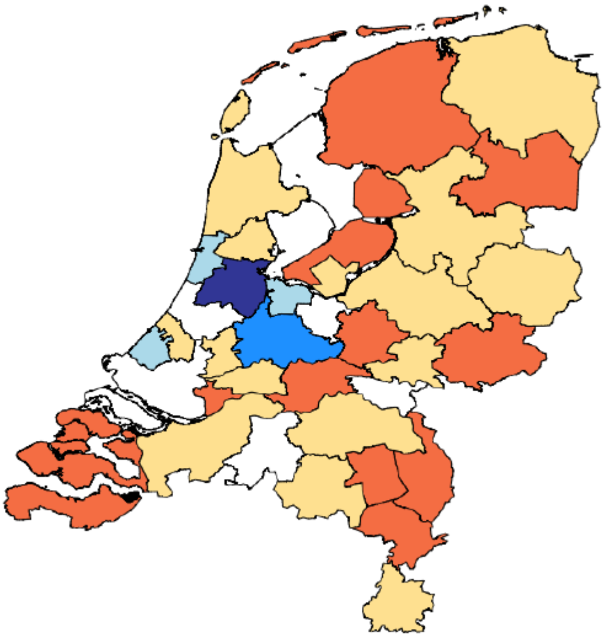
[9] In 2018, 11% of the Dutch population had obtained a master degree (Onderwijs in Cijfers, 2019).

is not always easy for organisations to obtain the required skill-level for their workforce. Therefore, the parts of these sectors requiring master educated workers benefit greatly from reduced moving cost as a result of clustering. In other words, these activities benefit relatively more from a `thick' labour market. If they would simply locate themselves randomly through the country, they would have a hard time finding enough qualified employees. Another obvious reason why the parts of these sectors which require master educated employees benefit relatively more from external economies, is that knowledge spillovers are relatively important for these economic activities. The argument for this is pretty obvious, as the highly educated nature of these activities illustrates the need for knowledge. Note that the line of argumentation presented here is a nice application of how the theories underlying the New Economic Geography can explain within-sector specialisation, and hence can explain differences in occupational and educational structures.

Figure 10: Deviation of regional sectoral share of people with master - economics and society education from its corresponding national share - financial services sector.

Source: own calculations based on EBB (Statistics Netherlands).

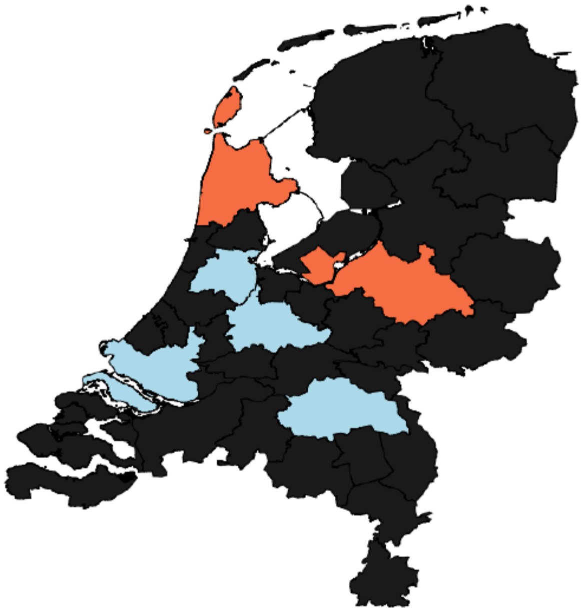
Figure 11: Deviation of regional sectoral share of people with master - economics and society education from its corresponding national share - specialist business services sector.



Source: own calculations based on EBB (Statistics Netherlands).

A last combination of education and sector this paper wishes to highlight, is bachelor – economics and society educated people in the culture, sports and recreation sector. Figure 12 shows that regional deviations from the national share for this combination. Also for this sector, a substantial number of regions have been excluded as a result of the selection based on the coverage. A potential reason for the observed deviation pattern can be found in the deviation pattern from Figure 6. Here it was observed that the regions Groot Amsterdam and Midden-Utrecht have a relatively large share of writers and artists. From Figure 12 it is observed that both these regions have a higher than average share of workers with bachelor - economics and society. These observations would confirm the expectation that the `cultural' part of this sector is more highly educated as compared to the `sports and recreational' part. This can probably also explain the deviation pattern for the remaining observed regions in Figure 12.

Figure 12: Deviation of regional sectoral share of people with bachelor - economics and society education from its corresponding national share - culture, sports and recreation sector.



Source: own calculations based on EBB (Statistics Netherlands).

9 Extension

In constructing the labour market forecasts for POA, ROA requires forecasts for the regional sectoral occupational and educational compositions. As indicated in the introduction, ROA currently assumes the regional sectoral compositions are the same as the national sectoral compositions. This justifies the current procedure of only forecasting national sectoral occupational and educational shares. This research has shown, however, that some heterogeneity in regional sectoral structures is present. Therefore, the assumption that regional and national structures are always the same does not seem to be entirely justified. A natural question which arises then, is whether we could improve the forecasts of the regional sectoral shares by directly modelling the regional shares themselves. This section provides an exploratory analysis of this question. In order to do so, for both occupational and educational shares two simple forecasting models based on regional shares are developed. The forecasting performance of these two models should be able to give a reasonable indication of the value which lies in the regional data. After developing these models based on the regional shares, the forecasting performance will be compared to a naive forecasting model which mimics POA's forecasting model. This allows us to conclude whether the regional data are of added value or not.

Just like in the previous sections, the incompleteness of the dataset also plays an important role here. The models are biased towards shares based on at least ten observations, and it is unclear whether the estimation results (and forecasting performance) of these models can be extrapolated to these unobserved shares. Nevertheless, the results will provide a general idea of the added value of modelling regional shares as compared to only modelling national shares. As the same aggregation levels are used as before (for the same reasons as before), the shares which are included account on average for roughly 80% of sectoral occupational and educational structures. This implies that even though the model is not based on all shares, it does contain those shares that allow one to almost fully forecast regional sectoral structures. The section begins with developing the forecasting models for the regional occupational shares. Subsequently, the forecasting performance of these models will be compared to the naive model exclusively based on national shares. Afterwards, a completely similar analysis is performed for educational shares. The section ends with a more detailed look at where the regional models perform well, and where they do not. This gives some indication as to how these models could potentially be improved.

9.1 Forecasting Regional Sectoral Occupational Shares

Model Development

As indicated, the aim here is to properly model the occupational shares (i.e. the X_{krst}), and thereby obtain a suitable prediction model for these shares. In the simplest form, one could try to explain the X_{krst} by exclusively using the past value of the national share $Q_{ks,t-2}$. Furthermore, it is also a good idea to include the lagged value of the coverage $coverage_{rs,t-2}$ in the model, as the data shows that the coverage increases over time. As the coverage increases, the value of the shares becomes lower on average as relatively large shares would have been big enough to be observed in the first place. It is mainly the small shares which start appearing as the coverage increases. Note furthermore, that as t takes on the values 1996, 1998, 2000 etc., taking the value of a variable at t-2 indeed corresponds to using the value of a variable in the previous period. More formally, define the following panel model (model 3):

$$X_{krst} = \alpha + \theta Q_{ks,t-2} + \psi coverage_{rs,t-2} + \eta_{krs} + \epsilon_{krst} \text{ (Random Effects)} \qquad (7)$$

Firstly, note that this model also assumes an individual specific effect η_{krs}. Furthermore, it can be seen that this model makes a random effects (RE) assumption. This means that it is assumed that the individual effects are mean independent of past, present and future values of all regressors (Wooldridge, 2010). Let T denote the total number of time periods. Then define $\widetilde{Q_{ks}}$ as the $(T-1)\times 1$ matrix containing the random variables $Q_{ks,1996}$ up until $Q_{ks,2014}$, and define $coverage_{rs}$ as the $(T-1)\times 1$ matrix containing the random variables $coverage_{rs,1996}$ up until$coverage_{rs,2014}$. Formally, this RE assumption can then be written as follows (assumption 1):

$$\mathbb{E}\left(\eta_{krs} | \widetilde{Q_{krs}}, \widetilde{coverage}_{rs}\right) = 0, \text{ for all unique combinations of k, r and s} \qquad (8)$$

This assumption is sufficient for the RE estimator to be consistent, and necessary for it to be efficient. Let us for the moment assume that the assumption holds, and consider the RE estimator estimation results for model 3. These results are displayed in the second column of Table 12. Note that below the estimated coefficients again some regression diagnostics are reported, where $\hat{\sigma}_{\epsilon}^2$ is the estimated variance of the error term ϵ, $\hat{\sigma}_{\epsilon}^2$ is the estimated variance of the individual effects, and $\hat{\rho}$ is the estimated fraction of the variance explained by the individual effects. It can be seen from the table that the lagged value of the national share is highly significant. Moreover, it is able to explain the majority of the variance in the dependent variable, which can be seen from the overall R-squared value of 0.91. This is thus probably already a pretty good prediction model. Also note that the coefficient for the $coverage_{rs,t-2}$ has the expected negative sign. This implies that, all else being constant, those combinations of k, r, s and t with a relatively high coverage have relatively lower shares. However, as indicated, one should be cautious in interpreting these results if assumption 1 does not hold. If this assumption does not hold, it can be said that the model (model 4) has fixed effects (FE) instead of random effects:

$$X_{krst} = \alpha + \theta Q_{ks,t-2} + \psi coverage_{rs,t-2} + \eta_{krs} + \epsilon_{krst} \text{ (fixed effects)} \qquad (9)$$

Table 12: Estimation results regional occupation share modelling.

	MODEL 3 (RE)	MODEL 4 (WI)	MODEL 5 (RE)	MODEL 6 (POLS)	MODEL 6 - Sample subset (POLS)
Parameter estimates					
$Q_{ks,t-2}$	0.956**	0.831**	0.523**	0.523**	0.519**
	(0.004)	(0.0180)	(0.00115)	(0.0115)	(0.0142)
$coverage_{rs,t-2}$	-0.042**	-0.021**	-0.026**	-0.026**	-0.014**
	(0.0017)	(0.0019)	(0.0025)	(0.0025)	(0.0024)
$X_{krs,t-2}$			0.451**	0.451**	0.462**
			(0.0017)	(0.0117)	(0.0142)
Regression diagnostics					
R-squared overall	0.914	0.913	0.935	0.935	0.948
R-squared within	0.162	0.165	0.119		
R-squared between	0.940	0.936	0.978		
$\hat{\sigma}_\eta$	0.012	0.024	0		
$\hat{\sigma}_\epsilon$	0.027	0.027	0.028	0.031	0.027
$\hat{\rho}$	0.165	0.450	0		
N	70639	70639	54283	54283	47685
Number of groups	13126	13126	9564	9564	8032
Hypothesis test					
H_0: $\theta + \beta = 1$, Resulting p-value			0	0	0

Robust standard errors in parentheses,
clustered standard errors for MODEL 6.
* $p < 0.05$, ** $p < 0.01$

Source: own calculations based on EBB (Statistics Netherlands).

To formally test assumption 1, a Hausman test can be employed (Wooldridge, 2010). Such a test compares the estimation results from estimating a certain regression equation using both the RE estimator and a certain FE estimator. Its null hypothesis is that the assumption holds. Employing this test for the model at hand, using the Within (WI) estimator as the FE estimator, a p-value of zero is found. Hence, this provides substantial evidence against assumption 1, and thus against model 3. Instead, model 4 seems to be the more appropriate model.

The estimation results of estimating model 4 using the WI estimator are displayed in the third column of Table 12. It can be seen from the table that the results do not change much. The explanatory power of the model, as measured by the overall R-squared, decreases a bit. Furthermore, the major difference with the RE estimation is that coefficient for the national share has become quite a bit smaller. Note that what is perhaps most striking about these estimation results, is the small value for the R-squared within relative to the R-squared between. Apparently, the model is pretty good at explaining differences between individual panel units (i.e. specific combinations of occupation, region, sector and time) but not so good at explaining changes within individual panel units. It thus seems to be the case that share dynamics on the national level are not very good predictors of share dynamics on the regional level. To try and capture some of these region specific dynamics, let us look at what happens when the lagged value of the region specific share $X_{krs,t-2}$ is added to the model. Perhaps this enables us to obtain an even better model. The resulting model (model 5), with an RE assumption, looks then as follows:

$$X_{krst} = \alpha + \theta Q_{ks,t-2} + \psi coverage_{rs,t-2} + \beta X_{krs,t-2} + \eta_{krs} + \epsilon_{krst} \quad \text{(random effects)} \tag{10}$$

The fourth column in Table 12 provides the estimation results of estimating this model with the RE estimator. Firstly, note that the number of observations has decreased quite drastically because of adding the lagged regional shares to the model. This is thus a downside of using this extended model. Also note that the estimated variance of the individual effects $\hat{\sigma}_\eta^2$ is equal to zero. This should always alert one of the possibility that there are no individual effects present in the model in the first place. If this is indeed the case, the RE estimator will no longer be efficient (although it will still be consistent). To formally test the null hypothesis of no individual effects, a Breusch-Pagan Lagrange multiplier test has been used (Wooldridge, 2010). The resulting p-value is equal to one. This means that we cannot reject the null hypothesis of no individual effects. Therefore, it is chosen to omit these from the model. This leaves us with the following model (model 6):

$$X_{krst} = \alpha + \theta Q_{ks,t-2} + \psi coverage_{rs,t-2} + \beta X_{krs,t-2} + \epsilon_{krst} \tag{11}$$

This model can be estimated consistently and efficiently using the POLS estimator. The estimation results from doing so, are presented in the fifth column of Table 12. It can be seen that all three explanatory variable are highly significant. The lagged value of the regional shares thus seems to provide some added value to the model. Moreover, it can be seen that the estimated coefficient for θ and β are both close to 0.5.

Therefore, the model's prediction becomes kind of an average of the past regional share and the past national share. This is a nice intuitive interpretation. To take this one step further, it was chosen to test the hypothesis that the sum of the coefficients for the past regional and national share sum up to one (i.e. testing that $\theta + \beta = 1$). The resulting p-value can be found in the last row of the table. It can be seen that the hypothesis is strongly rejected for both model 5 and model 6 (for the moment one can ignore the model in the last column of the table). Therefore, it is best to keep the model as it is, and not imposes this linear restriction when estimating it. Furthermore, it can be seen from the table that the overall R-squared has slightly increased as compared to the model only including the national share. Perhaps the model is now better able to capture some of the region specific dynamics. However, purely based on these numbers one cannot simply say whether model 4 or model 6 is the `best' forecasting model. As indicated, a major advantage of model 4 is that is estimated using roughly 16000 more shares than model 6. It is therefore likely to be more representative of the entire distribution of shares. So even though model 6 has a slightly better fit, it is probably more biased towards shares based on a relatively large number of people. Another benefit of using model 4 as a forecasting model, is that it only requires the national share in the previous period and the coverage in the previous period as inputs, but no regional shares. This would make model 4 more practical to use as a forecasting model, given the substantial number of missing regional shares. To more formally evaluate both models' forecasting performance, the next subsection compares both models' out-of-sample forecasting performance.

Evaluating Forecasts

To provide us with some idea of to what extent modelling the regional sectoral occupational shares provides added value in making forecasts for regional sectoral occupational shares, this section compares the out-of-sample forecasting performance of the two models developed in Section 9.1.1. with the out-of-sample forecasting performance of a `naive' model which mimics POA's forecasting behaviour. As indicated, ROA forecasts regional shares simply by forecasting national shares and assuming regional shares are equal to national shares. In order to be able to mimic this behaviour, a simple auto-regressive model (model ROA) of order one is estimated for the national shares:

$$Q_{kst} = \alpha + \omega Q_{ks,t-2} + \epsilon_{kst} \quad (12)$$

Note that a Breusch-Pagan Lagrange multiplier test has been performed to check for the presence of individual effects. The results of this test have suggested that it is probably better to not include individual effects, hence the resulting specification. This also implies that the model can be estimated using POLS. The estimation results are displayed in Table 13. The R-squared of 0.99 indicates that a model including the lagged value of the national shares is able to explain almost all variation in the national shares. Even though in practice ROA uses a more advanced forecasting model to predict national shares, the R-squared of 0.99 indicates that it is not likely that ROA's model has a much better forecasting performance. Therefore, this simple AR(1) model is deemed to properly represent the potential value of forecasting regional shares by modeling national shares. In order to calculate the fitted (predicted) value of a certain regional share $\hat{X}_{krst}$, the only step that remains is to set these regional fitted values equal to the corresponding national fitted values (i.e.$\hat{X}_{krst} = \hat{Q}_{kst}$).

Table 13: Estimation results national occupation share modelling.

	MODEL ROA *(POLS)*
Parameter estimates	
$Q_{ks,t-2}$	0.981**
	(0.0043)
Regression diagnostics	
R-squared	0.992
$\hat{\sigma}_{\epsilon}$	0.006
N	6103
Number of groups	797

Clustered standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Source: own calculations based on EBB (Statistics Netherlands).

To evaluate how well these three different models are able to predict regional sectoral occupational shares, Table 14 displays their out-of-sample forecasting performance for the period 2016. This implies that the models have been estimated using data for the period 1996-2014 only. The chosen performance metric to evaluate the models' performance is the Mean Absolute Error (MAE), which is nothing else than the average absolute value of the prediction error. Note furthermore that as each model requires different inputs and the dataset is not complete, the number of prediction instances varies per model. To facilitate a fair comparison between the models, it was chosen to calculate the MAE only over those observations which had a corresponding prediction from all three models. The total number of predicted regional shares is then equal to 6043.

Table 14: Out-of-sample forecasting performance for t=2016 for three candidate forecasting models - occupations.

	MODEL ROA	**MODEL 4**	**MODEL 6**
MAE	0.0209	0.0206	0.0192

Source: own calculations based on EBB (Statistics Netherlands).

The table indicates that model 4 has almost no added value as compared to a model based on national shares only. Model 6 however does seem to yield slightly better forecasts. On average, the forecasts resulting from this model are around 0.2 percentage points closer to target as compared to the proxy for ROA's model. This indicates that modelling the regional occupational shares themselves can potentially slightly improve one's forecasts for the regional occupational shares. If one would do so by using model 6, one can of course only make a forecast for those shares for which the value in the previous period is observed. It could then for example be a possibility to use MODEL ROA to predict the shares for which the previous value is not observed. In this way one exploits the information that is present in the regional shares as optimally as possible, while still obtaining complete forecasts.

9.2 Forecasting Regional Sectoral Educational Shares

Model Development

This subsection is occupied with modelling regional sectoral educational shares. In a similar fashion to the occupational share modelling, let us start by estimating model 3 using the RE estimator. The results are displayed in the second column of Table 15. The estimated variance of the individual effects indicates that it may be better to impose a fixed effects assumption on the model. Therefore, like before, a Hausman test has been employed to test the random effects assumption. The p-value resulting from this test is zero. This provides substantial evidence against the random effects assumption. Therefore, it was chosen instead to estimate model 4 using the WI estimator. The results from this estimation are displayed in the third column of Table 15. It can be seen that the lagged value of the national share is highly significant. Moreover, the overall R-squared indicates that by using the lagged value of the regional educational shares one is already able to explain a large portion of the variance in the shares. Also note that the coverage is again highly significant and has the expected negative sign.

Let us now examine the added value of adding the lagged value of the regional share to the model. The fourth column of Table 15 shows the estimation results of estimating model 5 for the educational shares, using the RE estimator. Similarly to before, the estimated variance of the individual effects puts doubt on the presence of individual effects in the model. To test for the presence of individual effects, a Breusch-Pagan Lagrange multiplier test is performed. The p-value which results is 1. Hence, clearly the absence of individual effects in model 5 (for the educational shares) cannot be rejected. Therefore, it is chosen to omit these individual effects from the model and instead estimate model 6 using POLS.

The estimation results of this estimation can be found in fifth column of Table 15. It can be seen that also for the educational shares, the lagged value of the regional share enters the model highly significantly. Moreover, both the lagged value of the regional share and the lagged value of the national share are again close to 0.5 in value. The last row of the table shows the p-value which results from testing the null hypothesis that the sum of the coefficients of the lagged regional and national share is equal to one, i.e. the hypothesis that $\theta + \beta = 1$. This hypothesis is clearly rejected for both the estimations of model 5 and model 6 (for the moment one can ignore the estimation in the last column). Hence, it is chosen to keep the model as it is, and not impose this linear restriction when estimating it. Furthermore, note that the model's fit as measured by the R-squared has increased quite significantly compared to model 4.

Table 15: Estimation results regional education share modelling.

	MODEL 3 *(RE)*	MODEL 4 *(WI)*	MODEL 5 *(RE)*	MODEL 6 *(POLS)*	MODEL 6 - Sample subset *(POLS)*
Parameter estimates					
$Q_{ks,t-2}$	0.902**	0.766**	0.526**	0.526**	0.495**
	(0.0068)	(0.0126)	(0.0094)	(0.0094)	(0.0109)
$coverage_{rs,t-2}$	-0.039**	-0.022**	-0.022**	-0.022**	-0.012**
	(0.0011)	(0.0012)	(0.0016)	(0.0016)	(0.0017)
$X_{krs,t-2}$			0.432**	0.432**	0.472**
			(0.0092)	(0.0092)	(0.0107)
Regression diagnostics					
R-squared overall	0.764	0.762	0.812	0.812	0.849
R-squared within	0.142	0.144	0.095		
R-squared between	0.842	0.832	0.932		
$\hat{\sigma}_\eta$	0.012	0.021	0		
$\hat{\sigma}_\epsilon$	0.023	0.023	0.024	0.026	0.024
$\hat{\rho}$	0.211	0.451	0		
N	87304	87304	67841	67841	59363
Number of groups	15698	15698	11976	11967	10016
Hypothesis test					
H_0: $\theta + \beta = 1$, Resulting p-value			0	0	0

Robust standard errors in parentheses,
clustered standard errors for MODEL 6.
* $p < 0.05$, ** $p < 0.01$

Source: own calculations based on EBB (Statistics Netherlands).

Model 6 would thus seem to be the more appropriate forecasting model. However, using model 4 to forecast regional shares has some advantages relative to using model 6, as discussed in the context of the occupational share modelling. Therefore, just like for the occupational shares, it is chosen to compare ROA's forecasting model's performance to the performance of both model 4 and model 6. This will be done in the next subsection. A last thing which is interesting to note, is that the optimal fit resulting from modelling the educational shares is substantially lower than the optimal fit resulting from modelling the occupational shares (0.935 as compared to 0.812). Apparently, the educational shares are less easy to predict by such models.

Evaluating Forecasts

To examine how much added value modelling regional sectoral educational shares has in predicting their value, this section compares the out-of-sample forecasting results of the regional models (i.e. model 4 and model 6) with the performance of a model which is estimated exclusively using national shares (i.e. in the style of the current forecasting model used in POA). The procedure used to do so is exactly the same as the procedure used to do this for the occupational share models.

The estimation results of estimating MODEL ROA using POLS for the educational shares are displayed in Table 16. Note that again a Breusch-Pagan Lagrange multiplier test has been performed to check for the presence of individual effects. The results of this test have suggested that it is probably better to omit the individual effects from the model, hence the resulting specification. Furthermore, note that even though in practice ROA uses a more advanced forecasting model to predict national educational shares, the R-squared of 0.98 indicates that it is not likely that ROA's model has a much better forecasting performance. Therefore, this simple AR(1) model is deemed to properly represent the potential value of forecasting regional shares by modeling national shares.

Table 16: Estimation results national education share modelling.

	MODEL ROA *(POLS)*
Parameter estimates	
$Q_{ks,t-2}$	0.982**
	(0.0050)
Regression diagnostics	
R-squared	0.979**
$\hat{\sigma}_{\epsilon}$	0.006
N	6066
Number of groups	712

Clustered standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Source: own calculations based on EBB (Statistics Netherlands).

To evaluate how well the three different models are able to predict regional sectoral educational shares, Table 17 displays their out-of-sample forecasting performance for the period 2016. Note that the value for the MAE is calculated over 7379 prediction instances. The results indicate that model 6 yields slightly better forecasts as compared to MODEL ROA and model 4. On average, the forecasts resulting from this model are around 0.2 percentage points closer to target as compared to the forecasts resulting from the other two models. It can thus be said that also for educational structures there seems to be some added value in using a prediction model based on regional shares, at least if one uses model 6. Note that model 4 now even performs slightly worse than MODEL ROA.

Table 17: Out-of-sample forecasting performance for t=2016 for three candidate forecasting models - education.

	MODEL ROA	MODEL 4	MODEL 6
MAE	0.0201	0.0206	0.0183

Source: own calculations based on EBB (Statistics Netherlands).

9.3 Further Improvements to the Forecasting Model

Given that there seems to lie some added value in forecasting regional sectoral occupational and educational shares by modelling the regional shares themselves (more specifically by using model 6), it is interesting to see whether model 6 can still be improved further. To get an idea of how this could possibly be done, it is useful to have a closer look at the residuals resulting from estimating model 6 using POLS. These can give us an idea of which type of observations are difficult to predict by the model.

Tables 18 and 19 display respectively the average absolute value of the residuals per sector resulting from estimating model 6 for the occupational shares and the educational shares. The most striking observation from these tables is that the sectors with the worst fit are roughly the same nine sectors which had been partially excluded from the analysis earlier on in this paper as a result of insufficient coverage. Why would precisely these sectors have the worst fit? A straightforward explanation would be that for these sectors in many regions relatively small amounts of people are observed for relatively large shares, which naturally results in more sampling variability of the corresponding shares, thereby making these shares more difficult to predict. Some evidence for this statement has already been given in the methodology section, by observing the strong negative correlation between the average deviation and the coverage (AD_{rst} and $coverage_{rst}$ respectively). Therefore, let us examine what happens to the models' fit if for these nine sectors the same regions are excluded as before (i.e. only regions with a sufficiently high coverage are included). Dropping these region/sector combination causes us to loose around ten percent of the sample for both the occupational and educational samples. Also note that this selection implies again that the energy sector is dropped from the sample in total. Lastly, note that essentially these tables are indicating the presence of heteroskedasticity in the data. It shows us that probably not for all sampled units the variance of the error term is the same. To ensure that the reported hypothesis tests are still reliable, it has therefore been chosen to estimate all models in this section (i.e. also the models in the previous subsections) using robust standard errors. Also the estimation on the subset of the sample which follows now is constructed using robust standard errors, even though the panel units with a supposedly larger variance of the error term are taken out in this sample.

Table 18: Average absolute residual per sector from estimating model 6 using POLS - occupation.

Sector	Average absolute residual
Agriculture	0.038
Energy	0.037
Culture, sports and recreation	0.034
Other services	0.032
Food and stimulants industry	0.031
Chemical industry	0.030
Catering	0.029
Financial services	0.025
Information and communication	0.024
Transport and storage	0.020
Rental and other business services	0.020
Metal industry	0.019
Other industry	0.018
Wholesale	0.018
Healthcare	0.017
Construction	0.017
Specialist business services	0.015
Well-being	0.015
Public governance	0.015
Education	0.013
Retail	0.011

Source: own calculations based on EBB (Statistics Netherlands).

The estimation results from estimating model 6 using the POLS estimator on the subset of the sample are displayed in the sixth column of the Tables 12 (for occupational shares) and 15 (for educational shares). Note that again a Breusch-Pagan Lagrange multiplier test has been used to test for the presence of individual effects. For both the occupation and the education model the resulting p-value is one. This suggests that it is better to omit the individual effects from the models. Therefore, the estimation results are simply obtained using POLS. From the tables it can be seen that the estimated coefficients and significance of them does not change much as compared to the estimations using the full sample. Furthermore, it can also be seen that the fit of both the occupation and the education model increases slightly as compared to the full sample model. This increase is approximately one percentage point for the occupation model and four percentages points for the education model. Especially for the education model, the increase in fit is quite substantial. Of course, by simply omitting these sector/region combinations from the model one is not really solving a problem. The respective shares still need to be predicted. However, the results presented in this section indicate that it may be better to not estimate these shares using a prediction model based on regional shares. Instead, for these shares it may be better to e.g. simply use MODEL ROA as a forecasting model. This is a question that remains for future research.

Table 19: Average absolute residual per sector from estimating model 6 using POLS - education.

Sector	**Average absolute residual**
Energy	0.034
Culture, sports and recreation	0.030
Other services	0.026
Chemical industry	0.025
Information and communication	0.024
Agriculture	0.023
Food and stimulants industry	0.023
Financial services	0.023
Catering	0.021
Metal industry	0.020
Other industry	0.019
Healthcare	0.017
Rental and other business services	0.017
Construction	0.017
Transport and storage	0.017
Wholesale	0.016
Education	0.016
Specialist business services	0.016
Well-being	0.014
Public governance	0.014
Retail	0.011

Source: own calculations based on EBB (Statistics Netherlands).

10 Conclusion and consequences for the regional forecast

10.1 Conclusions

This research examines heterogeneity in regional sectoral occupational and educational structures in the Netherlands. It is found that, on average, current regional sectoral occupational distributions differ at least 23 percentage points from the corresponding national sectoral occupational distribution as measured by the sum of absolute values of individual shares' deviations. This implies that for a given combination of region and sector, on average, one would currently have to reallocate at least 11.5% of the workforce to a different occupation in order to obtain the national occupational structure for that sector. Furthermore, the current cross-sector average deviation (in absolute value) of the observed occupations' regional sectoral share from their corresponding national sectoral share is equal to 2.2. For most individual sectors this average deviation lies close to 2.2, implying that the observed differences in the total percentage point deviation (i.e. the differences in $\overline{KSI_{.st}}$) across sectors are mainly due to certain sectors having more heterogeneous occupational structures in terms of number of relevant occupations, instead of individual observed occupations within those sectors having larger deviations on average. Furthermore, this value gives us an indication of how large deviations are at the extremes of the total unobserved distribution of occupational deviations for each sector. Looking at current average deviations from the national share for individual occupations across regions, it turns out that there are only a few combinations of occupation and sector for which there seems to be relatively much heterogeneity across regions. These are vehicle drivers and mobile machine operators for the transport and storage sector, writers and artists for the culture, sports and recreation sector, and personal services employees for the other services sector. Most other combinations' deviations lie relatively close to the average of 2.2. Lastly, it is found that heterogeneity in regional sectoral occupational structures as measured by the KSI_{rst} has been increasing at a rate of 0.23 percentage points every year. This hints at the existence of structural σ-divergence in regional sectoral occupational structures. Further analysis has shown that this observed trend in divergence is mainly driven by divergence in just a few regions and sectors. Whether these trends will continue in the future remains to be seen.

With regard to heterogeneity in regional sectoral educational structures, it is found that, on average, regional structures currently deviate at least 27 percentage points from their national counterpart as measured by the sum of absolute values of individual shares' deviations. This means that for a specific combination of region and sector, on average, currently at least 13.5% of the regional sectoral workforce would have to have a different educational background in order to obtain the educational composition of the corresponding national sectoral workforce. Furthermore, the current cross-sector average deviation of the observed educations' regional sectoral shares from their corresponding national sectoral shares is equal to 2.0. For most individual sectors this average deviation lies close 2.0, implying that the observed differences in the total percentage point deviation (i.e. the differences in $\overline{KSI_{.st}}$) across sectors are mainly due to those sectors having more heterogeneous educational structures in terms of number of relevant types of education, instead of individual types of education within those sectors having larger deviations on average. Furthermore, this value gives us an indication of how large deviations are at the extremes of the total unobserved distribution of educational deviations for each sector. Looking at current average deviations from the national share for specific types of education across regions, it turns out that there are only a few combinations of education and sector for which there seems to be relatively much heterogeneity across regions. It concerns the combinations mbo4 - technical and ict for the chemical industry, master - economics and society for the financial services sector, master - economics and society for the specialist business services sector, and bachelor - economics and society for the culture, sports and recreation sector. Most other combinations' deviations lie relatively close to the average of 2.0. Lastly, it is found that heterogeneity in regional sectoral educational structures as measured by the KSI_{rst} has been increasing at a rate of 0.26 percentage points every year. This points towards the existence of structural σ-divergence in regional sectoral educational distributions. Further analysis has shown that such a trend in divergence is prevalent in almost all regions and sectors. It remains to be seen though whether these trends in divergence will continue in the future. Note that the most important conclusions from the heterogeneity analysis for both occupational and educational structures are recapitulated in Table 20.

An important remark in the context of these findings for heterogeneity in regional occupational and educational structures in the Netherlands as measured by the KSI, is that the observed heterogeneity strongly depends on the type of classification one uses (and directly relevant to this paper, thus also the type of aggregation level). The more detailed the aggregation level which is chosen, the higher the KSI values will be (conditional on observing the full dataset). The observed heterogeneity for occupational segments can thus be seen as a lower bound for heterogeneity on the level of occupational groups. A similar argument applies to the educational aggregation levels. However, we also need to remark that the goal of this study is to research whether the assumption of equal regional occupational and educational composition holds, and this needs to be tested at the same level of detail as the one at which the regional labour market forecasts are computed. Given data constraints, the forecasts are computed at an intermediate level of detail.

Furthermore, this paper has done an exploratory analysis of potential ways in which we can improve the POA forecasts of regional sectoral occupational and educational distributions. More specifically, we looked at the potential added value of using data on the regional structures to forecast the regional structures, as compared to only using data on the national level to forecast regional structures. It is found that using regional data can improve the forecasting accuracy of regional sectoral shares with around 0.2 percentage points on average, for both occupational and educational shares. Given the large number of shares that need to be forecasted, this seems to be too much of an improvement to simply ignore. It should be said though, that the newly developed models are not compared to the actual forecasting model used in POA, but instead to a proxy for the forecasting model in POA. Furthermore, ROA normally forecasts yearly shares and not two-yearly shares and they also forecast more than one period ahead. Therefore, the forecasting analysis in this paper is merely meant to give a first indication of the added value of modelling regional shares. A more thorough analysis is still recommended.

The conclusions presented in this paper are only valid for the occupational and educational shares which are observed. Shares based on less than ten observations are omitted from the data and it therefore remains questionable to what extent these conclusions can be extrapolated to the unobserved part of the data. However, this paper has made significant efforts to indicate as well as possible how representative the available data are of the whole. Moreover, there is no alternative dataset on Dutch workers' occupation and education available. Therefore, when analysing these kind of questions, there is no other option than using this data.

A last finding from this research which is worth pointing out, is the relatively high heterogeneity which is observed for the occupational and educational regional sectoral shares belonging to region/sector/period combinations with a low coverage. It has been shown that there is a strong negative correlation between a regional share's deviation from the national share and the coverage of the corresponding region/sector/period combination. This observation is likely due to the fact that for such region/sector/period combinations, even relatively large shares are based on a small number of observations. Therefore, small changes in the number of observed people can cause relatively large deviations in share values. In other words, such shares are likely to be characterised by a large sampling variability. This `unpredictability' of such shares is confirmed by the findings presented in Section nine of this paper, where these appeared to be the shares which were most difficult to forecast. This information sheds some extra light on the EBB data.

As with any research paper, there are still aspects of the topic which could not be covered. Most importantly, this paper has put little effort in explaining the observed heterogeneity in regional sectoral occupational and educational structures. Instead, it has primarily focused on how to measure such heterogeneity in the first place. To be able to really grasp the economic mechanisms generating the observed heterogeneity, one has to ask oneself the question why certain regions or sectors show more heterogeneity than other sectors.

10.2 Implications for ROA’s regional forecast models

In this last section, we consider the implications of the results presented in this research for the regional labour market forecasts of ROA. The findings in this paper show there is some added value in including information on regional specific sectoral educational and occupational structures. However, implementing regional data on these structures in the forecasting models in practice is not a very straightforward task, especially in the presence of missing data.

Generally, at the level of educational (and possible occupational) aggregation at which the regional labour market forecasts are produced, we find that the regional differences in sectoral educational composition are not very large and only change very slowly over time. So based on this finding, using the national sectoral educational and occupational composition for the regional labour market forecasts is a practical working assumption that can be continued. A second finding in this research is that the accuracy of the regional labour market forecasts can be increased if information on regional sectoral structures would be included in the forecasts in addition to the information on the national sectoral structures. So based on these findings, ROA could improve the regional forecasts by taking regional differences in sectoral educational and occupational structures into consideration.

However, as is shown at various points in this research, when using LFS data as the only source to include sector of activity, occupation, and educational background, missing data is a big issue when including additional regional sectoral information in the forecasting models. Regional sectoral occupational and educational information is only available and reliable for larger sectors with large shares of workers in related occupational segments and related educational sectors. This research shows that if this is the case, the forecasts can be improved by including this information. However, this information on regional sectoral and educational structures is not available for all sectors and all regions in the times series of LFS, especially for smaller sectors and regions with a smaller residential population. So the trade-off for improving the forecasts would be either to improve only the forecasts for some educational sectors (or occupational segments) in some regions, or not improve the forecasts at all.

Given the level of aggregation in the ROA forecasts by educational background, the problem of missing data is smaller than in this research, and the regional deviations from the national sectoral educational and occupational structure will also be smaller than in this research. In addition, the ROA forecasting model is a supply and demand model in which the whole labour market is included to adequately depict inter-sectoral dependencies and labour market mobility and substitution by educations and occupations. Excluding small sectors, educational groups or occupational groups would compromise this model. Based on these arguments, for now, the loss of information by including regional sectoral differences in the forecasts will be too high compared to the gains of including this information. However, the growing availability of administrative data about educational attainments covering the whole population could potentially help improve the accuracy of the modelling in near the future.

Table 20: Main conclusions heterogeneity analysis regional sectoral occupational and educational structures (values x100).

		Occupational structures
Current heterogeneity	• $KSI_{.st}$	
	* Average across sectors:	23
	* Range across sectors:	13 - 32
	* Sectors at low end of distribution:	retail, catering, education
	* Sectors at high end of distribution:	culture, sports, and recreation; other services
	• $\overline{AD_{.st}}$	
	* Average across sectors:	2.2
	* Range across sectors:	1.4 - 3.0[a]
	• Occupation and sector combinations with relatively much heterogeneity:	*gardeners, arable farmers and cattle breeders* - agriculture *vehicle drivers and mobile machine operators* - transport and storage *writers and artists* - culture, sports and recreation *personal services employees* - other services
Heterogeneity over time	• Estimated yearly KSI_{rst} increase	0.23
		Educational structures
Current heterogeneity	• $\overline{KSI_{.st}}$	
	* Average across sectors:	27
	* Range across sectors:	21 - 32
	* Sectors at low end of distribution:	construction, retail, catering, well-being
	* Sectors at high end of distribution:	chemical industry; culture, sports, and recreation; other services
	• $\overline{AD_{.st}}$	
	* Average across sectors:	2.0
	* Range across sectors:	1.4 - 2.8
	• Education and sector combinations with relatively much heterogeneity:	*mbo4-technical and ict* - chemical industry *master-economics and society* - financial services *master-economics and society* - specialist business services *bachelor-economics and society* - culture, sports and recreation
Heterogeneity over time	• Estimated yearly KSI_{rst} increase	0.26

Source: own calculations based on EBB (Statistics Netherlands).
[a]The only exception being the agricultural sector with an $\overline{AD_{st}}$ of 3.9.

References

Barbour, E. and Markusen, A. (2007). Regional occupational and industrial structure: does one imply the other? International Regional Science Review, 30(1):72-90.

Barro, R. J. and Sala-i Martin, X. (1992). Convergence. Journal of political Economy, 100(2):223-251.

Barro, R. J. and Sala-i Martin, X. (1993). Galtons fallacy and the convergence hypothesis. Scandinavian Journal of Economics, 95(1):427-443.

Belke, A. and Heine, J. M. (2006). Specialisation patterns and the synchronicity of regional employment cycles in europe. International Economics and Economic Policy, 3(2):91-104.

Combes, P.-P. and Overman, H. G. (2004). The spatial distribution of economic activities in the european union. In Handbook of regional and urban economics, volume 4, pages 2845-2909. Elsevier.

Cörvers, F. and Meriküll, J. (2007). Occupational structures across 25 eu countries: the importance of industry structure and technology in old and new eu countries. Economic Change and Restructuring, 40(4):327-359.

Dixon, R., Freebairn, J., et al. (2009). Trends in regional specialisation in Australia. Australasian Journal of Regional Studies, The, 15(3):281.

Esteban, J. (2000). Regional convergence in europe and the industry mix: a shift-share analysis. Regional science and urban economics, 30(3):353-364.

Friedman, M. (1992). Do old fallacies ever die?

Glaeser, E. L. (2010). Agglomeration economics. University of Chicago Press.

De Groot (2015). Arbeids- en Woningmarktdynamiek, Platform31 Essay 6.

Höhenberger, N. and Schmiedeberg, C. (2008). Structural convergence of european countries. Technical report, CeGE Discussion Paper.

Hoover, E. M. (1948). Location of economic activity. McGraw-Hill Book Company, Inc., New York.

Krugman, P. (1991). Increasing returns and economic geography. Journal of political economy, 99(3):483-499.

Krugman, P. (1998). What's new about the new economic geography? Oxford review of economic policy, 14(2):7-17.

Krugman, P. R. (1993). Geography and trade. MIT press.

Leontief, W. et al. (1953). Domestic production and foreign trade; the american capital position re-examined.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. The quarterly journal of economics, 107(2):407-437.

Marelli, E. (2007). Specialisation and convergence of european regions. The European Journal of Comparative Economics, 4(2):149.

Marshall, A. (1961). Principles of economics: An introductory volume. Macmillan London.

Midelfart-Knarvik, K. H. and Overman, H. G. (2002). Delocation and european integration: is structural spending justified? Economic policy, 17(35):321-359.

Ohlin, B. (1935). Interregional and international trade. Harvard University Press, Cambridge.

Onderwijs in Cijfers (2019). Hoogst behaald opleidingsniveau. https://www.onderwijsincijfers.nl/kengetallen/onderwijs-algemeen/hoogst-behaald-opleidingsniveau. Accessed: 13-06-2019.

Palan, N. (2010). Measurement of specialization the choice of indices. Technical report, FIW working paper.

Ricardo, D. (1821). On the principles of political economy. J. Murray.

Scherer, C. E. M. and Folch, D. C. (2017). A comparative study of urban occupational structure:Brazil and united states.

Statistics Netherlands (2019a). CBS Statline regionale kerncijfers; nationale rekeningen 2010-2016. https://statline.cbs.nl/Statweb/. Accessed: 09-06-2019.

Statistics Netherlands (2019b). Meer hoogopgeleiden in alle beroepsklassen. https://www.cbs. nl/nl-nl/nieuws/2017/36/meer-hoogopgeleiden-in-alle-beroepsklassen. Accessed: 09-06-2019.

Suedekum, J. (2006). Concentration and specialization trends in germany since re-unification. Regional Studies, 40(8):861-873.

Theil, H. (1967). Economics and information theory. Technical report.

Verkade and Bakens (2020), Commuter Flow Predictions in POA: Evaluation Study, ROA Technical Report 005.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.

Young, A. T., Higgins, M. J., and Levy, D. (2008). Sigma convergence versus beta convergence: Evidence from us county-level data. Journal of Money, Credit and Banking, 40(5):1083-1093.

Appendix A

Bijlage 1
BRC2014: Beroepsklasse, -segment en -groep

Beroepsklasse	*Beroepssegment*	Beroepsgroep
01 Pedagogische beroepen		
	011 Docenten	
		0111 Docenten hoger onderwijs en hoogleraren
		0112 Docenten beroepsgerichte vakken secundair onderwijs
		0113 Docenten algemene vakken secundair onderwijs
		0114 Leerkrachten basisonderwijs
		0115 Onderwijskundigen en overige docenten
	012 Sportinstructeurs	
		0121 Sportinstructeurs
	013 Leidsters kinderopvang en onderwijsassistenten	
		0131 Leidsters kinderopvang en onderwijsassistenten
02 Creatieve en taalkundige beroepen		
	021 Auteurs en kunstenaars	
		0211 Bibliothecarissen en conservatoren
		0212 Auteurs en taalkundigen
		0213 Journalisten
		0214 Beeldend kunstenaars
		0215 Uitvoerend kunstenaars
	022 Vakspecialisten op artistiek en cultureel gebied	
		0221 Grafisch vormgevers en productontwerpers
		0222 Fotografen en interieurontwerpers
03 Commerciële beroepen		
	031 Adviseurs marketing, public relations en sales	
		0311 Adviseurs marketing, public relations en sales
	032 Vertegenwoordigers en inkopers	
		0321 Vertegenwoordigers en inkopers
	033 Verkopers	
		0331 Winkeliers en teamleiders detailhandel
		0332 Verkoopmedewerkers detailhandel
		0333 Kassamedewerkers
		0334 Callcentermedewerkers outbound en overige verkopers
04 Bedrijfseconomische en administratieve beroepen		
	041 Specialisten bedrijfsbeheer en administratie	
		0411 Accountants
		0412 Financieel specialisten en economen
		0413 Bedrijfskundigen en organisatieadviseurs
		0414 Beleidsadviseurs
		0415 Specialisten personeels- en loopbaanontwikkeling
	042 Vakspecialisten bedrijfsbeheer en administratie	
		0421 Boekhouders
		0422 Zakelijke dienstverleners
		0423 Directiesecretaresses
	043 Administratief personeel	
		0431 Administratief medewerkers
		0432 Secretaresses
		0433 Receptionisten en telefonisten
		0434 Boekhoudkundig medewerkers
		0435 Transportplanners en logistiek medewerkers
05 Managers		
	051 Algemeen directeuren	
		0511 Algemeen directeuren
	052 Managers op administratief en commercieel gebied	
		0521 Managers zakelijke en administratieve dienstverlening
		0522 Managers verkoop en marketing
	053 Managers productie en gespecialiseerde dienstverlening	
		0531 Managers productie
		0532 Managers logistiek
		0533 Managers ICT

Figure 13: Occupational classification BRC2014 (part 1). *Source: ROA website:* `http://roa.sbe.maastrichtuniversity.nl/?portfolio=poa-project-onderwijs-arbeidsmarkt-2`.

Beroepsklasse	*Beroepssegment*	Beroepsgroep
		0534 Managers zorginstellingen
		0535 Managers onderwijs
		0536 Managers gespecialiseerde dienstverlening
	054 Managers horeca, detailhandel en overige diensten	
		0541 Managers horeca
		0542 Managers detail- en groothandel
		0543 Managers commerciële en persoonlijke dienstverlening
	055 Managers z.n.d.	
		0551 Managers z.n.d.
06 Openbaar bestuur, veiligheid en juridische beroepen		
	061 Overheidsambtenaren en -bestuurders	
		0611 Overheidsbestuurders
		0612 Overheidsambtenaren
	062 Juristen	
		0621 Juristen
	063 Beveiligingswerkers	
		0631 Politie-inspecteurs
		0632 Politie en brandweer
		0633 Beveiligingspersoneel
		0634 Militaire beroepen
07 Technische beroepen		
	071 Ingenieurs en onderzoekers wis-, natuur- en technische wetenschappen	
		0711 Biologen en natuurwetenschappers
		0712 Ingenieurs (geen elektrotechniek)
		0713 Elektrotechnisch ingenieurs
		0714 Architecten
	072 Vakspecialisten natuur en techniek	
		0721 Technici bouwkunde en natuur
		0722 Productieleiders industrie en bouw
		0723 Procesoperators
	073 Bouwarbeiders	
		0731 Bouwarbeiders ruwbouw
		0732 Timmerlieden
		0733 Bouwarbeiders afbouw
		0734 Loodgieters en pijpfitters
		0735 Schilders en metaalspuiters
	074 Metaalarbeiders, machinemonteurs	
		0741 Metaalbewerkers en constructiewerkers
		0742 Lassers en plaatwerkers
		0743 Automonteurs
		0744 Machinemonteurs
	075 Voedselverwerkende beroepen en overige ambachten	
		0751 Slagers
		0752 Bakkers
		0753 Productcontroleurs
		0754 Meubelmakers, kleermakers en stoffeerders
		0755 Medewerkers drukkerij en kunstnijverheid
	076 Elektriciens en elektronicamonteurs	
		0761 Elektriciens en elektronicamonteurs
	077 Productiemachinebedieners en assemblagemedewerkers	
		0771 Productiemachinebedieners
		0772 Assemblagemedewerkers
	078 Hulpkrachten bouw en industrie	
		0781 Hulpkrachten bouw en industrie
08 ICT beroepen		
	081 Specialisten ICT	
		0811 Software- en applicatieontwikkelaars
		0812 Databank- en netwerkspecialisten
	082 Vakspecialisten ICT	
		0821 Gebruikersondersteuning ICT

Figure 14: Occupational classification BRC2014 (Part 2). *Source: ROA website:* `http://roa.sbe.maastrichtuniversity.nl/?portfolio=poa-project-onderwijs-arbeidsmarkt-2`.

Beroepsklasse	*Beroepssegment*	Beroepsgroep
		0822 Radio- en televisietechnici
09 Agrarische beroepen		
	091 Tuinders, akkerbouwers en veetelers	
		0911 Land- en bosbouwers
		0912 Hoveniers, tuinders en kwekers
		0913 Veetelers
	092 Hulpkrachten landbouw	
		0921 Hulpkrachten landbouw
10 Zorg en welzijn beroepen		
	101 Artsen, therapeuten en gespecialiseerd verpleegkundigen	
		1011 Artsen
		1012 Gespecialiseerd verpleegkundigen
		1013 Fysiotherapeuten
	102 Specialisten op maatschappelijk gebied	
		1021 Maatschappelijk werkers
		1022 Psychologen en sociologen
	103 Vakspecialisten gezondheidszorg	
		1031 Laboranten
		1032 Apothekersassistenten
		1033 Verpleegkundigen (mbo)
		1034 Medisch praktijkassistenten
		1035 Medisch vakspecialisten
	104 Sociaal werkers, groeps- en woonbegeleiders	
		1041 Sociaal werkers, groeps- en woonbegeleiders
	105 Verzorgenden	
		1051 Verzorgenden
11 Dienstverlenende beroepen		
	111 Medewerkers persoonlijke dienstverlening	
		1111 Reisbegeleiders
		1112 Koks
		1113 Kelners en barpersoneel
		1114 Kappers en schoonheidsspecialisten
		1115 Conciërges en teamleiders schoonmaak
		1116 Verleners van overige persoonlijke diensten
	112 Schoonmakers en keukenhulpen	
		1121 Schoonmakers
		1122 Keukenhulpen
12 Transport en logistiek beroepen		
	121 Bestuurders voertuigen en bedieners mobiele machines	
		1211 Dekofficieren en piloten
		1212 Chauffeurs auto's, taxi's en bestelwagens
		1213 Buschauffeurs en trambestuurders
		1214 Vrachtwagenchauffeurs
		1215 Bedieners mobiele machines
	122 Hulpkrachten transport en logistiek	
		1221 Laders, lossers en vakkenvullers
		1222 Vuilnisophalers en dagbladenbezorgers
13 Overig		
	131 Overig	
		1311 Overig
		1311 Overig

Figure 15: Occupational classification BRC2014 (part 3). *Source: ROA website:* `http://roa.sbe.maastrichtuniversity.nl/?portfolio=poa-project-onderwijs-arbeidsmarkt-2`.

	Original class	Translation
1	Pedagogische beroepen	Pedagogical occupations
2	Creatieve en taalkundige beroepen	Creative and linguistic occupations
3	Commerciële beroepen	Commercial occupations
4	Bedrijfseconomische en administratieve beroepen	Business economical and administrative occupations
5	Managers	Managers
6	Openbaar bestuur, veiligheid en juridische beroepen	Public governance, security and legal occupations
7	Technische beroepen	Technical occupations
8	Ict beroepen	Ict occupations
9	Agrarische beroepen	Agricultural occupations
10	Zorg en welzijn beroepen	Healthcare and well-being occupations
11	Dienstverlenende beroepen	Service occupations
12	Transport en logistiek beroepen	Transport and logistical occupations
13	Overig	Other

Table 21: Occupational classes according to BRC2014 and their translation.

ONRsector ONRsubsector ONRtype

110 Basisonderwijs

- 1100 Basisonderwijs
  - 11000 Basisonderwijs

121 Vmbo-b/k, mbo1 - economie en maatschappij

- 1211 Vmbo-b/k, mbo1 - economie en administratie
  - 12111 Vmbo-b/k, mbo1 - financieel-administratieve dienstverlening
  - 12112 Vmbo-b/k, mbo1 - secretariële dienstverlening
- 1212 Vmbo-b/k, mbo1 - handel, ondernemerschap, transport en logistiek
  - 12121 Vmbo-b/k, mbo1 - detailhandel en groothandel
  - 12122 Vmbo-b/k, mbo1 - transport en logistiek

122 Vmbo-b/k, mbo1 - landbouw en natuur

- 1221 Vmbo-b/k, mbo1 - voedsel, natuur en leefomgeving
  - 12211 Vmbo-b/k, mbo1 - landbouw en dierverzorging
  - 12212 Vmbo-b/k, mbo1 - slagerij, bakkerij, versindustrie
  - 12213 Vmbo-b/k, mbo1 - tuinbouw en groenvoorziening

123 Vmbo-b/k, mbo1 - techniek en ict

- 1231 Vmbo-b/k, mbo1 - ict, media en vormgeving
  - 12311 Vmbo-b/k, mbo1 - ict-ondersteuning
  - 12312 Vmbo-b/k, mbo1 - mediatechniek
- 1232 Vmbo-b/k, mbo1 - techniek, bouw en procesindustrie
  - 12321 Vmbo-b/k, mbo1 - bouw en infra
  - 12322 Vmbo-b/k, mbo1 - elektro- en installatietechniek
  - 12323 Vmbo-b/k, mbo1 - procestechniek en textiel
  - 12324 Vmbo-b/k, mbo1 - werktuigbouwkunde en metaalbewerking

124 Vmbo-b/k, mbo1 - zorg, onderwijs en dienstverlening

- 1241 Vmbo-b/k, mbo1 - dienstverlening
  - 12411 Vmbo-b/k, mbo1 - horeca
  - 12412 Vmbo-b/k, mbo1 - reiniging, sport en toerisme
  - 12413 Vmbo-b/k, mbo1 - schoonheids- en haarverzorging
  - 12414 Vmbo-b/k, mbo1 - veiligheid
- 1242 Vmbo-b/k, mbo1 - gezondheidszorg en welzijn
  - 12421 Vmbo-b/k, mbo1 - helpende zorg en welzijn
  - 12422 Vmbo-b/k, mbo1 - laboratorium en gezondheidstechniek

129 Vmbo-b/k, mbo1 - overig

- 1299 Vmbo-b/k, mbo1 - overig
  - 12999 Vmbo-b/k, mbo1 - overig

130 Vmbo-g/t, havo-, vwo-onderbouw

- 1300 Vmbo-g/t, havo-, vwo-onderbouw
  - 13000 Vmbo-g/t, havo-, vwo-onderbouw

210 Havo, vwo

- 2100 Havo, vwo
  - 21000 Havo, vwo

Figure 16: Educational classification ONR2019 (part 1). *Source: ROA website:* `http://roa.sbe.maastrichtuniversity.nl/?portfolio=poa-project-onderwijs-arbeidsmarkt-2`.

- *221 Mbo2 - economie en maatschappij*
  - 2211 Mbo2 - economie en administratie
    - 22111 Mbo2 - financieel-administratieve dienstverlening
    - 22112 Mbo2 - secretariële dienstverlening
  - 2212 Mbo2 - handel, ondernemerschap, transport en logistiek
    - 22121 Mbo2 - detailhandel en groothandel
    - 22122 Mbo2 - transport en logistiek
- *222 Mbo2 - landbouw en natuur*
  - 2221 Mbo2 - voedsel, natuur en leefomgeving
    - 22211 Mbo2 - landbouw en dierverzorging
    - 22212 Mbo2 - slagerij, bakkerij, versindustrie
    - 22213 Mbo2 - tuinbouw en groenvoorziening
- *223 Mbo2 - techniek en ict*
  - 2231 Mbo2 - ict, media en vormgeving
    - 22311 Mbo2 - ict-ondersteuning
    - 22312 Mbo2 - mediatechniek
  - 2232 Mbo2 - techniek, bouw en procesindustrie
    - 22321 Mbo2 - bouw en infra
    - 22322 Mbo2 - elektro- en installatietechniek
    - 22323 Mbo2 - procestechniek en textiel
    - 22324 Mbo2 - werktuigbouwkunde en metaalbewerking
- *224 Mbo2 - zorg, onderwijs en dienstverlening*
  - 2241 Mbo2 - dienstverlening
    - 22411 Mbo2 - horeca
    - 22412 Mbo2 - reiniging, sport en toerisme
    - 22413 Mbo2 - schoonheids- en haarverzorging
    - 22414 Mbo2 - veiligheid
  - 2242 Mbo2 - gezondheidszorg en welzijn
    - 22421 Mbo2 - helpende zorg en welzijn
    - 22422 Mbo2 - laboratorium en gezondheidstechniek
- *229 Mbo2 - overig*
  - 2299 Mbo2 - overig
    - 22999 Mbo2 - overig

- *231 Mbo3 - economie en maatschappij*
  - 2311 Mbo3 - economie en administratie
    - 23111 Mbo3 - financieel-administratieve dienstverlening
    - 23112 Mbo3 - secretariële dienstverlening
  - 2312 Mbo3 - handel, ondernemerschap, transport en logistiek

Figure 17: Educational classification ONR2019 (part 2). *Source: ROA website:* `http://roa.sbe.maastrichtuniversity.nl/?portfolio=poa-project-onderwijs-arbeidsmarkt-2`.

23121 Mbo3 - detailhandel en groothandel
23122 Mbo3 - transport en logistiek

232 Mbo3 - landbouw en natuur

2321 Mbo3 - voedsel, natuur en leefomgeving

23211 Mbo3 - landbouw en dierverzorging
23212 Mbo3 - slagerij, bakkerij, versindustrie
23213 Mbo3 - tuinbouw en groenvoorziening

233 Mbo3 - techniek en ict

2331 Mbo3 - ict, media en vormgeving

23311 Mbo3 - ict-ondersteuning
23312 Mbo3 - mediatechniek

2332 Mbo3 - techniek, bouw en procesindustrie

23321 Mbo3 - bouw en infra
23322 Mbo3 - elektro- en installatietechniek
23323 Mbo3 - procestechniek en textiel
23324 Mbo3 - werktuigbouwkunde en metaalbewerking

234 Mbo3 - zorg, onderwijs en dienstverlening

2341 Mbo3 - dienstverlening

23411 Mbo3 - horeca
23412 Mbo3 - reiniging, sport en toerisme
23413 Mbo3 - schoonheids- en haarverzorging
23414 Mbo3 - veiligheid

2342 Mbo3 - gezondheidszorg en welzijn

23421 Mbo3 - laboratorium en gezondheidstechniek
23422 Mbo3 - verzorgende ig

239 Mbo3 - overig

2399 Mbo3 - overig

23999 Mbo3 - overig

241 Mbo4 - economie en maatschappij

2411 Mbo4 - economie en administratie

24111 Mbo4 - commerciële dienstverlening
24112 Mbo4 - financieel-administratieve dienstverlening
24113 Mbo4 - juridisch-administratieve dienstverlening
24114 Mbo4 - secretariële en algemene administratieve dienstverlening

2412 Mbo4 - handel, ondernemerschap, transport en logistiek

24121 Mbo4 - management en ondernemerschap
24122 Mbo4 - transport en logistiek

242 Mbo4 - landbouw en natuur

2421 Mbo4 - voedsel, natuur en leefomgeving

24211 Mbo4 - landbouw en dierverzorging
24212 Mbo4 - slagerij, bakkerij, versindustrie
24213 Mbo4 - tuinbouw en groenvoorziening

Figure 18: Educational classification ONR2019 (part 3). *Source: ROA website:* `http://roa.sbe.maastrichtuniversity.nl/?portfolio=poa-project-onderwijs-arbeidsmarkt-2`.

- *243 Mbo4 - techniek en ict*
  - 2431 Mbo4 - ict, media en vormgeving
    - 24311 Mbo4 - ict- en mediabeheer
    - 24312 Mbo4 - mediavormgeving
  - 2432 Mbo4 - techniek, bouw en procesindustrie
    - 24321 Mbo4 - bouw en infra
    - 24322 Mbo4 - elektro- en installatietechniek
    - 24323 Mbo4 - procestechniek en textiel
    - 24324 Mbo4 - werktuigbouwkunde en metaalbewerking
- *244 Mbo4 - zorg, onderwijs en dienstverlening*
  - 2441 Mbo4 - dienstverlening
    - 24411 Mbo4 - facilitaire dienstverlening
    - 24412 Mbo4 - horeca
    - 24413 Mbo4 - schoonheids- en haarverzorging
    - 24414 Mbo4 - sport en bewegen
    - 24415 Mbo4 - toerisme en recreatie
  - 2442 Mbo4 - gezondheidszorg en welzijn
    - 24421 Mbo4 - laboratorium en gezondheidstechniek
    - 24422 Mbo4 - pedagogisch werk
    - 24423 Mbo4 - sociaal en maatschappelijk werk
    - 24424 Mbo4 - verpleegkunde en medische ondersteuning
- *249 Mbo4 - overig*
  - 2499 Mbo4 - overig
    - 24999 Mbo4 - overig

- *311 Bachelor - economie en maatschappij*
  - 3111 Bachelor - dienstverlening
    - 31111 Bachelor - horeca, vrije tijd en faciliteitsmanagement
    - 31112 Bachelor - transport en logistiek
  - 3112 Bachelor - economie en recht
    - 31121 Bachelor - economie en econometrie
    - 31122 Bachelor - financieel management en fiscaal recht
    - 31123 Bachelor - management bedrijfs- en personeelswetenschappen
    - 31124 Bachelor - marketing en public relations
    - 31125 Bachelor - recht
  - 3113 Bachelor - journalistiek, gedrag en maatschappij
    - 31131 Bachelor - communicatie en journalistiek
    - 31132 Bachelor - psychologie, sociale en maatschappijwetenschappen
  - 3114 Bachelor - kunst, taal en cultuur
    - 31141 Bachelor - kunst
    - 31142 Bachelor - taal en cultuur
- *312 Bachelor - landbouw en natuur*
  - 3121 Bachelor - landbouw, wiskunde en natuurwetenschappen
    - 31211 Bachelor - landbouw, biologie en biochemische technologie
    - 31212 Bachelor - wis-, schei-, natuurkunde en geologie

Figure 19: Educational classification ONR2019 (part 4). *Source: ROA website:* `http://roa.sbe.maastrichtuniversity.nl/?portfolio=poa-project-onderwijs-arbeidsmarkt-2`.

- *313 Bachelor - techniek en ict*
  - 3131 Bachelor - techniek en ict
    - 31311 Bachelor - bouwkunde, civiele techniek en architectuur
    - 31312 Bachelor - chemische technologie
    - 31313 Bachelor - elektrotechniek
    - 31314 Bachelor - informatica
    - 31315 Bachelor - werktuigbouwkunde
- *314 Bachelor - zorg, onderwijs en dienstverlening*
  - 3141 Bachelor - gezondheidszorg en welzijn
    - 31411 Bachelor - maatschappelijk werk
    - 31412 Bachelor - medische diagnostiek en medische technologie
    - 31413 Bachelor - sociaal pedagogisch werk
    - 31414 Bachelor - therapie en revalidatie
    - 31415 Bachelor - verpleeg- en verloskunde
  - 3142 Bachelor - onderwijs
    - 31421 Bachelor - lerarenopleiding algemene en beroepsgerichte vakken
    - 31422 Bachelor - lerarenopleiding basisonderwijs, speciaal onderwijs en basiseducatie
    - 31423 Bachelor - onderwijskunde
- *319 Bachelor - overig*
  - 3199 Bachelor - overig
    - 31999 Bachelor - overig

- *321 Master - economie en maatschappij*
  - 3211 Master - economie en recht
    - 32111 Master - economie en econometrie
    - 32112 Master - financieel management en fiscaal recht
    - 32113 Master - management bedrijfs- en personeelswetenschappen
    - 32114 Master - marketing en public relations
    - 32115 Master - recht
  - 3212 Master - journalistiek, gedrag en maatschappij
    - 32121 Master - communicatie en journalistiek
    - 32122 Master - psychologie
    - 32123 Master - sociale en maatschappijwetenschappen
  - 3213 Master - kunst, taal en cultuur
    - 32131 Master - kunst
    - 32132 Master - taal en cultuur
- *322 Master - landbouw en natuur*
  - 3221 Master - landbouw, wiskunde en natuurwetenschappen
    - 32211 Master - landbouw, biologie en biochemische technologie
    - 32212 Master - wis-, schei-, natuurkunde en geologie

Figure 20: Educational classification ONR2019 (part 5). *Source: ROA website:* `http://roa.sbe.maastrichtuniversity.nl/?portfolio=poa-project-onderwijs-arbeidsmarkt-2`.

323 Master - techniek en ict

3231 Master - techniek en ict

32311 Master - bouwkunde, civiele techniek en architectuur

32312 Master - chemische technologie

32313 Master - elektrotechniek

32314 Master - informatica

32315 Master - werktuigbouwkunde

324 Master - zorg, onderwijs en dienstverlening

3241 Master - gezondheidszorg

32411 Master - (dier)geneeskunde en tandheelkunde

32412 Master - farmacie en gezondsheidswetenschappen

3242 Master - onderwijs

32421 Master - lerarenopleiding algemene en beroepsgerichte vakken

32422 Master - onderwijskunde en pedagogische wetenschappen

329 Master - overig

3299 Master - overig

32999 Master - overig

999 Onderwijs overig

9999 Onderwijs overig

99999 Onderwijs overig

Figure 21: Educational classification ONR2019 (part 6). *Source: ROA website:* `http://roa.sbe.maastrichtuniversity.nl/?portfolio=poa-project-onderwijs-arbeidsmarkt-2`.

Appendix B

Sector	Translation
Landbouw, bosbouw en viserij	Agriculture, forestry and fishing
Voedings- en genotmiddelenindustrie	Food and stimulants industry
Chemische industrie	Chemical industry
Metaalindustrie	Metal industry
Overige industrie	Other industry
Energie	Energy
Bouwnijverheid	Construction
Detailhandel	Retail
Groothandel	Wholesale
Vervoer en opslag	Transport and storage
Horeca	Catering
Informatie en communicatie	Information and communication
Financiële dienstverlening	Financial services
Specialistische zakelijke dienstverlening	Specialist business services
Verhuur en overige zakelijke dienstverlening	Rental and other business services
Openbaar bestuur en overheidsdiensten	Public governance and government services
Onderwijs	Education
Zorg	Healthcare
Welzijn	Well-being
Cultuur, sport en recreatie	Culture, sports and recreation
Overige dienstverlening, huishoudens en extraterritoriale organisaties	Other services, households and extraterritorial organisations
*Onbekend**	*Unknown*

Table 22: Classification of employment sectors and their translation

* For some people Statistics Netherlands has not observed their sector of employment. These are classified in the category unknown.

Appendix C

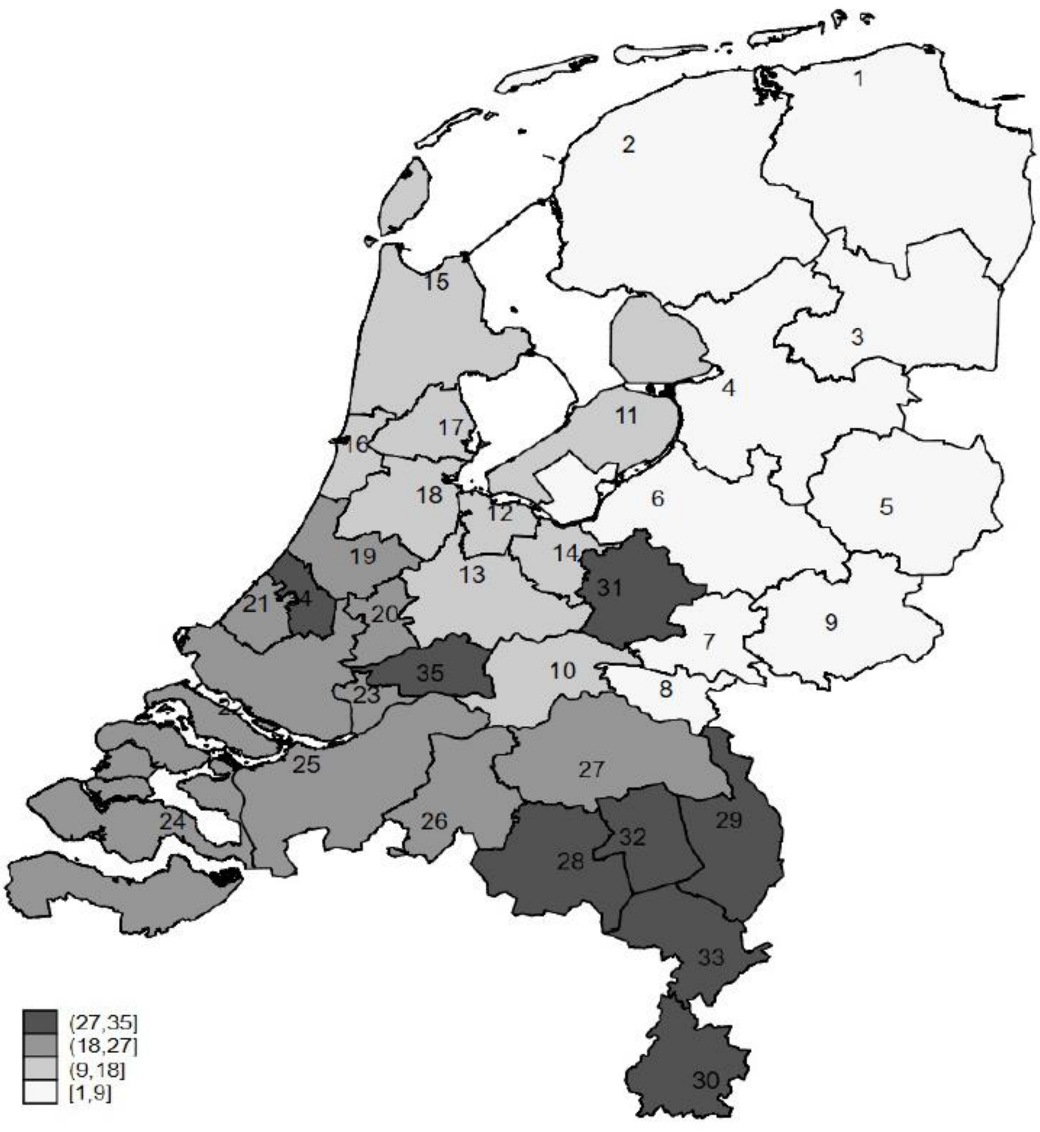


Figure 22: Regional classification - Overview of Dutch labour market regions (for legend see next page)

Label	Labour market region
1	Groningen
2	Friesland
3	Drenthe
4	Regio Zwolle
5	Twente
6	Stedendriehoek en Noordwest Veluwe
7	Midden-Gelderland
8	Rijk van Nijmegen
9	Achterhoek
10	Rivierenland
11	Flevoland
12	Gooi en Vechtstreek
13	Midden-Utrecht
14	Amersfoort
15	Noord-Holland Noord
16	Zuid-Kennemerland en IJmond
17	Zaanstreek/Waterland
18	Groot Amsterdam
19	Holland Rijnland
20	Midden-Holland
21	Haaglanden
22	Rijnmond
23	Drechtsteden
24	Zeeland
25	West-Brabant
26	Midden-Brabant
27	Noordoost-Brabant
28	Zuidoost-Brabant
29	Noord-Limburg
30	Zuid-Limburg
31	Food Valley
32	Helmond-De Peel
33	Midden-Limburg
34	Zuid-Holland Centraal
35	Gorinchem

Table 23: Overview of Dutch labour market regions

Appendix D

Labour market region	Sector	Agriculture	Food and stim.	Chemical indust.	Other indust.	Energy	Culture	Other services	Metal indus.	Information
Groningen		x	x		x		x		x	x
Friesland		x	x		x			x	x	x
Drenthe		x							x	x
Regio Zwolle		x			x				x	x
Twente		x	x	x	x				x	x
Stedendriehoek en Noordwest Veluwe		x	x		x				x	x
Midden-Gelderland									x	x
Rijk van Nijmegen									x	x
Achterhoek		x			x				x	x
Rivierenland										x
Flevoland		x								x
Gooi en Vechtstreek										x
Midden-Utrecht					x		x	x	x	x
Amersfoort										x
Noord-Holland Noord		x	x		x			x	x	x
Zuid-Kennemerland en IJmond									x	x
Zaanstreek/Waterland										x
Groot Amsterdam					x		x	x		x
Holland Rijnland		x								x
Midden-Holland										
Haaglanden		x						x		x
Rijnmond		x	x	x	x		x	x	x	x
Drechtsteden									x	
Zeeland		x		x					x	
West-Brabant		x	x	x	x				x	x
Midden-Brabant			x		x		x		x	x
Noordoost-Brabant		x	x		x				x	x
Zuidoost-Brabant					x				x	x
Noord-Limburg		x							x	
Zuid-Limburg				x	x			x	x	x
Food Valley									x	x
Helmond-De Peel			x						x	
Midden-Limburg									x	
Zuid-Holland Centraal										x
Gorinchem										

Table 24: Regions per sector with an average coverage (over t=2014 and t=2016) larger than 80% - based on occupational segments. *Source: own calculations based on EBB (Statistics Netherlands).*

Labour market region / **Sector**	Agriculture	Food and stim.	Chemical industry	Other industry	Energy	Culture	Other services	Metal industry	Information
Groningen				x			x	x	x
Friesland	x	x		x			x	x	x
Drenthe								x	
Regio Zwolle	x			x				x	x
Twente	x		x	x				x	x
Stedendriehoek en Noordwest Veluwe	x	x		x		x		x	x
Midden-Gelderland								x	x
Rijk van Nijmegen								x	x
Achterhoek				x				x	
Rivierenland								x	
Flevoland									x
Gooi en Vechtstreek									x
Midden-Utrecht						x	x	x	x
Amersfoort								x	x
Noord-Holland Noord	x	x		x		x		x	x
Zuid-Kennemerland en IJmond								x	x
Zaanstreek/Waterland									
Groot Amsterdam				x		x	x	x	x
Holland Rijnland	x			x					x
Midden-Holland									
Haaglanden	x						x	x	x
Rijnmond			x	x		x	x	x	x
Drechtsteden								x	
Zeeland			x						
West-Brabant	x	x	x	x			x	x	x
Midden-Brabant			x	x				x	x
Noordoost-Brabant	x	x		x		x	x	x	x
Zuidoost-Brabant				x				x	x
Noord-Limburg	x							x	
Zuid-Limburg			x	x			x	x	
Food Valley								x	x
Helmond-De Peel								x	
Midden-Limburg								x	
Zuid-Holland Centraal							x		x
Gorinchem									

Table 25: Regions per sector with an average coverage (over t=2014 and t=2016) larger than 80% - based on ONRsector aggregation level. *Source: own calculations based on EBB (Statistics Netherlands).*