Technische Universität Dresden
Fakultät Verkehrswissenschaften “Friedrich List”
Institut für Wirtschaft und Verkehr
Professur für Ökonometrie und Statistik, insb. im Verkehrswesen

Dissertation

Environmental Awareness and Labour Market Tightness
An Analysis on Regional Level

Freie wissenschaftliche Arbeit
zur Erlangung des akademischen Grades einer
"‘Doctor rerum politicarum"’

eingereicht von:
Dipl.-Verk.wirtsch. Stefanie Lösch (geb. Stock)

eingereicht am:
05.03.2020

verteidigt am:
30.11.2020

Gutachter:
Prof. Dr. rer. pol. Ostap Okhrin
Prof. Dr. rer. nat. habil. Bernhard Schipp

Abstract

This thesis reveals methods for estimating individual perception at the regional level, such as environmental awareness or wage effects due to labour market tightness. Environmental awareness belongs to individual attitudes, which is driven by socialization, culture and education. Consequently, it is difficult to compare environmental awareness between regions. Labour market tightness might be reflected in individual wages, but the latter is also triggered by a lot of exogenous variables. Given that a simple linear regression model fails in both cases, existing alternative approaches from the econometrics like Multiple Indicator Multiple Cause model and Least Absolute Shrinkage and Selection Operators are used.

First, a Multiple-Indicator Multiple-Causes model is introduced, which allows us to estimate a not-directly observable individual attitude, environmental awareness, for different regions and to rank them. The method is cost-effective and less time-consuming, it also allows for comparisons between regions. The study area serves 81 regions in Russia. The model is constructed in such a way that Internet queries from the search engine ©Yandex are assumed to be indicators, which are affected by the regional *environmental awareness index*. In addition, regional characteristics, such as Gross Regional Product per capita, the proportion of employees in specific industry sectors, and also the environmental situation within the regions are potential cause variables. The regional environmental awareness index is estimated for each of the Russian regions from January 2014 until April 2016. Furthermore, the findings shows a rather non-linear positive relationship between the regional environmental awareness and regional wealth, as well as a strong negative correlation with the temperature. The colder the region and the observed month, the higher the interest of the population in environmental topics. Furthermore, the regions can be grouped into four environmental awareness clusters by using k-means clustering algorithm. It seems that the environmental awareness index shrinks from the Eastern to the Western part of Russia. The highest values can be estimated in Chukotka, Kamchatka, and Magadan. The lowest values are found in the rather poor and warm Caucasus area.

Second, another issue concerns the estimation of an effect of an observable regional variable, such as labour supply shortage, on individual wages. This thesis investigates the ten year wage development of employees who first enter the labour market from 1995 until 2004 and looks for positive wage effects of labour market tightness in different occupational groups. Due to incomplete vacancy data, labour market tightness is measured as the number of unemployed people divided by the number of employees within an occupational field and region. Mean and quantile regression methods are applied. Because the number of right-hand side variables could lead to incorrect detected statistical significant coefficients, different Least Absolute Shrinkage and Selection Operators are used for reducing the variables set. The findings suggests that regional labour market tightness in occupational fields affects individual wages. Employees who start their carrier in a tighter labour market enjoy higher wage growth compared to workers from more relaxed labour markets. The wages in technical professions, such as several engineer groups, IT-occupations, technicians, and also in some commercial occupations are especially affected

by a shortage of labour supply. Health care occupations, such as nurse, reveals a complete reverse relationship. A shortage of workforce seems to be correlated with smaller wages.

Danksagung

An dieser Stelle möchte ich den Menschen meinen herzlichen Dank aussprechen, die maßgeblich an der Entstehung dieser Arbeit beteiligt waren. Mein großer Dank gilt Prof. Dr. Ostap Okhrin, da er sich bei der Übernahme des Lehrstuhls für *Ökonometrie und Statistik, insb. im Verkehrswesen* bereit erklärte meine Dissertation zu betreuen. Prof. Okhrin stand mir jederzeit mit fachlichen Ideen und Ratschlägen zur Seite und ermöglichte mir das Wahrnehmen notwendiger Dienstreisen. Des Weiteren möchte ich an dieser Stelle recht herzlich Prof. Dr. Bernhard Schipp danken, der kurzfristig das Zweitgutachten der Dissertation übernahm.

Die Dissertation basiert auf fünf wissenschaftlichen Fachartikeln, die nicht ohne Unterstützung entstanden wären. So danke ich recht herzlich Prof. Dr. Hans Wiesmeth für die interessante Forschungsidee eines *Environmental Awareness Index*, den Zugang zu den Daten der Suchmaschine *Yandex* sowie seine fortlaufende Unterstützung über den gesamten Zeitraum der Arbeit hinweg.

Des Weiteren möchte ich herzlich Prof. Dr. Stephan Brunow sowie dem Institut für Arbeitsmarkt und Berufsforschung (IAB) danken. Prof. Brunow ermöglichte mir den Zugang zu den Daten des IABs und unterstützte die Arbeit mit seinem umfangreichen Arbeitsmarktwissen.

In diesem Zusammenhang dürfen die Kollegen vom IAB Sachsen-Anhalt-Thüringen in Halle nicht unerwähnt bleiben. Dr. Michaela Fuchs und Birgit Fritzsche stellten mir eine professionelle Arbeitsumgebung zur Verfügung, in der ich mit den Daten arbeiten konnte und standen mir bei technischen Schwierigkeiten mit Rat und Tat zur Seite.

Ein besonderer Dank gilt außerdem Dr. Stefan Lämmer. Dr. Lämmer überzeugte mich bereits als Studentin davon am Lehrstuhl zu arbeiten und begeisterte mich für das Fach Statistik. Er glaubte schon damals an meine Fähigkeiten und ermutigte mich zum Anfertigen einer Dissertation.

Zum Schluss bedanke ich mich recht herzlich bei meiner Familie, allen voran meinem Mann Hendrik Lösch und meinem Sohn Erik Lösch, die mir Mut machten und es verstanden, mich von der Wissenschaft abzulenken, wann immer das nötig war. Beide ermöglichten mir Dienstreisen und Termine ohne schlechtes Gewissen wahrzunehmen.

Auch dank des unermüdlichen organisatorischen Einsatzes meiner Eltern, Margitta und Harald Stock, sowie meiner Schwiegereltern, Ursula und Anton Lösch, ließen sich Beruf, Dissertation und Kind gut vereinbaren. Ohne ihre Hilfe wäre diese Arbeit immer noch nicht abgeschlossen.

Teile der Dissertation wurden vom Proof Reading Service (`https://www.proof-reading-service.com`) lektoriert.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Two Issues in Empirical Regional Science

This thesis follows up on the issue of modelling not-directly measurable individual perceptions, such as *environmental awareness* or wage effects due to *labour market tightness*, at the regional level. If they were easy to quantify, then only a linear regression model would need to be applied. In the first case, *environmental awareness* is a more subjective individual attitude, which cannot be easily transferred at the regional level. The second case handles an object that is triggered by an enormous list of exogenous variables. This thesis adapts some existing methods from econometrics like Multiple-Indicator-Multiple-Causes (MIMIC) model and Least Absolute Shrinkage and Selection Operator (LASSO) to overcome the mentioned problems.

Chapter 2 discusses a measurement for *environmental awareness* of a population in a geographical area, which allows the compatibility of environmental consciousness with other regions. In some studies a very time-consuming and expensive survey is performed, see Hiramatsu, Kurisu, and Hanaki (2015). However, respondents are inclined to give socially desirable answers (the so-called "yes-bias"). Because the replies depend enormously on individual education, personal socialisation and local culture, they might differ between regions and countries. Thus, the answers are hardly comparable and it is not useful to sequence the regions in terms of the population's *environmental awareness*. Another method might be to compare objective environmental data (i.e. expenditure on environmental protection, waste per capita, waste sorting rate, etc.). However it is often difficult to get access into this kind of detailed environmental data for each region/area of interest. Furthermore the method of data record may vary between regions. Chapter 2 introduces an approach that avoids such problems. The proposed method allows a comparison of various regions in respect to their *environmental awareness* and is less time-consuming and less-expensive. The idea is to use the number of Internet queries of environmental-specific words and phrases, which might be affected by attitude of a population in terms of environmental issues. The more that a regional population is interested in such topics, the more frequent the Internet is asked for a certain word or phrases (i.e. relative to all other requests). This value can easily be compared with other regions. Furthermore the interest in environmental topics might be driven by regional specific characteristics, such as the regional wealth or industrialisation, which should be considered. The proposed method is explored by investigating the *environmental awareness* of 81 regions (Oblasts, Krais and Republics) of the largest country of the world—the Russian Federation. Given that a simple regression analysis fails estimating such an unobservable latent construct as *environmental awareness*, a MIMIC model from Jöreskog and Goldberger (1975) is performed and estimated through optimizing a Maximum Likelihood function. In Chapter 2 the advantages of the method are highlighted. Based on the theoretical concept, different empirical models are estimated. This allows a measurement to be made of an *environmental awareness* index for each of the 81 Russian regions and to rank them. The original method is extended by considering trend and seasonal effects. Therefore, climate or temperature circumstances, respectively, of the different regions deliver especially interesting results. A cluster analysis is performed for grouping the Russian regions in terms of their *environmental awareness*. In addition, environmental awareness is separately measured for (Sub-) Arctic areas and compared with all other Russian regions. Beside the climate data, further regional characteristics as drivers of environmental awareness are discussed. In addition, further application areas of the measurement are mentioned. A summary concludes the chapter.

A further issue in empirical regional science is discussed in the second part of this thesis. Chapter 3 occupied with a rather quantifiable variable *labour market tightness* and its effects on measurable individual wages. However, different not-directly measurable influencing factors on individual wages need to be considered, such as individual negotiating skills and productivity, for extracting the effect of a labour supply shortage. However, a very homogenous sample of employees may mitigate this issue. Consequently, a 10% sample of very young German employees who finished a vocational training is drawn. The Institute for Employment Research provides the data, including information about the training before, the time up to entering and the first 10 years on the labour market, such as wage changes, periods of unemployment, firm changes, etc. Thus, there might be less variations in the productivity level within the sample. In addition, employment specific and regional information are added. However, this leads to a sample of 320 thousand individuals with 628 observed characteristics. The individual log-wages at the first-ten years are predicted by estimating a linearised regression model through minimizing the sum of squared errors. Because almost the half of the variables are binary, the estimation could suffer from collinearity. Although this does not influence the prediction quality of the log-wages, it does make the extraction of the effect of occupational-specific labour market tightness on the log-wages difficult because collinearity leads to erroneously statistical significant estimated coefficients. Consequently, LASSO from Tibshirani (1996) and Zou (2006) is used to shrink the number of variables. Because occupational specific differences in the effects of labour market tightness on the log-wages might be partly driven by technological progress, stronger effects for technical professions are expected. Furthermore, differences between the quantile levels are investigated. On the one hand, it is conceivable that well-paid employees have a stronger negotiating position compared to low-paid workers because they have certain rather rarely skills. On the other hand, low-skilled people might rather accept a job for lower wages due to the competitive situation with machines. Therefore, the unconditional quantile regression method by Firpo, Fortin, and Lemieux (2009) is used to measure the effect of labour shortage on the different quantile levels of the log-wage distribution. Finally, Chapter 4 concludes the thesis and gives a prospect for future work.

1.2 Published and Submitted Articles

Five papers have been published as part of this research project. Chapter 2 quantifies regional *environmental awareness* and includes parts of the following four published papers:

1. Khakimova et al. (2019): Index of environmental awareness through the MIMIC approach. *Papers in Regional Science*, 98, 3, 1419 – 1441. `https://doi.org/10.1111/pirs.12420`

This paper introduces a structural equation model for estimating a latent construct, the regional *environmental awareness* (EA), by using Internet queries as indicators, which are affected by EA, as well as regional characteristics as potential causes for EA. A so-called EA index is estimated for each of the 81 Russian regions for two time points: the years 2014 and 2015. This paper highlights the relationship between regional wealth and EA. It also investigates the correlation between the environmental situation (for example the regional contaminated water per km area) and EA of the population. The regions can be ranked in terms of EA.

My part of the paper includes the performance of the empirically analysis, the editing of the results in tables and graphics, as well as the interpretation of the results.

2. Lösch, Okhrin, and Wiesmeth (2018a): Awareness of climate change: differences among Russian regions. *Area Development and Policy, Routledge*, 4, 3, 284 – 307. `https://doi.org/10.1080/23792949.2018.1514982`

   This paper is basically an extension of the previous paper Khakimova et al. (2019). The temporal and spatial development of EA is investigated. The research area is again the 81 Russian regions during the time period between January 2014 and April 2016. In contrast to Khakimova et al. (2019), more data are available and the EA index development can be analysed for each of the 28 months. We find seasonal and spatial variations of the EA, which are strong correlated with the regional temperature. The colder the region, the higher the EA. Furthermore, the EA is higher in the winter months than in the summer months. This effect can be observed in all regions. It seems that there is a regional spread from East Siberia to the Western part of Russian in Europe. The highest EA indices can be measured in the Eastern part of the Russian Federation. The lowest in the poor, rather warm Caucasus area. Four EA clusters are found.

   Again, my main part lies on the development and estimation of the empirical models as well as the statistical analysis and interpretation of the results.

3. Lösch, Okhrin, and Wiesmeth (2018b): Awareness of climate change: focus on the Russian Arctic Zone. *Proceedings of the International Research*

Workshop on Information Technologies and Mathematical Modeling for Efficient Development of Arctic Zone, ceur-ws.org, Vol. 2109, 38-42. `http://ceur-ws.org/Vol-2109/paper-07.pdf`

This paper focus on the EA in the Arctic area, which is also significantly different from other regions. We suspect that the stronger regions are concerned by environmental problems (e.g. shrinking of the permafrost areas in the Northern part of Russia) and they have a larger interest in environmental topics.

In this paper, I performed the empirical analysis and edited the results.

4. Lösch, Okhrin, and Wiesmeth (2017): Diffusion of environmental awareness. *diffusion-fundamentals.org*, Vol. 30, No. 2, 1-16. `http://diffusion.uni-leipzig.de/pdf/volume30/diff_fund_30(2017)02.pdf`

   This paper is a preparatory work to Lösch, Okhrin, and Wiesmeth (2018a). The temporal analysis is on a quarterly instead of monthly level. The results are more inexact. However, the seasonal variation of the EA, as well as four spatial clusters are found. In comparison to Lösch, Okhrin, and Wiesmeth (2018a), a link between regional temperature and EA is not yet visible.

   As in the previous papers, my main work consists of the empirical analysis and interpretation of the results.

Furthermore, Chapter 3 analyses the influence of a labour market shortage on individual wages. Parts of this chapter come from a joint project with

5. Brunow, Lösch, and Okhrin (submitted): Labour Market Tightness and Individual Wage Growth – Evidence from Germany.

   The paper investigates the 10-year-wage growth of employees who finished a vocational training and it looks for positive wage effects of labour market tightness in different occupational groups. Different mean and quantile regression methods, as well as dimension-reduction approaches are applied. The findings suggest that regional labour market tightness in occupational fields significantly explains wages. Individuals that start their carrier in a tighter labour market enjoy higher wage growth compared to workers from more relaxed labour markets. The effect is especially strong in technical professions, such as several engineer groups, IT-occupations, technicians, and also in some commercial occupations. Interestingly, health care occupations, such as nursing, reveal a complete reverse relationship. A shortage of workforce seems to be correlated with smaller wages.

My role in this paper was to develop the motivation and literature overview (partly), the data preparation (partly), empirical work of this paper, including estimation and interpretation of the results (mainly).

Chapter 2

Quantifying Regional Environmental Awareness

This chapter contains parts, figures and tables, which are published in Khakimova et al. (2019), Lösch, Okhrin, and Wiesmeth (2018a), Lösch, Okhrin, and Wiesmeth (2018b) and Lösch, Okhrin, and Wiesmeth (2017). The citations is unified and the format is adjusted.

2.1 Introduction

This Section 2.1 and 2.2 is adopted by Khakimova et al. (2019), pages 1419 – 1424.

Environmental degradation remains a global issue, and the Russian Federation is no exception. Various regions and cities continue to suffer from poor environmental conditions due to, or in spite of, the economic decline after the breakdown of the Soviet Union and the economic recovery in the years thereafter (Kozeltsev et al. 2013). Air pollution, soil and water contamination are likely responsible for unnecessarily high rates of morbidity and even mortality. With emissions a little above 1.7 Gt CO_2 equivalents per year, as plotted on Figure 2.1, Russia is still among the largest five emitters of greenhouse gases after China, the United States, European Union and India.

There are various signals and actions, which document the political willingness to reduce environmental pollution in Russia. For example, in 1991 a system of fees on emissions, discharges, and solid waste became a key element of Russia's

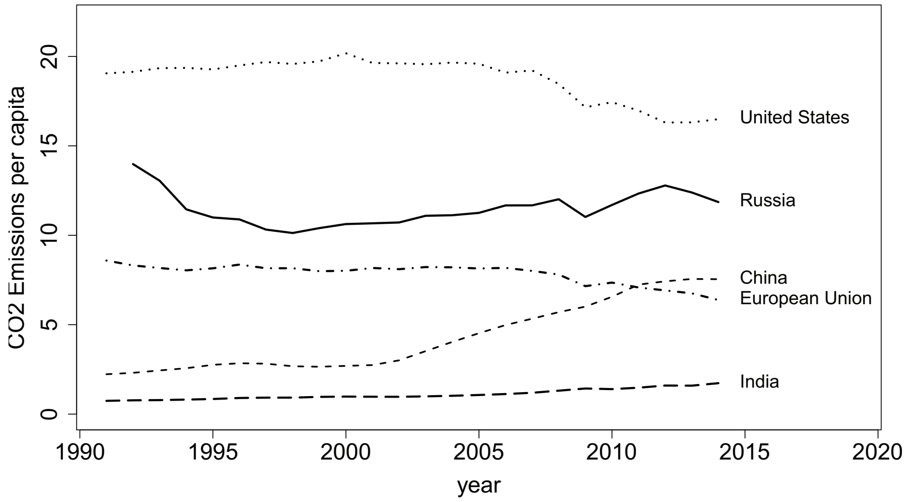

Figure 2.1: Left: CO_2 emissions (in metric tonnes) per capita of the five largest greenhouse gas emitters from 1991-2014. *Source:* World Bank (2018); Khakimova et al. (2019).

environmental policy. There were even differentiated fees per unit of emissions of hazardous substances (Kozeltsev et al. 2013). Moreover, in 2004 Russia was the last party to join the Kyoto Protocol and was instrumental regarding its successful implementation, and in 2009, the Russian government passed a resolution *On the Measures Stimulating Reduction of Atmospheric Pollution by Products of Associated Gas Flaring*, which limits associated gas flaring levels to 5% of the entire output from 2012 (IEA Statistics 2017).

The economic recession following the breakdown of the Soviet Union, significantly reduced per capita emissions of greenhouse gases in Russia. In 2015, due to this and other reasons, these emissions were still some 30% below their 1990 level, despite some smaller more recent increases of 7% in 2010 and 5% in 2011 (IEA Statistics 2017). As a consequence, it was relatively easy for Russia to reach the Kyoto target of a 0% reduction of GHG emissions in the period from 1990 to 2012. With regard to the UN Climate Change Conference in Paris in 2015, Russia declared the willingness to stabilize CO_2 emissions on a level of 75% of the emissions in 1990.

Thus, Russia seems to be ready to reduce environmental pollution and to participate in the provision of global environmental commodities, such as the reduction of greenhouse gas emissions. In this context the question arises to the extent of the environmental awareness (EA) of the people living in the Russian regions. To be more concrete: to what extent are people concerned about environmental issues? Can the government expect a certain “willingness” in the population to support corresponding initiatives? How to measure this willingness, this “awareness”? Are there differences regarding this awareness in the regions of the Russian Federation? If so, which regional factors are responsible for these differences? Answers to these questions provide valuable insight into the current and, in particular, the future state of the environment in Russia and the Russian regions. After all, documented EA in the population could raise pressure on local politics to improve the environmental situation and/or invest in environmental protection projects.

As environmental issues such as global warming, air pollution, soil and water contamination and others are of relevance in many regions, EA in the context of these questions is of relevance for an investigation, the research agenda of this paper.

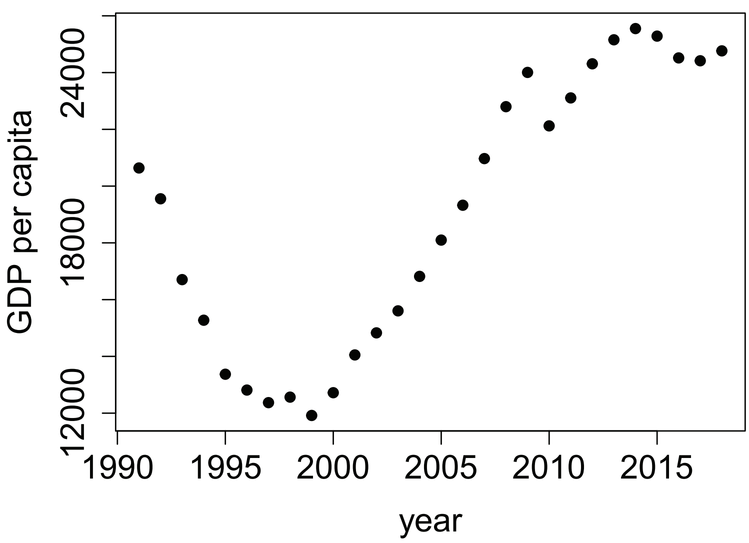

Figure 2.2: Left: GDP per capita in 2011 international Dollars, purchasing power parity, from 1990-2017. *Source:* World Bank (2018); Khakimova et al. (2019).

The Russian Federation provides a particularly interesting study area for these research questions. If, for example, there were a relationship between economic development and whatever concept of EA, then a somewhat more environmentally focused behaviour could be expected from a growth-stimulating economic policy, mirroring also relevant differences between the regions. As Figure 2.2 shows, GDP per capita in Russia increased from 12,051 Dollars in 2000 to 24,516 Dollars in 2015 (international Dollars based on 2011; data source World Bank (2018)). The substantial differences between the regions of the Russian Federation in terms of economic and demographic development, and in terms of ethnic composition should, thus, allow a careful analysis of the research questions.

First of all, however, the basic question referring to the concept of EA has to be addressed. How can "awareness" be measured in the framework outlined above? Moreover, what are the relevant differences among the Russian regions in this context? Which structural characteristics of the various regions are decisive for differences regarding EA? Is there, in particular, a relationship between GRP per capita of a region and EA? The direct effects of economic, demographic and political conditions on the environment have been investigated in many empirical studies. However, the sentiments or understanding of environmentally relevant issues in the population are usually not taken into account, although they are important for a successful implementation of environmental policies. This refers, for example, to the concept of "stakeholder integration", which is of utmost importance for environmental policies (Wiesmeth and Häckl 2017).

For this reason, we develop a new index of EA and provide a ranking for 81 Russian regions for two periods in 2014 and 2015. This index of EA is a latent construct, which affects various indicators derived from search entries in ©Yandex, the most prominent Russian search engine. The data refers to some 200 phrases, both in Russian and in English, which are related to environmental issues and, thus, mirror the environmental concerns of the people. These search phrases are clustered into five categories covering, among others, direct climate change queries as well as literature searches. In the average, the environmental phrases generated between approximately 4750 (July 2014) and 7400 (January 2015) clicks per month and Russian region in our data set.

We assume that these indicators are depended on or, rather, result from certain economic and societal developments in the Russian regions. Therefore, we integrate these "causes" into the model in order to investigate their influence on a specific level of the index. The resulting model is the Multiple Indicators-Multiple Causes (MIMIC) model, which allows us to estimate the latent variable from various indicators and various causes of EA (Jöreskog and Goldberger 1975).

In the next section, we briefly investigate the development of the concept of EA in the literature. The formal aspects of the MIMIC model are presented in Section 2.3 and 2.4, including a discussion on the applicability of this model to the concept of EA. The data are introduced in 2.5. Section 2.6 and 2.7 presents the empirical results, and some final in Section 2.8 remarks conclude the paper.

2.2 Environmental Awareness in the Literature

In this Section we first consider briefly the development of the concept of EA in the literature. Thereafter the focus is on the literature describing factors influencing EA.

2.2.1 Emergence of the Concept of EA

EA is usually considered a prerequisite for environmentally friendly behaviour. However, it is not so straightforward to conceptualise it in order to make stringent use of it in academic research. In particular, in order to understand the EA index developed in this paper, it is necessary to have a brief look at the literature documenting the emergence of the concept of EA.

Not too surprisingly, interest in the concept of EA or environmental consciousness originated with the ecological movement in the 1960s. According to Soyez et al. (2009), in order to explore the business potentials of environment-related commodities, researchers in marketing and social psychology focused first on the personal characteristics of environmentally conscious people, such as socio demographic variables. In the 1970s and 1980s, environmentally friendly behaviour was explained more in terms of environmentally friendly attitudes, which were measurable by means of multi-item scales. The "theory of planned behaviour", which is applicable to this context (Ajzen 1991), allows an integration of a variety of factors or attitudes (self interest, norms, situational barriers, etc.) that influence a specific action or behaviour related, for example, to the environment.

Personal value orientation as a precursor of sustainable behaviour was considered in a further stream of research followed by cultural values, which have been investigated approximately for the last 15 years (Soyez et al. 2009). Of course, cultural values form the basis for cross-cultural studies on environmentally friendly behaviour, which are, again, of particular interest for researchers in marketing and social psychology. In this context, Soyez (2012) analyses how environmentally friendly behaviour is influenced by cultural values and how national cultural values can be linked to personal pro-environmental behaviour.

These studies present further links to the earlier literature, and are among the first to provide insight into environmentally friendly consumer behaviour in Russia in comparison to Germany. They make use of methods of factor analysis and structural equation modelling to conceptualize environmentally friendly value orientation.

Besides behavioural characteristics, indices of environmental quality have been introduced to categorise EA. There is, for example, the prominent Environmental Performance Index (EPI), a project affiliated with the Yale Center for Environmental Law and Policy and the Center for International Earth Science Information Network at Columbia University (Hsu et al. 2016). The EPI ranks countries' performance on high-priority environmental issues: protection of human health and protection of ecosystems. Within these policy areas the EPI scores the performance of the countries by means of issue areas comprised of a multitude of indicators. Each indicator is weighted within the issue categories to create an issue category score with the weightings respecting the indicator's relevance, among other things. Then, these subindices are weighed approximately equally, thus yielding indices of protection of human health and protection of ecosystems. Finally, those two indicators are summarized into the EPI by applying equal weights. Cf. Hsu et al. (2016), pp. 23. Lisciandra and Migliardo (2017), for example, use the EPI to explain the effect of corruption on environmental degradation. They find a negative relation of corruption level as well as a positive correlation of regional income on the regional environmental quality.

For another example, Expert RA (2018) introduces an Investment Index for Russia focusing on, among other topics, ecological risks. This index relates investment risks in Russian regions to regional environmental quality. It considers costs of pollutant removal, the elimination of inherited environmental liabilities and other environmental issues on a regional level to measure investment risks. Some aspects of this index and the associated ranking of the Russian regions are discussed in Section 2.6.2.

Clearly, the choice of the various weights in the construction of such indices requires a lot of experience on the one hand, and some stability regarding the influence of certain variables on the indices on the other. In addition to that the EPI and the

Investment Index are performance indicators. If interpreted as indices of EA, they are measuring EA through past environment-relevant activities of the population. They do not directly respect the attitude of the population regarding environmental issues and means to influence this attitude and environmentally friendly behaviour thereafter. Performance indices and awareness indices should therefore be expected to yield different rankings, in general. Moreover, as we do not really know the importance of the various indicators in our approach, it might be a good idea to refrain from exogenously assigned weights, at least in this first step. All these issues are of relevance for the design of environmental policies.

These considerations guide us to our approach based on structural equation modelling. Soyez (2012) uses this methodology and derives conclusions on environmentally friendly value orientation for cross-cultural marketing. There are, however, only few words on possibilities to influence these orientations (cf. p. 641). But this is an important aspect for designing and implementing environmental policies in areas, which require the cooperation of the individuals. As most environmental issues belong to these areas, we are looking for economic and socio-economic causes affecting EA.

2.2.2 Influencing EA

Buehn and Farzanegan (2013) are among the first to construct an environmental index, which is dependent on certain determinants. Their "new index of air pollution" is calculated by means of a MIMIC model and shows the influence of economic, demographic, and governance factors. Some of these factors can be modified by economic policies and can, thus, also affect the future level of the index. This, as already indicated, might be of importance for certain activities of relevance for the environment.

The analysis of Buehn and Farzanegan (2013) supports the Environmental Kuznets Curve (EKC) hypothesis, pointing to a significant influence of the level of GDP per capita on the index (cf. p. 109). The economic background for observations like this is the intrinsic nature of environmental commodities. They are, at least in industrialised and newly industrialised countries, characterised by a high income elasticity, resulting from a gradual transformation of societies with increasing economic prosperity (Inglehart 1990). Consequently, demand for these commodities should rise and environmental pollution should be reduced as real GDP per capita increases.

In order to provide support for this hypothesis, Grossman and Krueger (1995) find no evidence that environmental quality deteriorates steadily with economic growth. In other words, for most indicators, economic growth brings an initial phase of deterioration followed by a subsequent phase of improvement. The turning points for the different pollutants vary but in most cases they come before a country reaches a per capita income of $8000 (for USD in 1985). Grossman and Krueger (1995) use urban air pollution, the state of the oxygen regime in river basins, fecal contamination of river basins, and contamination of river basins by heavy metals.

Going one step further, the EKC is a hypothesised relationship between various indicators of environmental pollution and GDP per capita (Stern 2004). The concept emerged in the early 1990s with studies of the potential environmental impacts of NAFTA. Stern (2004) provides an interesting survey on "the rise and the fall" of the EKC, characterising the EKC as an essentially empirical phenomenon with not much support from econometrics. Similarly, Huang, Lee, and Wu (2008) do not find empirical evidence supporting the EKC hypothesis for greenhouse gas emissions.

Nevertheless, in the last few years increasingly advanced econometric techniques were employed to investigate the existence or non-existence of the EKC with respect to aspects of global warming. Fosten, Morley, and Taylor (2012), for example, analyse the EKC with respect to CO_2 and SO_2 emissions in the United Kingdom, and provide a useful literature survey on the econometric methods used in this context. Cf. also Brajer, Mead, and Xiao (2011), He and Richard (2010), Wang (2013), Yang, He, and Chen (2015). Maddison (2006) shows that the countries' emissions per capita are affected by events in neighbouring states. These effects should be considered by creating an EKC. Rupasingha et al. (2004) find a more cubic than inverted-U-shaped relationship between county per capita income and toxic pollutants, also affected by ethnic diversity. A comprehensive survey of the EKC hypothesis up to the year 2004 is provided by Dinda (2004).

Beyond GDP per capita, there are other factors, which could affect the attitude towards environmental issues. Diederich and Goeschl (2014), for example, uncover education, the information structure among the population, and exogenous environmental conditions as causes of voluntary climate action. Similarly, Rabe, Borick, and Lachapelle (2011), and Lorenzoni and Pidgeon (2006) study public views on climate change in the United States and Canada, and in Europe and the United States, respectively. The focus of the special survey "European's attitude towards climate change” in the “Eurobarometer” from the European Commission (2008) was on, among other issues, the extent to which citizens feel informed about climate change, and reveals the influence of information on EA.

Thus, there is some justifiable interest in factors determining EA, in particular regarding the economic factors such as GDP per capita.

2.3 Multiple Indicators-Multiple Causes Model

Khakimova et al. (2019) pick up the approach used by Buehn and Farzanegan (2013) processing an index of air pollution for 122 countries for the period between 1985 and 2005. Their index is based on a MIMIC model, which was originally developed by Jöreskog and Goldberger (1975) and is a special case of a structural equation model. Buehn and Farzanegan (2013) use three different indicators, different kind of measurable emissions, which are influenced by the air pollution index and several cause variables affecting the index. Similarly, EA can hardly be described by means of a single indicator, and different framework conditions might lead to different levels of awareness.

The MIMIC approach helps to overcome the difficulty with a priori fixed weights for the various indicators in a situation where awareness is still emerging. This might currently characterize the situation in the Russian Federation and the Russian regions. It uses well defined indicators to measure a latent construct (index of EA) with associated properties and regresses them against theoretically discovered causes, as Buehn and Farzanegan (2013) did.

The two parts of the model can be explained as a measurement model for the latent construct and a structural part, which describes the causal structure of the model. The measurement part looks as follows:

$$y = \lambda\,\eta + \varepsilon, \tag{2.1}$$

where $y = (y_1, y_2, \ldots, y_p)^\top$ is a set of observable endogenous indicators. They are affected by EA, which is the latent variable η, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^\top$ being a vector of p random errors. The factor loadings are summarized in the p-vector λ.

The structural part follows the theoretical assumptions and can be written as

$$\eta = \beta^{\top} x + \zeta, \tag{2.2}$$

where $x = (x_1, x_2 \ldots, x_k)^{\top}$ are exogenous causes, $\beta = (\beta_1, \beta_2, \ldots, \beta_k)^{\top}$ is a set of model parameters, and ζ being a random error term. Inserting (2.2) into (2.1) results in

$$y = \lambda \left(\beta^{\top} x + \zeta\right) + \varepsilon = \Pi^{\top} x + v, \tag{2.3}$$

with $\Pi = \beta \lambda^{\top}$ and $v = \lambda \zeta + \varepsilon$. It is assumed that the random errors ε and ζ are mutually independent and normally distributed

$$\varepsilon \sim \mathrm{N}(0, \Theta^2) \qquad \text{and} \qquad \zeta \sim \mathrm{N}(0, \sigma^2), \tag{2.4}$$

with $\mathrm{E}(\zeta \varepsilon^{\top}) = 0$ where $\Theta = \mathrm{diag}(\theta_1, \theta_2, \ldots, \theta_p)$. The parameters of λ and β, and the variances θ^2 and σ^2 of the error terms ε and ζ, can be estimated using a ML approach. There is indeterminacy in the structural parameters (if λ is multiplied by a scalar and β and σ^2 are divided by the same scalar parameters do not change). Avoiding this indeterminacy, one parameter is set fixed, see Goldberger and Hauser (1971).

Following Jöreskog and Goldberger (1975) the fixed case, where x is $[N \times k]$ fixed matrix, y is multivariate normally distributed $[N \times p]$ matrix, $x \in \mathcal{M}_{N \times k}$ and $y \in \mathcal{M}_{N \times p} \sim \mathrm{N}(x\Pi, \Omega)$, with the sample size $i = 1, \ldots N$.

The multivariate probability density function of y is

$$f(y) = \frac{1}{(2\pi)^{p/2} |\Omega|^{1/2}} \exp\left\{-(y - x\Pi)^{\top} \Omega^{-1} (y - x\Pi)/2\right\}. \tag{2.5}$$

Thus the likelihood function can be described by

$$
\begin{aligned}
\mathcal{L}(y|x\Pi,\Omega) &= \prod_{i=1}^{N} f(y_i;\lambda_2,\dots\lambda_p,\beta,\theta^2,\sigma^2) \\
&= \prod_{i=1}^{N} \frac{1}{(2\pi)^{p/2}|\Omega|^{1/2}} \exp\left\{-(y_i - x_i\Pi)^\top \Omega^{-1}(y_i - x_i\Pi)/2\right\} \\
&= \frac{1}{(2\pi)^{Np/2}} \frac{1}{|\Omega|^{N/2}} \exp\left\{-\sum_{i=1}^{N}((y_i - x_i\Pi)^\top \Omega^{-1}(y_i - x_i\Pi)/2)\right\}
\end{aligned}
$$

The natural logarithms simplified the estimation.

$$
\begin{aligned}
\log\mathcal{L}(y|x\Pi,\Omega) = l(y|x\Pi,\Omega) &= -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log(|\Omega|) - \frac{1}{2}\left(\sum_{i=1}^{N}(y_i - x_i\Pi)^\top \Omega^{-1}(y_i - x_i\Pi)\right) \\
l(y|x\Pi,\Omega) &= -\frac{N}{2}\left\{\log|\Omega| + \operatorname{tr}\left(\Omega^{-1}W\right)\right\} - \frac{Np}{2}\log(2\pi) \rightarrow \ \text{max!}
\end{aligned}
$$

with the sample size N, $\Omega = \mathrm{E}\left(vv^\top\right) = \sigma^2\lambda\lambda^\top + \Theta^2$ and $W = (y - x\Pi)^\top(y - x\Pi)$.

The function can be maximized by setting the first partial derivative of $\mathcal{L}(\beta,\ \lambda,\ \theta^2,\ \sigma^2)$ with respect to all arguments to zero.

$$
\frac{\partial l(\beta,\lambda,\theta^2,\sigma^2)}{\partial\beta,\partial\lambda,\partial\theta^2,\partial\sigma^2} = 0
$$

The estimators of the coefficients and variances $\hat{\lambda},\hat{\beta},\hat{\theta}^2,\hat{\sigma^2}$ can be computed through an iterative approach.

1. Set initial values $\hat{\lambda}_0$, $\hat{\Omega}_0$;
2. Compute $\hat{\beta} = \left(\frac{1}{\hat{\lambda}^\top\hat{\Omega}^{-1}\hat{\lambda}}\right)(x^\top x)^{-1}x^\top y\hat{\Omega}^{-1}\hat{\lambda}$;
3. $\hat{\lambda} = \left(1 + \hat{\beta}^\top x^\top x\hat{\beta}\right)\hat{\lambda}$;
4. $\hat{\theta}_j^2 = r_{jj} - \left(1 + \hat{\beta}^\top x^\top x\hat{\beta}\right)\hat{\lambda}_j^2 \quad$ for $(j = 1,\dots,p)$;
   r_{jj} is the jth diagonal element of the matrix $R = y^\top y$;
5. Repeat step 2 until 4 until the convergence.

The estimated parameters are plugging in (2.2) and η is computed for each observation i. The normality is a strong assumption, which leads to consistent, but inefficient estimators. The variances of the parameters cannot be exactly determined. This leads to more inexact indication of the parameters significance. For that reason, Khakimova et al. (2019) use two different methods to correct the estimated parameter variances. The first method from Satorra and Bentler (1994) weights the incorrect standard errors by the asymptotic covariance-matrix. The approach from Yuan and Bentler (2000) uses a kind of sandwich estimator correcting the standard errors. Determining the parameter significance, the largest corrected standard error is taken into account. For the extended research in Section 2.6.1 the approach from Satorra and Bentler (1994) is used.

2.4 Empirical Models

Having the mathematical construction of the MIMIC model in mind, Khakimova et al. (2019) look for eligible indicators y, which are affected by EA, and some regional causes x, that might have a direct influence on EA. In the age of the Internet and search engines, queries of relevant environmental phrases might be good indicator for the interests in environmental topics of a regional population. ©Google is one of the most popular search engines world wide. Currently, ©Google's market share is approximately 92.04% world wide in June 2019, following by ©Yahoo! with 2.67%, ©Bing with 2.39%, ©Baido with 0.89% and ©Yandex with 0.51%. Unfortunately, there is not currently publicly available access to the number of Internet queries as frequency data from the ©Google search engine. However, the search engine ©Yandex provides such data. Since the market share of ©Yandex is quite small world wide, Russia is used as study location. The market share of ©Yandex is measured by 50.42% in Russia (May 2019) compared with 46.06% for ©Google. In the study period, 2014 to 2016, the market share of ©Yandex was a bit smaller, approximately 41%, and ©Google a bit larger with 51%. Notwithstanding, the influence of ©Yandex is large enough to ensures statistical representativeness of the query data, if the assumption hold that the EA of the ©Yandex users is similarly to the EA of the ©Google users. For more information about the market shares of Internet search engines see StatCounter Global Stats (2019).

Following Khakimova et al. (2019) the indicators y of the MIMIC model are clustered queries of environmental phrases from the Internet search engine ©Yandex. The requested phrases are in English and Russia available.

The queries are grouped by different environmental categories:

y_1: Climate Change Queries,

y_2: Endangered Environment Queries,

y_3: Political Queries,

y_4: Science Queries,

y_5: Renewable Energies and Technologies Queries,

The *Climate Change* cluster includes words and phrases such as "global warming", "greenhouse effect", "temperature record" and others. The *Endangered Environment* cluster contains "acid rain", "forest fire", "overfishing" and others. *Political queries* summarizes words and phrases in the context of political engagement and environmental contracts such as "Asia-Pacific Partnership on Clean Development and Climate", "Kyoto Protocol" as well as "green movement", "carbon trading", etc.. The *Science queries* cluster includes words or phrases such as "geoengineering", "cloud reflectivity enhancement" as well as "El Nino" and summarizes terms from environmental and ecological science. The last group summarizes words and phrases about energy and technology like "ecologically clean energy" and "electric mobility". A list of the terms and the related categorisation can be found in the supplementary material of online version of Lösch, Okhrin, and Wiesmeth (2018a). In few cases, a clear assignment of a word or a phrase into one group is rather difficult. Another clustering might be possible. Thus, we collected the number of such phrases monthly, from January 2014 to April 2016 for each region. For comparability purposes, the indicator variables y are calculated as follows,

$$y_{in} = \frac{\text{number of queries of category } i \text{ in region } n}{\text{number of all queries in region } n}, \tag{2.6}$$

where $i = 1, \ldots, p$ denotes the indicator index with $p = 5$ categories and $n = 1, \ldots, N$ is the index of Russian region with $N = 81$. As in Khakimova et al. (2019)

the measurement model (2.1) of region n is described in terms of the EA η as follows:

$$\begin{pmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \\ \mathrm{y}_3 \\ \mathrm{y}_4 \\ \mathrm{y}_5 \end{pmatrix} = \begin{pmatrix} 1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{pmatrix} \times \eta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}. \tag{2.7}$$

The structure of the equations follows logical considerations. Of course, it could be a conceivable solution to investigate, whether these indicators can be measured by a single variable like η. For examination purposes, we perform a factor analysis for these five indicators. As Figure 4.2 in the appendix reveals, the four variables can be explained by factor 1. The fraction of the explained variance is measured by 0.41 for this factor. The Endangered Environment (EE) group is only strong related with factor 2. This factor shows a variance fraction of 0.25. However, all five categories are summarized as one single factor η, because all five categories are expression of interest and knowledge building of a regional population as well as an awareness of the environmental issues and possibilities.

Furthermore, cause variables are used to explain the EA index η. The gross regional product (GRP) per capita in purchasing power parity in first, second and third order are included capturing a possible non-linear relationship between the regional wealth and the interest in environmental topics of the population. Additional, regional characteristics about industry, social status in population and environmental situation are considered.

From (2.3) it follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{pmatrix} \underbrace{\left[\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}^{\top} \begin{pmatrix} \text{GRP per capita} \\ \text{GRP per capita}^2 \\ \text{GRP per capita}^3 \\ \text{set of control variables} \end{pmatrix} + \zeta \right]}_{\eta = \text{Index of Environmental Awareness}} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}. \tag{2.8}$$

The construction of the MIMIC model and the assumed direction of causality are shown in Figure 2.3. After estimating the weights of the causes, β, in the regression

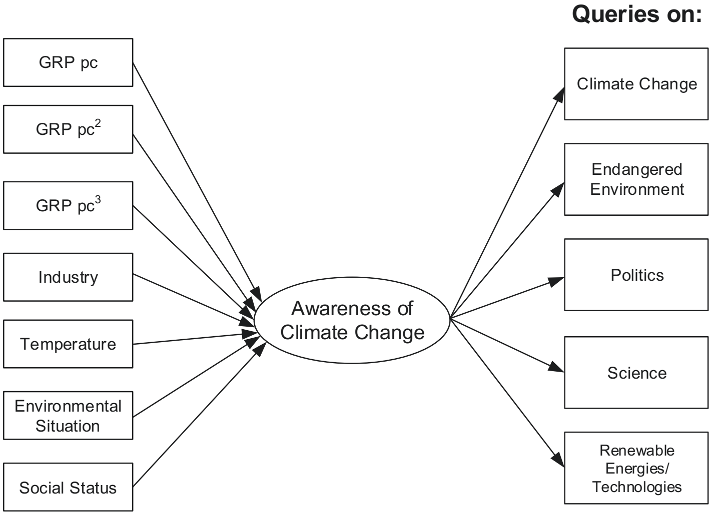

Figure 2.3: Path Diagram of the MIMIC Model inspired by Khakimova et al. (2019).

model, the environmental EA index can be computed for each region. The index allows to calculate a ranking of environmental consciousness among the regions of the Russian Federation.

Considering the EA status of regions to a specific time point by using cross sectional data, the *Default MIMIC model* is adequate to estimate the regional EA index as in Khakimova et al. (2019). However, the search engine provider of ©Yandex gives access to time series of Internet queries and thus the possibility to estimate the EA index over the course of time. Parts of this Subsection are adopted from Lösch, Okhrin, and Wiesmeth (2018a) and Lösch, Okhrin, and Wiesmeth (2017).

Since the relative number of Internet queries are subject to seasonal deviations, the primary structural part of the MIMIC model is extended through a seasonal component, which captures quarter fixed effects. Taking a year-to-year development into account, a trend component is also included. The structural model part, which explains the latent EA variable η quarterly, looks now as follow:

$$\tilde{\eta} = \beta^\top x + \gamma^\top z + \zeta, \tag{2.9}$$

Thereby, x is again a set of the over-all-years standardized cause variables. The vector z includes three binary variables for the quarter (reference is the 4th quarter) and two binary variables for the years (reference is the year 2016). Moreover, β and γ are coefficient vectors and ζ is the normally distributed random error. The MIMIC model (2.8) is extended and looks as follow:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{pmatrix} \underbrace{\left[\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}^\top \begin{pmatrix} \text{GRP per capita} \\ \text{GRP per capita}^2 \\ \text{GRP per capita}^3 \\ \text{set of control variables} \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_5 \end{pmatrix}^\top \begin{pmatrix} 1^{st} \text{ quarter} \\ 2^{nd} \text{ quarter} \\ 3^{rd} \text{ quarter} \\ 1^{st} \text{ year} \\ 2^{nd} \text{ year} \end{pmatrix} + \zeta \right]}_{\tilde{\eta}=\text{Index of Environmental Awareness}} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}. \tag{2.10}$$

The whole model is again estimated in one step, all the theory holds from the used ML estimation. For examination purposes, the average EA-index is separately estimated by each month using (2.8) and compare it with the estimation of the full-model (2.10). The results are shown in Section 2.7.

Seasonal dummies allow capturing the variation during the time, but differences between the regions are not considered. It is conceivable that the population in regions with long cold winters are particular interested in climate change, or general in environmental awareness topics, than in warmer regions. There are two possible scenarios. First, people believe in climate change and are interested in that topic, because their region is especially affected by the impact of the environmental change. Or second, they do not believe in climate change issues, because the people in the cold regions experience very cold winters. More or less, both scenarios might be displayed of interest in environmental topics and should be revealed in a larger number of internet queries. Thus, the average monthly temperature might be captured seasonal variation in time as well as differences between the geographical position of the regions. For examination purposes, the EA index time series is estimated with and without seasonal dummies and with as well as without average monthly temperature for each region.

After estimating the EA index for each region and each available time point, commonalities of regions with similar awareness structure are analysed using a k-means clustering algorithm from Hartigan and Wong (1979). The main idea is to minimize the within-cluster sum of square for a given number of cluster. The optimal number of cluster can be found by using the Silhouette method by Rousseeuw (1987) or the elbow criterion, see for example Thorndike (1953), Ketchen Jr and Shook

(1996). In addition, different region groups with similar EA - structure are composed for 2014, 2015 and 2016, separately, and their development compared over the time. This method is here called "Cluster-Break-Analysis". Furthermore, the regions are grouped by assorted characteristics, for example by GRP per capita quantile levels, and compared regarding their awareness indices.

2.5 Procedure and Data Preparation

First, the results from Khakimova et al. (2019) are presented and discussed in Section 2.6.1. For this analysis, the monthly data of the indicator (query) variables from ©Yandex are aggregated at year level, for 2014 and 2015. The data are available for each of 81 Russian regions. Beside the indicators, observable causes are needed to explain the index variable η. Since we expect a positive relationship between the wealth of a region and the EA in the population, the GRP per capita in purchasing power parity is considered. As already mentioned, the variable is included in the first, second and third order to allow for a non-linear correlation. Furthermore, in order to capture regional economic effects, we control for the structure of the industry (the share of workforce in mining and manufacturing sector). Differences in the EA might also arise from the quality of the regional environment. For that reason, regional air pollution per capita, contaminated water per square kilometre surface, as well as the identified costs for environmental protection are taken into account. In addition, controls for the infrastructure (access to the Internet, vehicle rate), agglomeration (population density), proportion of people older than 65 years, and education of the regional workforce are included. The collected dataset is in Table 4.1 in the Appendix. The data are provided by Organisation for Economic Co-operation and Development (2016), Federal Statistics Service of Russia (2016) as well as Ministry of Natural Resources of Russia. For some regional variables, data were only until 2014 available. For this reason, the data for 2015 are predicted by linear extrapolations. However, in some case there are hardly data from the past available, for that reason only data from 2014 are used. For example, since the share of employees in manufacturing sector are currently not available for 2015, we use the 2014 data for estimating the EA index in 2015. This leads to a biased awareness index for these years. However, the bias relates equally to all regions and hardly affects the ranking of the awareness index for these years.

All variables are standardized for each year separately. The standardization of the variables compensates for different scaling and helps to avoid problems in the convergence process applying the iterative ML approach. Since all of the variables are standardized, the EA index is standardized for each year as well. That means, the estimated mean of η, and thus of the EA index is 0 for each year. Consequently, the previous MIMIC model allows to evaluate a year-to-year change of the EA-position of a region taking the ranks of all other regions into account. Avoiding indeterminacy in the MIMIC-model, $\lambda_1 = 1$ fixed for all estimations. Then model 2.8 is estimated for 2014 and 2015. Making the values of the λ-parameter comparable and interpretable, all λ_i are standardized to have $\sum_{i=1}^{p} \lambda_i = 1$ after each ML estimation.

Second, seasonal effects are considered as in 2.10 by using monthly query data from ©Yandex from January 2014 to April 2016 for each of the 81 Russian regions, see also Lösch, Okhrin, and Wiesmeth (2018a). The words and phrases are again clustered as in Section 2.4 and divided by the total number of all Internet queries in region n in a certain month. The previous regional data are used as cause variables. In addition, information about nitrogen oxides and carbon dioxide are included. It allows to consider pollutants, as well as air pollutants, which are released into the atmosphere during the year from fuel combustion (for generating electricity and heat) and from stationary sources (as index values). Furthermore, the average temperature in °C for each region and each month from Meteo (2018) is included in the dataset. Since there were two events during the observation time, which might affect the interest in EA topics: the enormous Russian Rouble (RUB) devaluation and the Climate Change Conference in Paris (COP). For that reason binary variables for the RUB devaluation (=*1* from December 2014 until April 2015) and COP (=*1* from November until December 2015) are generated. Taking seasonal variations into account, the data are stacked, which leads to a sample size of $N = 81$ regions $\times$ 28 month $= 2268$ observations. Not removing the time component, all variables are here standardized over all years (all 2268 observations). Again $\lambda_1 = 1$ is fixed before each estimation. After the estimation, λ_i are standardized as above. The results are presented in Section 2.7.

The statistical language R Core Team (2016) (version 3.3.1), in particular the package Rosseel (2012), is used for the analysis of the data.

2.6 Results of the Default MIMIC Model Without Considering Seasonal Effects

After methods and data are introduced, the findings are revealed. Subsection 2.6.1 presents the results of the *Default MIMIC model* without considering seasonal effects, which are already published in Khakimova et al. (2019), p. 1427–1430. The estimated model is discussed. Afterwards, the EA index is measured as well as ranked for 81 Russian regions for 2014 and 2015 in Subsection 2.6.2, which comes from Khakimova et al. (2019), p. 1430–1432. A non-linear relationship between EA and regional wealth is revealed in Subsection 2.6.3, which can partly be found in Khakimova et al. (2019), p. 1432.

2.6.1 Model Fit and Estimated Coefficients

The *Default MIMIC model* without seasonal components is estimated with all specifications discussed in the previous Section. The results of the estimations are given in Table 2.1, as well as in Tables 4.2 and 4.3 provided in the Appendix. We consider different model structures, indicated by a capital letter, each of them estimated three times: separately for 2014 and 2015, as well as together for the two years by using a pooled data pattern. In the latter case, a dummy variable for the year 2015 is additionally included capturing year fixed effects. Tables 2.1 and 4.2 reveal the estimated parameters of the MIMIC models, which differ according to the included cause variables. The model patterns in Table 4.3 are similar to those in Table 2.1, except for the indicator variables: in view of the results from the factor analysis in Section 2.4, we also estimate η by considering only the *Climate Change*, *Science* and *Renewable Energies and Technologies* indicator variables.

The Akaike Information Criterion (AIC), Comparative Fit Index (CFI), as well as Root Mean Square Error of Approximation (RMSEA) reveal the statistical significance of the different models. Small values of AIC and RMSEA show a good model fit. Interestingly, AIC and RMSEA differ between and within the model structures (between 2014, 2015 and the pooled model). A RMSEA value < 0.1 shows a good model fit for a structural equation model, which is only the case for 2015 (A), 2014 (E), 2014 (G) and 2015 (J). It seems that the yearly EA is differently affected through the cause variables. Further information about the model fit indices can be found in Table 4.4 in the Appendix.

Various estimated parameters are fairly robust. The indicators of EA have a positive sign, as expected, or are not significant. This supports our intuition that the index stands for high interests in environmental issues. The results hold, even if we consider only three instead of five indicator variables. The standardization $\sum_{i=1}^{p} \lambda_i = 1$ facilitates the comparison between 2014 and 2015: it seems that the interests in *Climate Change* and *Renewable Energies/Technologies* topics have risen, the interest in *Endangered Environment* and *Science* has remained constant, and the interest in *Politics* has fallen. However, we cannot measure statistically significant differences of the indicator parameters between 2014 and 2015. Interestingly, the estimated parameter of the *year* dummy is negative or not significant. However, it is difficult to interpret this result, because the variable captures all unobservable differences between the years.

The first cause variable, GRP per capita, indicates a strong positive relation between regional wealth and EA. We assume a non-linear relationship between regional wealth and EA considering the EKC hypothesis in the literature. For that reason, GRP per capita^2 and GRP per capita^3 are also included as cause variables. However, the estimated coefficients reveal a monotone growing slope for the EA index function of the GRP per capita, since the values from Table 1 (pooled B) are $\beta_{\text{GRP}} = 2.430$, $\beta_{\text{GRP}^2} = -4.788$ and $\beta_{\text{GRP}^3} = 2.755$. There are no maxima or minima of the curve. Since the coefficients are standardised before the estimation, the coefficient values can hardly be interpreted. The mining variable, which controls for the importance of the mining and oil industry in each region, shows a positive sign as the volume of this sector increases. However, there is also a strong positive correlation between GRP per capita and the mining variable (Pearson = 0.710, p-value=0.000). More manufacturing implies a smaller EA, with a negative correlation with GRP per capita (Pearson = -0.523, p-value=0.000). The share in manufacturing sector and mining sector are also related with Pearson's $r = -0.518$ (p-value=0.000) as Figure 4.1 in the Appendix shows. In other words, the richer the region the higher the share in mining sector, and thus indirectly, the higher the EA index. The poorer the region the higher the share in manufacturing sector. Since the effect from mining sector as well as manufacturing sector seems not to be very robust, we suspect, there is an indirect wealth-effect on EA.

In particular, EA seems to be related to the wealth of the regions (cf. Section

2.6.3). Furthermore, we find a positive relation between the environmental quality and the EA index: a lower level of the air pollution indicates a higher EA index with a similar result for contaminated water. The effect disappears in models with many cause variables, as in model D. It might be a collinearity problem.

As already mentioned, these are only correlations and it is not possible to derive direct causalities. Perhaps, the population in regions with a high EA improved the environmental situation already in the past. This could explain the insignificant effect of the environmental protection costs. In regions with high EA and a good environmental quality there is no urgent need for further and higher environmental expenses.

Other structural factors, such as the share of old people or the labour force in the tertiary sector, seem not to be significant. Unfortunately, there is no information on the environmental education in the regions, and whether there are differences between the regions. As mentioned above, we also control for the regional access to the Internet. The coefficients are negative in some models, but mostly not significant. Interestingly, there is no linear correlation between GRP per capita and access to the Internet (Pearson = 0.153, p-value=0.172).

Table 2.1: Estimated parameters from the MIMIC model.

	2014 (A)	2015 (A)	pooled (A)	2014 (B)	2015 (B)	pooled (B)	2014 (C)	2015 (C)	pooled (C)	2014 (D)	2015 (D)	pooled (D)
λ												
	2014 (A)	2015 (A)	pooled (A)	2014 (B)	2015 (B)	pooled (B)	2014 (C)	2015 (C)	pooled (C)	2014 (D)	2015 (D)	pooled (D)
Climate Change	0.258	0.335	0.258	0.256	0.333	0.274	0.256	0.304	0.27	0.256	0.304	0.274
Endangered Environment	0.135**	0.121**	0.145**	0.136**	0.12**	0.134**	0.134**	0.123**	0.125**	0.135	0.115**	0.115**
	(0.233)	(0.173)	(0.231)	(0.247)	(0.155)	(0.215)	(0.256)	(0.158)	(0.181)	(0.27)	(0.163)	(0.179)
Politic	0.107**	0.059	0.088**	0.103**	0.057	0.088**	0.102**	0.063	0.087**	0.102**	0.066	0.09**
	(0.161)	(0.107)	(0.108)	(0.164)	(0.095)	(0.107)	(0.152)	(0.126)	(0.119)	(0.153)	(0.139)	(0.126)
Science	0.34***	0.252**	0.311**	0.348**	0.257**	0.296**	0.345***	0.261**	0.283***	0.345**	0.262**	0.276***
	(0.368)	(0.257)	(0.389)	(0.431)	(0.268)	(0.346)	(0.335)	(0.279)	(0.257)	(0.436)	(0.273)	(0.235)
Renewable Energies/Technologies	0.16**	0.233***	0.198***	0.157**	0.233***	0.208***	0.163**	0.248***	0.234***	0.162**	0.253**	0.246***
	(0.228)	(0.164)	(0.169)	(0.228)	(0.159)	(0.159)	(0.238)	(0.229)	(0.197)	(0.26)	(0.26)	(0.235)
β												
GRP per capita in ppp	0.394***	0.315**	0.307***	2.1	3.006**	2.43**	1.993**	3.29**	2.74***	2.113	3.178***	2.713***
	(0.094)	(0.129)	(0.092)	(1.172)	(0.93)	(0.781)	(1.008)	(1.2)	(0.761)	(1.091)	(0.9)	(0.585)
GRP per capita^2 in ppp				-4.824**	-5.751**	-4.788**	-6.122**	-7.79**	-6.858***	-6.551**	-8.255**	-7.31***
				(2.429)	(2.305)	(1.723)	(2.562)	(2.986)	(1.915)	(2.933)	(2.514)	(1.633)
GRP per capita^3 in ppp				3.238**	3.126**	2.755**	4.621**	5.052**	4.647***	4.926**	5.6**	5.104***
				(1.447)	(1.476)	(1.065)	(1.763)	(2.036)	(1.319)	(2.056)	(1.785)	(1.169)
Access to Internet	-0.283**	-0.225	-0.235**	-0.244	-0.292**	-0.292**	-0.058	-0.048	-0.063	-0.065	-0.074	-0.09
	(0.127)	(0.185)	(0.109)	(0.151)	(0.142)	(0.104)	(0.076)	(0.109)	(0.071)	(0.075)	(0.102)	(0.065)
Mining							0.527**	0.38	0.381**	0.503**	0.427	0.4**
							(0.193)	(0.246)	(0.16)	(0.229)	(0.274)	(0.186)
Manufacturing							-0.09	-0.318**	-0.276**	-0.054	-0.153	-0.169**
							(0.093)	(0.115)	(0.091)	(0.078)	(0.091)	(0.063)
Air pollution per capita							-0.472**	-0.681**	-0.607***	-0.468	-0.749**	-0.666***
							(0.22)	(0.25)	(0.17)	(0.269)	(0.236)	(0.163)
Environmental Protection Costs							-0.001	-0.136	-0.134**	-0.03	-0.171	-0.148
							(0.054)	(0.112)	(0.064)	(0.078)	(0.14)	(0.078)
Contaminated Water							-0.176**	-0.16	-0.111	-0.197	-0.276	-0.128
							(0.073)	(0.108)	(0.065)	(0.101)	(0.213)	(0.094)
Population Density										0.002	0.15	0.018
										(0.068)	(0.169)	(0.072)
People with Age +65										-0.058	-0.296	-0.247
										(0.152)	(0.207)	(0.143)
Labour Force with Tertiary Education										0.107	0.099	0.09
										(0.055)	(0.08)	(0.05)
Vehicle Rate										-0.001	0.233	0.187**
										(0.127)	(0.151)	(0.092)
Year Dummy (2015==1)			-0.109			-0.088			-0.127**			-0.114**
			(0.064)			(0.074)			(0.046)			(0.044)
Observations	81	81	162	81	81	162	75	75	150	75	75	150
Degrees of Freedom	13	13	17	21	21	25	41	41	45	57	57	61
AIC	1012	1061	2094	998	1043	2067	896	943	1855	900	939	1846
CFI	0.874	0.94	0.809	0.802	0.937	0.782	0.697	0.755	0.699	0.62	0.702	0.644
RMSEA	0.143	0.078	0.137	0.155	0.069	0.13	0.163	0.122	0.137	0.164	0.121	0.136

Standard errors in parentheses; significance level: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

2.6.2 A Ranking of Environmental Awareness in Russia

The construction of the EA index gives us some new insight into the environmental situation of the Russian regions. Given the estimated EA index for each region and each year, a ranking of these regions in terms of their environmental consciousness is possible. We also provide the ranking of the Investment Index regarding ecological risks in Russian regions mentioned in Section 2.2.1.

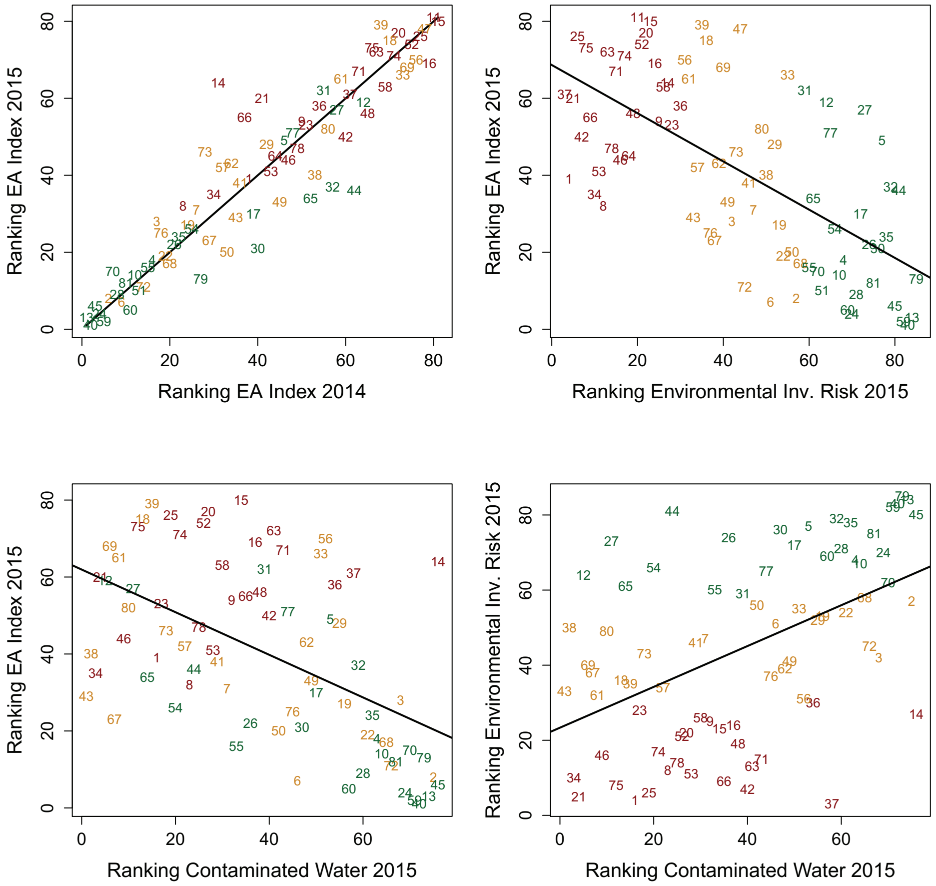

Figure 2.4: Left above: relationship between the ranking of the EA index in 2014 and 2015 by using model B (pooled) in Table 2.1 for estimating the EA index. Right above: negative relation between the ranked regions corresponding to the Investment Index in 2015 and estimated EA index in 2015. Left below: relation between the ranking of the EA index and regional contaminated water per km^2 in 2015. Right below: relation between the ranking according to the Investment Index and contaminated water in 2015. The pictures right above, left below and right below illustrate also the corresponding linear regression line. The colours comply with the environmental investment risk (red: high, orange: middle, green: low environmental investment risk). The region with the highest EA is on place 1. Analogously, the region with the highest environmental investment risk is on the 1st rank. The Ids corresponds to the regional keys in Table 4.5 in the Appendix.

As seen from Figure 2.4 (left above), there is a linear relationship between the ranking of the EA index in 2014 and 2015. The results hardly change from period to period and they seem to be robust. The regions with the highest EA in 2014, Chukotka AO (13), and in 2015, Magadan (40) are on place 1. A highly ranked region in 2014 is most likely well-placed in 2015, as well. Nevertheless, there are some regions, which occupy a lower rank in 2015: the regions Chuvashia (14), Kaliningrad (21), and Tambovsk (66). In contrast, Leningrad with St. Petersburg (38), Komi (32), Murmansk (44), and Sverdlovsk (65) reveal a higher EA in 2015. The highest EA index values are observed in relatively rich regions, such as Chukotka AO (13), Magadan (40), Nenets AO (45), Kamchatka (24), and Sakha (59), located more or less in the North of Russia. The regions with the smallest EA values belong to the poor and conflictual Caucasus area, such as Karachay-Cherkessia (25), North Ossetia-Alania (47), Ingushetia (16), Chechnya (11), and Dagestan (15).

Figure 2.4 (right above) resembles that regions, which are ranked high according to EA tend to be ranked low regarding the ecologically based investment risk developed by Expert RA (2018). The Spearman's ρ between this Environmental investment risk ranking and the estimating ranking of the EA index amounts -0.647 (p-value = 0.000) for 2015, see Table 2.3. The higher the environmental investment risk, the worse the quality of environment and the higher the costs for corresponding hazardous waste disposal. Interestingly, in this case our forward oriented EA index corresponds to the Investment Index based on ecological risks resulting from environmental pollution in the past.

According to Figure 2.4 (left below), there is a negative relationship between regional contaminated water per km^2 and the EA index (see in Table 2.3 Spearman's ρ = -0.563, p-value = 0.000 for 2014 and ρ = -0.539, p-value = 0.000 for 2015). A similar result holds for air pollution and EA. It is conceivable that due to a high EA an increasing pressure on local politics in the past has helped to improve the environmental situation. In view of the remarks on ecological investment risks, a comparable result holds for the reaction between the ranking according to the ecological investment risks and contaminated water per km^2 (see Figure 2.4, right below).

The colours in all four pictures correspond to the ranking regarding the ecological investment risk (red: high, orange: middle, green: low environmental investment risk). The calculated ranks of the EA indices for 2014 and 2015 are illustrated as maps (cf. Figure 4.3 and Figure 4.4 in the Appendix).

Table 2.2: The EA index and the corresponding ranking for 81 Russian regions for the years 2014 and 2015. The keys correspond to the Ids in Table 4.5 in the Appendix. The EA indices are estimated by using model *pooled (B)* from Table 2.1.

Key	Region	Index '14	Rank '14	Index '15	Rank '15	Key	Region	Index '14	Rank '14	Index '15	Rank '15
13	Chukotka AO	331.0	1	191.3	3	29	Khakassia	-3.7	42	-34.6	48
40	Magadan	311.1	2	280.2	1	53	Penza	-5.4	43	-29.7	41
45	Nenets AO	273.3	3	60.1	6	64	Stavropol	-7.5	44	-30.8	45
24	Kamchatka	223.8	4	172.3	4	49	Novosibirsk	-7.6	45	-21.3	33
59	Sakha	213.3	5	192.6	2	5	Arkhangelsk	-8.8	46	-35.4	49
2	Altai	188.1	6	41.7	8	46	Nizhny Novgorod	-8.8	47	-30.3	44
70	Tuva	89.2	7	26.3	15	77	Vologda	-9.4	48	-37.2	51
28	Khabarovsk	76.8	8	39.9	9	78	Voronezh	-9.5	49	-33.6	47
6	Astrakhan	71.9	9	45.9	7	9	Bryansk	-11.3	50	-38.4	54
81	Zabaykalsky	61.1	10	30.2	12	23	Kaluga	-11.6	51	-38.0	53
60	Sakhalin	59.2	11	71.5	5	65	Sverdlovsk	-13.4	52	-22.5	34
10	Buryatia	58.9	12	27.4	14	38	Leningrad/St.Petersburg	-15.7	53	-28.7	40
51	Orenburg	54.3	13	33.1	10	36	Kurgan	-16.0	54	-41.8	58
72	Tyumen	45.9	14	31.6	11	31	Kirov	-18.0	55	-50.6	62
55	Primorsky	43.2	15	19.1	16	80	Yaroslavl	-18.6	56	-37.3	52
4	Amur	37.6	16	8.0	18	32	Komi	-19.1	57	-26.7	37
3	Altai	34.2	17	-11.9	28	27	Kemerovo	-20.7	58	-40.1	57
76	Volgograd	33.0	18	-6.4	25	61	Samara	-22.2	59	-53.8	65
22	Kalmykia	30.8	19	5.5	19	42	Mordovia	-23.3	60	-36.5	50
68	Tomsk	28.3	20	15.2	17	37	Kursk	-25.6	61	-45.4	61
26	Karelia	27.4	21	-2.2	22	44	Murmansk	-26.0	62	-24.3	36
35	Krasnoyarsk	22.8	22	-5.8	24	71	Tver	-27.4	63	-55.4	67
8	Belgorod	15.2	23	-21.1	32	12	Chelyabinsk	-28.7	64	-42.2	59
19	Jewish	14.5	24	-11.8	27	48	Novgorod	-30.6	65	-39.8	56
54	Perm	14.2	25	-6.7	26	75	Vladimir	-31.0	66	-62.6	73
7	Bashkortostan	12.7	26	-19.4	31	63	Smolensk	-34.3	67	-58.0	72
79	Yamalo-Nenets AO	11.0	27	28.0	13	39	Lipetsk	-34.3	68	-76.3	79
73	Udmurtia	10.2	28	-31.2	46	58	Ryazan	-36.2	69	-52.2	63
67	Tatarstan	8.0	29	-5.4	23	18	Ivanovo	-39.1	70	-66.5	75
34	Krasnodar	6.7	30	-22.9	35	74	Ulyanovsk	-46.1	71	-57.9	71
14	Chuvashia	6.6	31	-52.4	64	20	Kabardino-Balkaria	-48.5	72	-72.9	77
57	Rostov	5.5	32	-29.9	42	33	Kostroma	-48.5	73	-53.8	66
50	Omsk	4.9	33	1.6	20	69	Tula	-52.9	74	-55.8	68
62	Saratov	3.2	34	-30.0	43	52	Oryol	-54.7	75	-65.3	74
43	Moscow	1.1	35	-16.2	29	56	Pskov	-56.8	76	-57.7	70
41	Mari El	-0.5	36	-27.9	38	25	Karachay-Cherkessia	-77.5	77	-71.5	76
66	Tambovsk	-0.7	37	-39.0	55	47	North Ossetia-Alania	-79.8	78	-75.9	78
1	Adygea	-1.6	38	-28.2	39	16	Ingushetia	-90.0	79	-56.4	69
17	Irkutsk	-2.9	39	-17.5	30	11	Chechnya	-91.3	80	-126.3	81
30	Khanty-Mansi AO	-3.4	40	-1.7	21	15	Dagestan	-96.9	81	-118.1	80
21	Kaliningrad	-3.4	41	-42.3	60						

Correlation between $\hat{\eta}$ and ...		Pearson's cor		Spearman's rho	
		coef.	p - value	coef.	p - value
GRP per capita	2014	0.434	0.000	0.414	0.000
	2015	0.353	0.001	0.532	0.000
Environment Protection Costs	2014	-0.155	0.176	-0.338	0.002
	2015	-0.147	0.199	-0.396	0.000
Contaminated Water	2014	-0.266	0.019	-0.563	0.000
	2015	-0.161	0.162	-0.539	0.000
Environmental Inv. Risk Rank	2015			-0.647	0.000

Table 2.3: Pearson's and Spearman's correlation coefficient for measuring linear as well as monotone relationship between the EA index and GRP per capita, Environment protection costs as well as Contaminated Water in 2014 and 2015. The EA indices are estimated by using model *pooled (B)* in Table 2.1.

2.6.3 Dependence of EA on GRP per capita

The results of the estimation yield a curvilinear relationship between GRP per capita and EA in the regions: the parameter values β_1 for (GDP per capita), β_2 for (GDP per capita)^2, and β_3 for (GDP per capita)^3 with $\beta_1 > 0$, $\beta_2 < 0$, and $\beta_3 > 0$ are significant in almost all estimated model variations. Nevertheless, the relationship might be rather linear, especially for poorer regions.

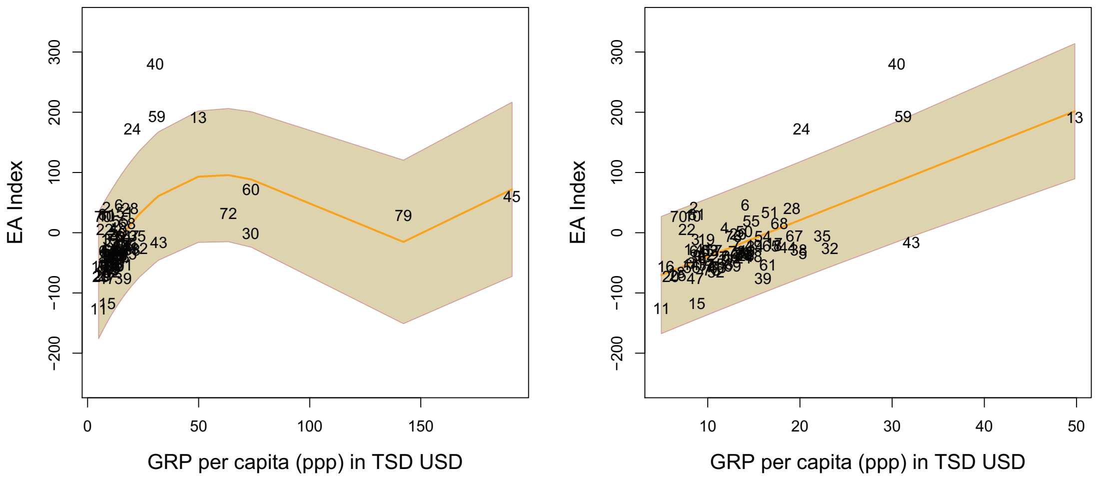

Figure 2.5: Relationship between GRP per capita in purchasing power parity in thousand USD and the EA index using local polynomial regression fit (left, orange) as well as linear regression fit (right, orange), 95% pointwise confidence interval (orange area). The keys correspond to the Ids in Table 4.5 the Appendix. The EA indices are estimated by using model *pooled (B)* in Table 2.1.

The average regional EA indices are plotted in Figure 2.5 against the GRP per capita (in purchasing power parity in 2014) of the Russian regions. The curve on the left hand side is fitted by local polynomial regression enclosed by a 95% point-wise confidence interval. On the right hand side the scatter plot is constrained by regions with a GRP per capita smaller than 50 thousand USD in 2014, the five richest regions Khanty-Mansi AO (30), Nenets AO (45), Sakhalin (60), Tyumen (72) and Yamalo-Nenets AO (79) are excluded. Thus, we cannot confirm the EKC hypothesis in general, but we find a positive correlation between awareness in environmental topics and regional wealth, see also Table 2.3.

There is a moderate relationship between the EA index and GRP per capita. For the data in 2014, the Pearson's correlation coefficient is approximately similar to the Spearman's ρ. In 2015 the Spearman's coefficient for monotone correlation is larger than the Pearson's coefficient for a linear relation. On the one hand, there is the leverage effect of the outliers. There are some very rich regions with rather small EA index values, as Nenets AO, Yamalo-Nenets Autonomous Obkrug (AO), Khanty-Mansi AO and Sakhalin. For that reason, a linear relationship between wealth and the regional interest in environmental topics is rather conceivable for the poorer regions, as Figure 2.5 shows. In Russia, many regions obtain a GRP per capita less than 50 Tsd. USD in 2014. A few of them generate more such as Chukotka AO (13), Khanty-Mansi AO (30), Sakhalin (60), as well as Yamalo-Nenets AO (79) and Nenets AO (45). The last two mentioned regions are especially rich and generate the steep rising slope in the curve on the left hand side picture. Regarding poorer regions like the Kamchatka (24), Oblast Magadan (40) and Sakha Republic (59) are above the 95% Confidence Interval. These regions have above average high EA Index related to their GRP per capita. Kamchatka, Oblast Magadan and Sakha Republic are neighbours positioned in the far East of the Russian Federation. The Sakha Republic is the largest federal area in Russia with almost the geographic size of central Europe. It is part of the central Siberian mountain region bordered by the Arctic Ocean in the North and the Aldan Highlands in the South. Oblast Magadan is situated in the East of Sakha Republic and part of the East Siberian mountain region bordered by the Sea of Okhotsk in the West. Kamchatka is the largest peninsula in the Eastern part of Asia. It is bordered by Chukotka to the north and Magadan to the east. On the peninsula are 29 active volcanoes and many geysers. Large parts of Kamchatka belongs to of the UNESCO world heritage site. All three areas are very thin populated and most of the locals live in cities. The regions are very rich in raw materials (precious metal, diamonds, mineral oil, gas, coal). The locals live especially from the mining sector, but as well from fishing (especially in Kamchatka), forestry and energy production.

Some regions, such as Oblasts Ivanovo (18), Kostromo (33) and Kermerovo (27), lie on the regression fit in Figure 2.5 (right above). Oblasts Ivanovo and Kostroma are part of Central Russia. Oblast Ivanovo is situated in the North East of Moscow and lies in the East European lowland. It is bordered by Kostroma Oblast in the North. Both regions live from textile production, but also from chemical and food processing industry. The tourism sector plays for Kostroma Oblast an economic role, too. Oblast Kermerovo lies between the West Siberian lowland and the South Siberian mountains. It is one of the most dense areas in Siberia. The locals live especially from the coal production, but also from metal processing, chemical industry and steel production, see also Kerneck and Oertel (2016).

Notwithstanding, the geographical position seems to play a role as well. The largest EA indices can be measured in the East of Siberia. This comes to no surprise, since the locals there are said to be close to nature.

2.7 The Results of the Extended MIMIC Model with Seasonal Variation

The current section shows the estimated EA index for the extended model by considering seasonal effects. Moreover, the advantages of using temperature data instead of seasonal dummies capturing seasonal effects are discussed. In addition, some events, such as the Rubel devaluation (RUB) at the end of 2014 and beginning of 2015 as well as the Climate Change Conference (COP) in December 2015 are investigated, which could influence the interest in environmental topics. The findings are already published in Lösch, Okhrin, and Wiesmeth (2018a) and Lösch, Okhrin, and Wiesmeth (2017).

The last part of this section addresses the EA situation in the Arctic and Subarctic regions and is published in Lösch, Okhrin, and Wiesmeth (2018b). All three publications focus on the awareness of climate change, which belongs to environmental issues. The *Environmental Awareness Index* is there called *Climate Change Index*, because of the strong correlation between environmental, especially climate, queries and regional temperature, which is going to be shown later in detail. Staying in line with concept of the chapter of this thesis, *Environmental Awareness Index* (EA) is further on used instead of *Climate Change Index* (CC) as in the published articles.

In order to investigate the development of EA in the Russian regions, the index $\tilde{\eta}$ is estimated through the MIMIC model (2.10) considering variation in time for the 2268 stacked data as described in 2.5. The estimated coefficients are shown in Table 2.4. All models from m_1 to m_{11} include the five indicators y and the GRP per capita-cause variable x. We consider either GRP per capita, (GRP per capita)^2 and (GRP per capita)^3 or log(GRP per capita) capturing the expected non-linear relationship between income and the regional EA. The binary variables z for the time components: quarters (z_1, z_2 and z_3) and years (z_4, z_5) are included in model m_4, m_5 and m_{12}. The other models consider the temperature for capturing seasonal and spatial differences. In addition, binary variables capturing geographical affiliation are included in model m_5, m_6, m_{10} and m_{11}. Further cause variables follow as mentioned above.

For examination purposes, we also estimate the EA index separately for each

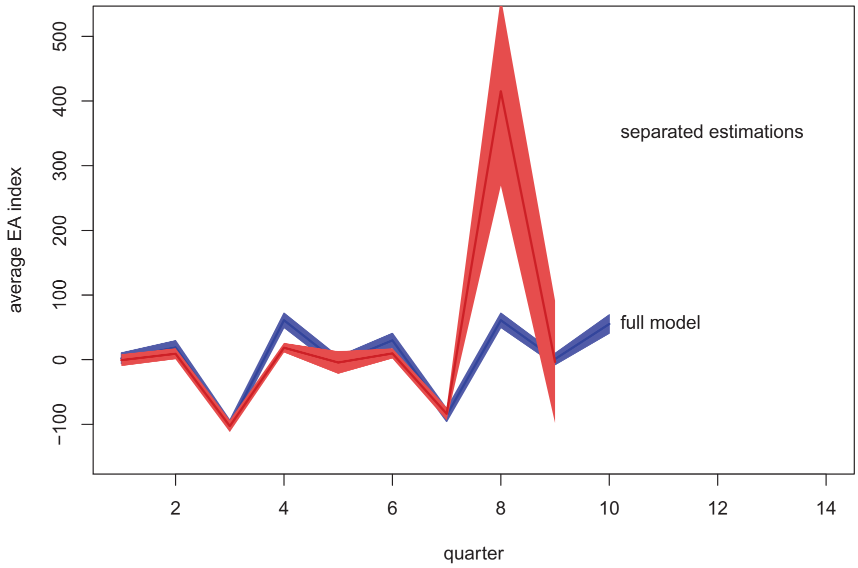

Figure 2.6: Concurrent (blue; stacked data) and sequential estimation of the quarter EA index by using m_3 in Table 2.4.

quarter by using model m_3 without quarter dummies in Table 2.4 and compare it with the joint estimation of m_3 using the stacked data (including all quarter) with quarter dummies. As Figure 2.6 shows, with the exception of quarters four and eight, there are hardly differences between the average quarter EA indexes of both methods. Only there are some convergence problems for the separate estimation the average EA index of the 8th quarter (red peak). For that reason, a joint estimation of the pooled data is rather suitable.

The reduced model m_1 in Table 2.4 shows the lowest Akaike information criterion (AIC) value and the lowest robust Root Mean Square Error of Approximation (RMSEA), thus m_1 has the best model fit. The model fit of model m_2 hardly differ from m_1. Both models only differ in the form of the GRP variables. Interestingly, the models, which include temperature variables instead of time dummies, reveal a better model fit. Thus, the temperature might adequate capture seasonal deviations and, additional, spatial conditioned climate characteristics. Information about the used model fit measurements are in Table 4.4 in the Appendix.

A look at signs and variances of the estimated coefficients might be very interesting. The coefficient of the temperature is highly negatively significant and very robust over all 11 model variations, even if we additionally consider time dummies as model m_4 shows. This result confirms our expectation of stronger EA interest in colder regions. If the temperature variable is included instead of time dummies, the coefficient of the COP dummy becomes positively significant. If we could analyse a longer time-series, it might be easier to differ between seasonal effects and the actual influence of the event. However, the first results meet our expectations. Especially, the coefficient for the RUB devaluation event, approximately one year before the COP, is in almost all models insignificant. All λs are significantly positive. Thus, as expected, there is a positive relationship between the index and environmental queries. The most interesting Internet queries seem to belong to the *Climate Change* group, because λ_1 shows the highest coefficient value in each model. Beside the EA queries, all the other indicator variables include the relative number of phrases, which are closed to the EA topics, too. Thus, the awareness index can be interpreted as high interest in the environment, especially in EA.

To sum up, model m_1 fits the data well enough. The other cause variables, besides the GRP per capita, the temperature and COP dummy, seem to have a weak impact of the awareness index. The most of them have significant coefficients, but hardly affect the model fit.

Nonetheless, it is worthwhile looking at the signs of the other coefficients. As before, the coefficients of the GRP per capita are positive in the first and third order, β_1 and β_3, but negative in the second order β_2. This suggests a curvilinear relationship between wealth of a region and the awareness index. The models with log(GRP per capita) show a positive sign for its highly significant coefficient, which supports this hypothesis. From the geographic aspect, the coefficient value of the North-Eastern Europe dummy is negative significant compared with the reference (Far) Eastern Siberia. A overview of the geographical groups can be found in Table 4.7 in the Appendix.

Moreover, other socio-economic causes such as the share of employees in manufacturing or the rate of unemployment have a certain influence on awareness. Consequently, as in Khakimova et al. (2019), promising efforts to stimulate EA are fundamentally related to the economic system.

Table 2.4: Results for Eleven Different Variations of the Extended MIMIC Model.

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}
λ_1 Climate Change	0.391	0.394	0.487	0.467	0.365	0.375	0.320	0.318	0.318	0.315	0.389
λ_2 Endangered Environment	0.102***	0.106***	0.092***	0.100***	0.104***	0.109***	0.088***	0.091***	0.088***	0.097***	0.124***
	(0.034)	(0.035)	(0.026)	(0.028)	(0.034)	(0.034)	(0.042)	(0.041)	(0.043)	(0.046)	(0.037)
λ_3 Politic Queries	0.118***	0.117***	0.109***	0.113***	0.104***	0.107***	0.125***	0.125***	0.121***	0.105***	0.106***
	(0.041)	(0.040)	(0.032)	(0.031)	(0.043)	(0.041)	(0.056)	(0.055)	(0.055)	(0.054)	(0.038)
λ_4 Renewable Energy	0.216***	0.209***	0.173***	0.178***	0.251***	0.233***	0.300***	0.301***	0.312***	0.329***	0.238***
	(0.052)	(0.050)	(0.036)	(0.038)	(0.052)	(0.049)	(0.069)	(0.069)	(0.070)	(0.070)	(0.049)
λ_5 Science Queries	0.174***	0.174***	0.138***	0.142***	0.175***	0.176***	0.166***	0.165***	0.160***	0.154***	0.144***
	(0.046)	(0.045)	(0.027)	(0.028)	(0.046)	(0.044)	(0.066)	(0.065)	(0.065)	(0.064)	(0.039)
β_1 GRP per capita	1.329***		2.006***	1.306***	1.326***		1.043***	1.076***	0.856***	0.985***	1.138***
	(0.181)		(0.192)	(0.178)	(0.183)		(0.159)	(0.194)	(0.186)	(0.187)	(0.209)
β_2 GRP per capita^2	-2.992***		-4.339***	-2.989***	-3.049***		-2.860***	-2.906***	-2.793***	-3.099***	-3.429***
	(0.436)		(0.474)	(0.447)	(0.424)		(0.410)	(0.461)	(0.469)	(0.484)	(0.527)
β_3 GRP per capita^3	1.816***		2.609***	1.844***	1.850***		1.754***	1.776***	1.838***	1.995***	2.217***
	(0.281)		(0.311)	(0.297)	(0.268)		(0.256)	(0.279)	(0.300)	(0.308)	(0.342)
β_4 log(GRP per capita)		0.163***				0.152***					
		(0.024)				(0.026)					
β_5 Temperature	-0.415***	-0.425***		-0.514***	-0.376***	-0.389***	-0.341***	-0.334***	-0.311***	-0.289***	-0.409***
	(0.024)	(0.024)		(0.039)	(0.025)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)	(0.043)
β_6 Climate Conference Paris	0.290***	0.293***		0.076	0.381***	0.412***	0.171*	0.176*	0.187**	0.222**	0.232*
	(0.111)	(0.112)		(0.132)	(0.108)	(0.111)	(0.092)	(0.092)	(0.091)	(0.092)	(0.129)
β_7 Rubel devaluation	-0.054	-0.059		0.018	0.062	0.075	-0.065*	-0.062	-0.053	-0.015	-0.002
	(0.043)	(0.043)		(0.059)	(0.050)	(0.050)	(0.039)	(0.039)	(0.039)	(0.046)	(0.058)
β_8 North-Eastern Europe					-0.271***	-0.247***				-0.231***	-0.141***
					(0.025)	(0.025)				(0.031)	(0.034)
β_9 South-Eastern Europe					0.035	-0.005				0.064*	0.089
					(0.041)	(0.041)				(0.037)	(0.069)
β_{10} Western Siberia					-0.037	-0.097**				0.235***	-0.151***
					(0.043)	(0.046)				(0.057)	(0.034)
β_{11} Manufacturing							-0.248***	-0.236***	-0.232***	-0.189***	-0.025
							(0.023)	(0.027)	(0.030)	(0.029)	(0.038)
β_{12} Mining							0.123**	0.124**	0.039	0.028	-0.116***
							(0.061)	(0.062)	(0.064)	(0.063)	(0.030)
β_{13} Fishery								0.013	-0.019	-0.058**	-0.312***
								(0.025)	(0.024)	(0.025)	(0.069)
β_{14} Air pollution							-0.156***	-0.195***	-0.195***	-0.185***	-0.112
							(0.024)	(0.027)	(0.028)	(0.029)	(0.074)
β_{15} Carbon dioxide								0.057**	0.061**	0.029	0.223***
								(0.026)	(0.027)	(0.031)	(0.076)
β_{16} Nitrogen dioxide								-0.002	-0.006	-0.088***	-0.017
								(0.019)	(0.019)	(0.024)	(0.031)
β_{17} Unemployment rate									-0.159***	-0.161***	-0.167***
									(0.037)	(0.035)	(0.046)
β_{18} Share of old people									-0.205**	-0.195**	-0.067
									(0.082)	(0.082)	(0.098)
β_{19} Share of young people									0.026	0.013	-0.194**
									(0.082)	(0.081)	(0.096)
β_{20} Education									-0.043	-0.068	-0.045
									(0.052)	(0.051)	(0.055)
γ_1 January – March			-0.248***	-0.551***							-0.301***
			(0.038)	(0.043)							(0.056)
γ_2 April – June			-0.023	0.443***							0.398***
			(0.051)	(0.064)							(0.071)
γ_3 July – September			-1.300***	-0.568***							-0.531***
			(0.061)	(0.082)							(0.085)
γ_4 Year 2014			0.343***	0.136***							0.230***
			(0.045)	(0.048)							(0.056)
γ_5 Year 2015			0.382***	0.203***							0.208***
			(0.042)	(0.054)							(0.059)
N	2268	2268	2268	2268	2268	2268	2268	2268	2268	2268	2268
AIC	40040	43993	43927	46629	44940	49136	53751	66680	77568	78446	84416
RMSEA	0.057	0.059	0.093	0.089	0.107	0.119	0.077	0.073	0.074	0.083	0.098

Standard errors in parentheses; significance level: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Furthermore, we investigate the dependence of EA in the various regions of the Russian Federation on certain regional characteristics. Also, the possible effects of the RUB devaluation end of 2014 and their dependence on regional aspects are of interest, as well as the COP end of 2015.

2.7.1 Regional Differences of EA

Russian regions differ substantially regarding GRP per capita, the share of manufacturing, level of air pollution, and the share of the older generation, just to name a few. Using estimation results from version m_1 MIMIC model (2.10), we again aim at investigating, how these additional causes variables affect EA. Apart of this we are also interested in the effect of temperature on EA. The regions are grouped into four clusters according to the size of the index of EA by using the k-means algorithm from Hartigan and Wong (1979).

As Figure 2.7 shows, there is a linear relationship between the average temperature and the average EA index. This result confirms the expectations of the high interests in EA topics in colder than in warmer areas. However there is a larger temperature variation in regions with higher EA indices. This result might be driven by the other cause variables, such as GRP per capita.

Furthermore, Figure 2.8 refers to the relevance of the share of employees in manufacturing in the various regions. Interestingly, we observe an increase of awareness with a rising share of manufacturing in the lowest cluster (red), and a decrease in the other cluster. Regions with a relatively low share of employees in the manufacturing sector and low EA indices, such as Dagestan, Cechnya and Ingushetia in the lowest (red) cluster, have a strong agriculture sector and are very traditional. The population seems not to be so interested in more "modern" issues such as EA. The population in (higher) cluster (grey and blue) with a relatively weak manufacturing sector and high EA index values are rather richer regions, which live especially from oil-extracting industry (for example Nenets Autonomous Okrug, Yamalo-Nenets Autonomous Okrug). The regions with a relatively high share of employees in manufacturing and low EA indices are rather poor. There is a relation between GRP per capita and the share of employees in manufacturing sector.

Surprisingly, the situation is quite different when we look at the effect on awareness of higher levels of "air pollution" in Figure 2.9. Especially the second and third cluster with a large variance of these emissions show no sensitivity regarding awareness with respect to increasing levels. Similar results hold for the influence of the rate of unemployment, the share of employees in mining, and the share of older people in the population of the regions. This result is also reflected through some unstable coefficients (changing signs)

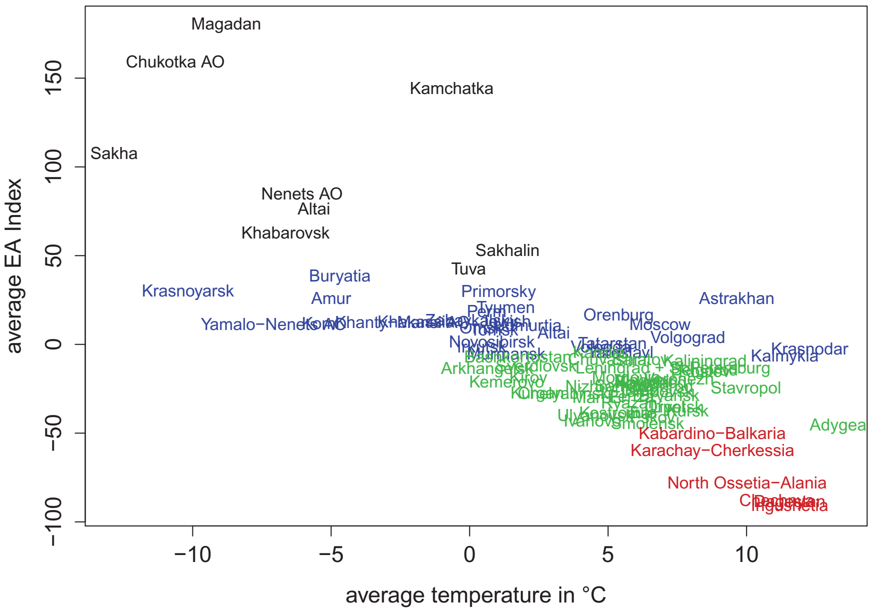

Figure 2.7: Estimated average awareness index depending on the average temperature for the regions assembled in four clusters according to the level of index. The affiliation of the regions to the cluster can be found in Table 4.6 in the appendix. Model m_1 is used for estimating the EA index.

in Table 2.4.

2.7.2 Devaluation of RUB and Climate Change Conference

The substantial fall of the world market prices of natural oil and gas starting end of 2014 implies the decline of the RUB-Dollar exchange rate and other negative consequences for the Russian economy, which is largely dependent on the export of oil and gas. Moreover, there might be environmental issues associated with these developments: efforts to reduce the consumption of oil and to mitigate climate change might be weakened.

On the other hand, end of November 2015 the COP inspired many countries to contribute towards mitigating EA and support efforts to adapt to consequences of climate change, especially in the developing countries.

Thus, for both these issues the question arises, whether there are also measurable effects regarding EA, perhaps dependent on the characteristics of certain regions of the Russian Federation.

Again, the estimation of version m_1 of model (2.10) yields interesting results regarding awareness in different contexts. Figure 2.10 shows the time-series of the EA (left) and the average temperature (right) per month and region, whereby the regions are clustered

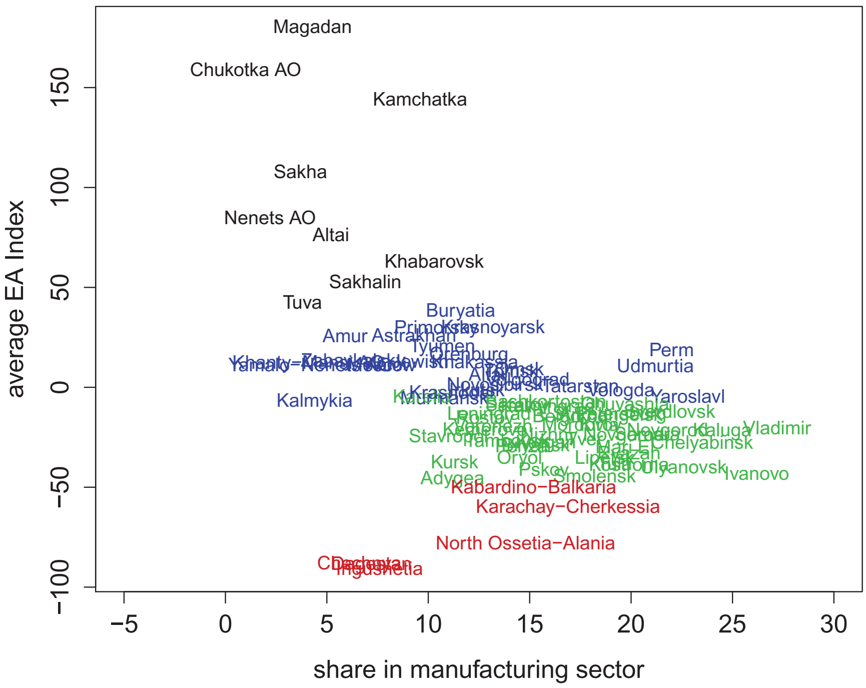

Figure 2.8: Estimated average awareness index depending on the share of employees in manufacturing for the regions assembled in four clusters according to the level of the index. The affiliation of the regions to the cluster can be found in Table 4.6 in the appendix. Model m_1 is used for estimating the EA index.

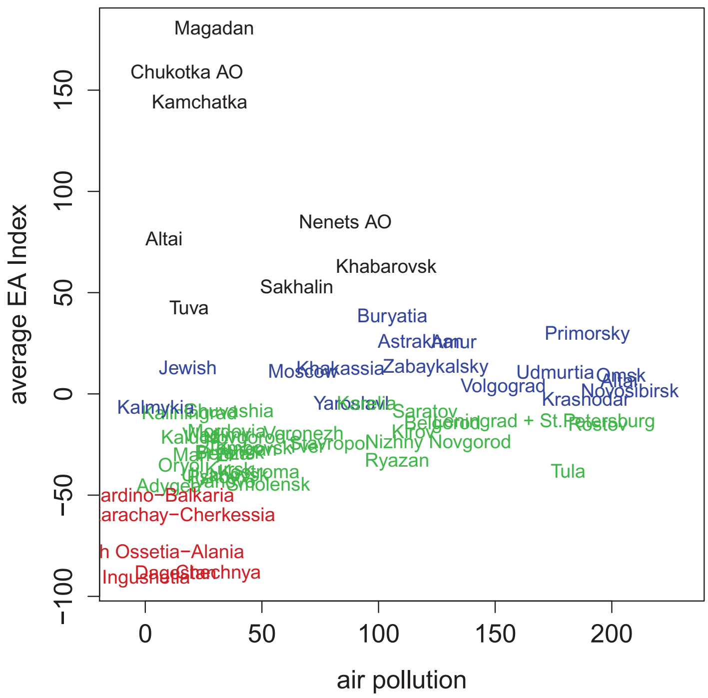

Figure 2.9: Estimated average awareness index depending on per capita greenhouse gas emissions for the regions assembled in four clusters according to the level of the index. The affiliation of the regions to the cluster can be found in Table 4.6 in the appendix. Model m_1 is used for estimating the EA index.

regarding their awareness index (left) or their average temperature by using a k-means algorithm. The colours reveal the rank related to the EA index: red < green < blue < black (grey). The coloured areas are the 95% Confidence Intervals. Table 4.6 in the appendix lists the corresponding regions within the EA index cluster. The pictures in Figure 2.10 confirm the relationship between EA and temperature as in Figure 2.7 shown. In addition, Figure 2.10 reveals seasonal variations in the EA, which are related to the monthly temperature. The higher the temperature, the less the interests in EA topics. The results remain, if we cluster the regions for each year separately. Figure 2.11 shows the result of this "Cluster-Break-Analysis". From the algorithm of clustering, it can happen that cluster g includes a different set of regions in 2014 as in 2015, but the within-cluster variances are minimized for each year and each cluster. It seems that, the currency devaluation at the end of 2014 until the first half year in 2015 does not affect the interest in EA topics significantly. The awareness index seems to be on a high level. This result is confirmed in Table 2.4 through the insignificant coefficient β_7. The devaluation of the RUB might have deflected interest to other areas. However, in Figures 2.10 a rising of the EA index is observable between August and December 2015, in particular in cluster 2 until cluster 4, which might be interpreted as a increase in EA topics before and during the COP. But after the conference, a decrease in the interest in EA topics is measurable. This result could be also driven by season, or rather temperature effects. However the deepest temperature point is reached in February. The coefficient for the COP, β_6 in Table 2.4 is positively significant in the most models, but if we include season dummies there, than the coefficient becomes insignificant, as m_4 shows.

If we cluster the regions differently, for example, regarding the height of their GRP per capita, as in Figure 2.12, then this results remains.

This result also holds, if we cluster the regions with respect to other variables: the share of employees in mining or manufacturing, or the emissions of greenhouse gases.

A further question refers to regional differences of awareness and their development over time. In this context, Figure 2.13 shows that the regions located in the eastern part of the Russian Federation tend to reveal a higher level of awareness in comparison to the regions located in Europe or in Central and Western Siberia; see also Table 4.7 in the Appendix for the corresponding regions within the geographical cluster. The results hardly change from Figure 2.10. The intensity seems to decrease from east to west. Already the estimated coefficient values in Table 2.4 confirm these results. This phenomenon is already mentioned in Lösch, Okhrin, and Wiesmeth (2017).

Figure 2.14 details the location of the clustered regions with the highest (grey) and lowest (red) values of EA. Lösch, Okhrin, and Wiesmeth (2017) provide more information on the "diffusion" of EA, which seems to spread from the eastern parts of the country to

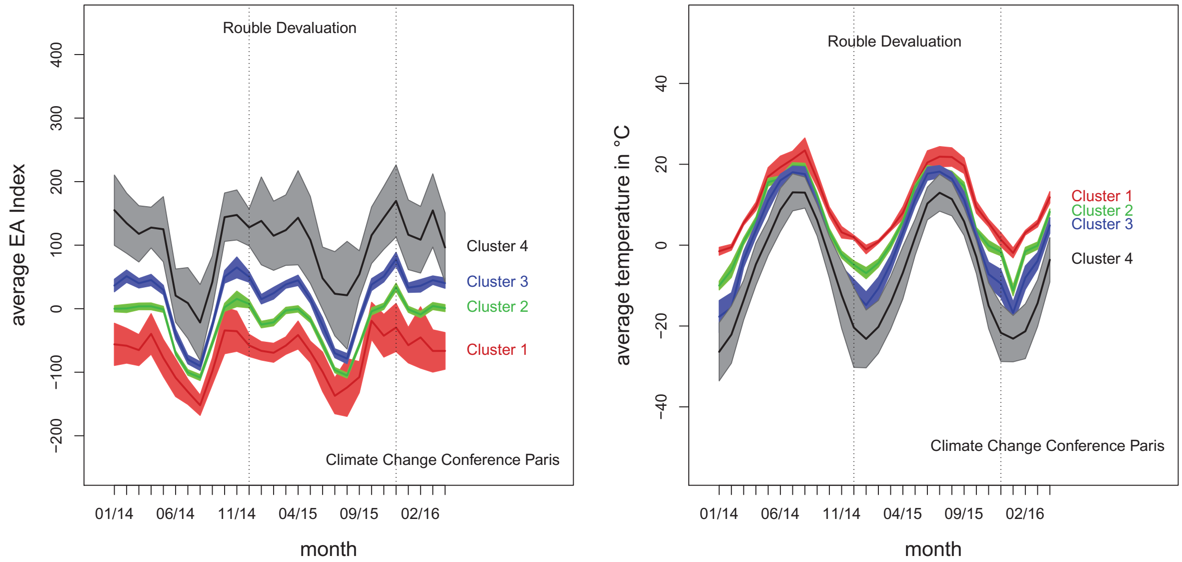

Figure 2.10: Estimated average awareness index for 28 month for the regions assembled in four clusters according to the EA index (left) and temperature (right). The affiliation of the regions to the cluster (left picture) can be found in Table 4.6 in the appendix. Model m_1 is used for estimating the EA index.

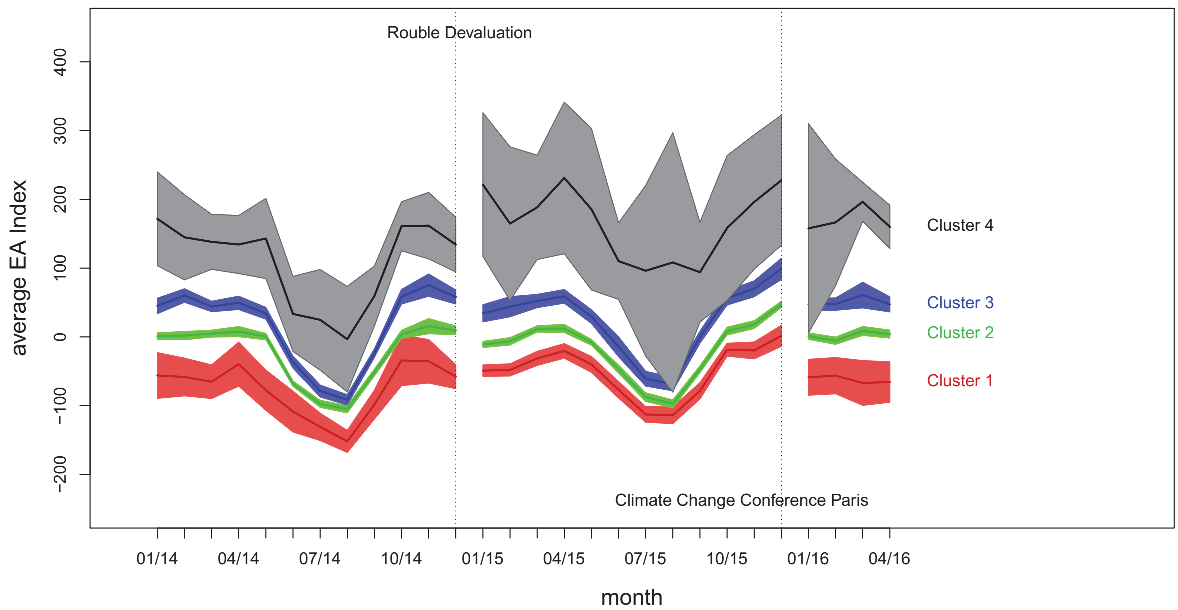

Figure 2.11: Estimated average awareness index for 28 month for the regions assembled in four clusters according to the temperature (right). Model m_1 is used.

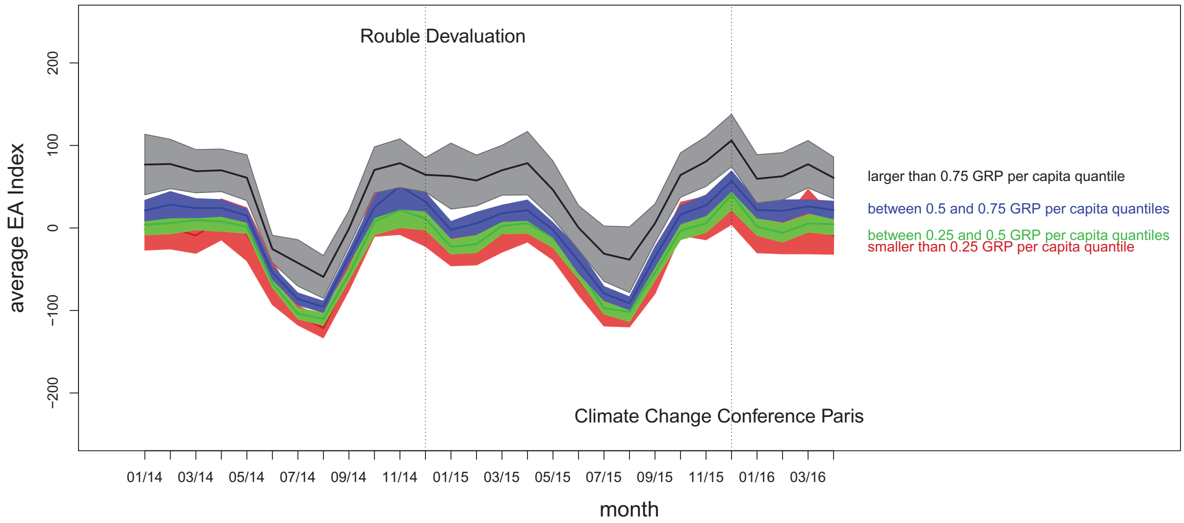

Figure 2.12: Estimated average awareness index for 28 month for the regions assembled in four clusters according to the size of the GRP per capita. Model m_1 is used.

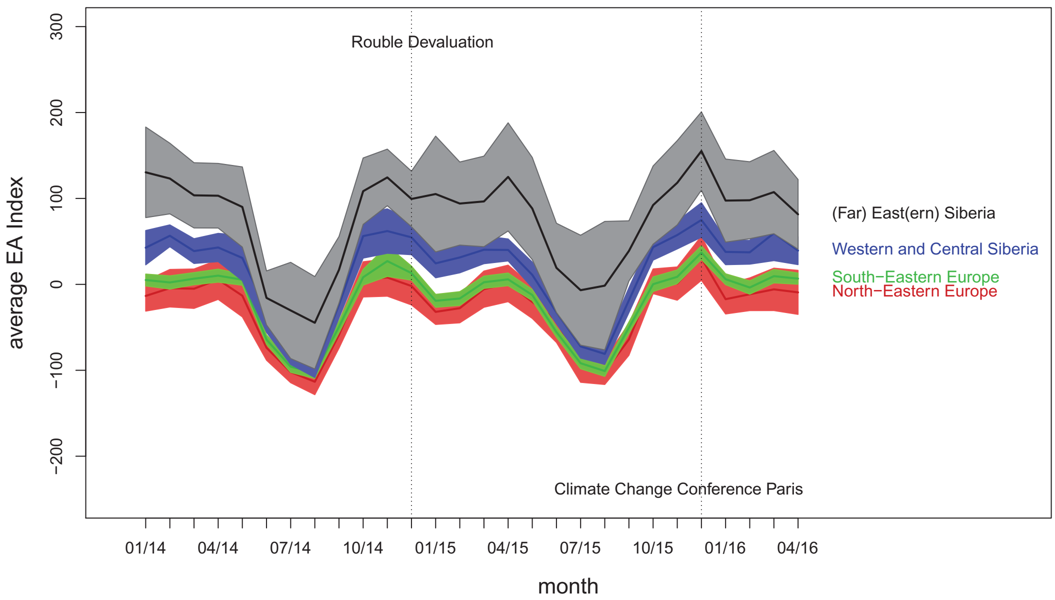

Figure 2.13: Estimated average awareness index for 28 month for the regions assembled in four clusters according to the geographical position. Model m_1 is used.

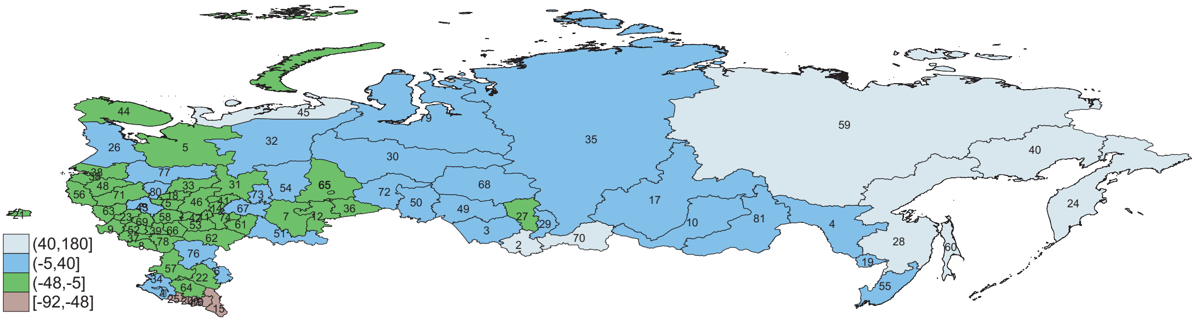

Figure 2.14: Map with associated average levels of awareness index for the regions in 2014 assembled in four clusters according to the level of awareness of EA. Model m_1 is used for estimating EA index per month. The average EA index for 2014 in each region is subsequent computed. The Ids corresponds to the regional keys in Table 4.5 in the Appendix.

the western parts.

The final Figure 2.15 in this context shows some box plots for the changes of the average levels of awareness in the regions clustered according to their awareness levels. The graphic depicts the changes in the levels of awareness from 2014 to 2015. Although the composition of the clusters changes a little bit, taking into account this effect would not alter the picture radically.

We thus observe a light upward development of average awareness levels in each cluster when moving from 2014 to 2015 levels. The variability within the clusters decreases.

2.7.3 Environmental Awareness in the Arctic Zone

The previous results might confirm the presumption, that the population of colder regions are especially interested in EA topics. For that reason the current section focuses on the EA of Arctic and Subarctic regions in the North of the Russian Federation. The well-being of the citizens in the Arctic regions depends on the environmental situation particularly. The impacts of the global climate change are already apparent. They reveal through the speed of melting permafrost areas as well as the changing picture of the forests, agriculture and so forth. The 81 Russian regions are grouped into Arctic, Subarctic and all the others (rest) regions. The Arctic zone comprises regions, which are, at least partially, north of the parallel of 66°north latitude. These are Arkhangelsk, Nenets AO, Yamalo-Nenets AO, Chukotka AO, Murmansk, Sakha, and Krasnoyarsk. Moreover, Khanty-Mansi AO, Karelia, Komi, Magadan, Kamachatka, and Leningrad with St. Petersburg, lying between the parallel of 66° and 60° north latitude, are here defined as Subarctic zones. As mentioned above, the Russian regions differ substantially in their characteristics. As before, MIMIC model 2.10 is estimated and the average EA indices per regions are

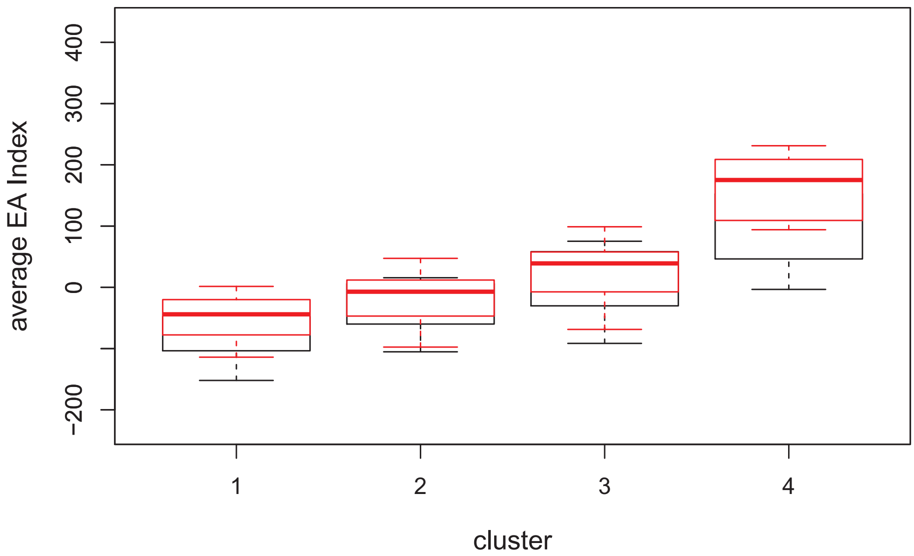

Figure 2.15: Box plot with associated levels of awareness index for the regions assembled in four clusters for 2014 (black) and 2015 (red) according to the level of EA. Model m_1 is used.

computed. In contrast to Lösch, Okhrin, and Wiesmeth (2018b), which used model m_3 from Table 2.4 estimating the awareness index, the better data-fitted model m_1 is here exploited. However, the results hardly change. The most of the Arctic or Subarctic regions have an above average GRP per capita and/or higher EA indices.

Figure 2.16 shows the development of awareness in the Arctic and Subarctic zones in comparison to that in the other regions. Interestingly, awareness tends to be higher in the Arctic regions in comparison to the rest of the country. Since there does not seem to be statistically significant differences between the defined Arctic and Subarctic zones, a common spatial group is generated. The average EA index time-series for this Arctic-Subarctic group can be found in Figure 2.17 including 95% CI. Since the joint variance shrinks with a rising number of observations, the CI of the Arctic-Subarctic group hardly touch the CI of all other regions. In particular, there are statistically significant differences in the winter terms observable. This result confirms the hypothesis above. Figure 2.18 details the regions with Arctic (blue) and Subarctic zones (orange) as well as all others (grey). Actually the region Krasnoyarsk extends from the north to the south of Siberia, which could strongly influence the variability of the average Arctic EA index. But if we exclude this region as an Arctic area, the results hardly change from Figure 2.16.

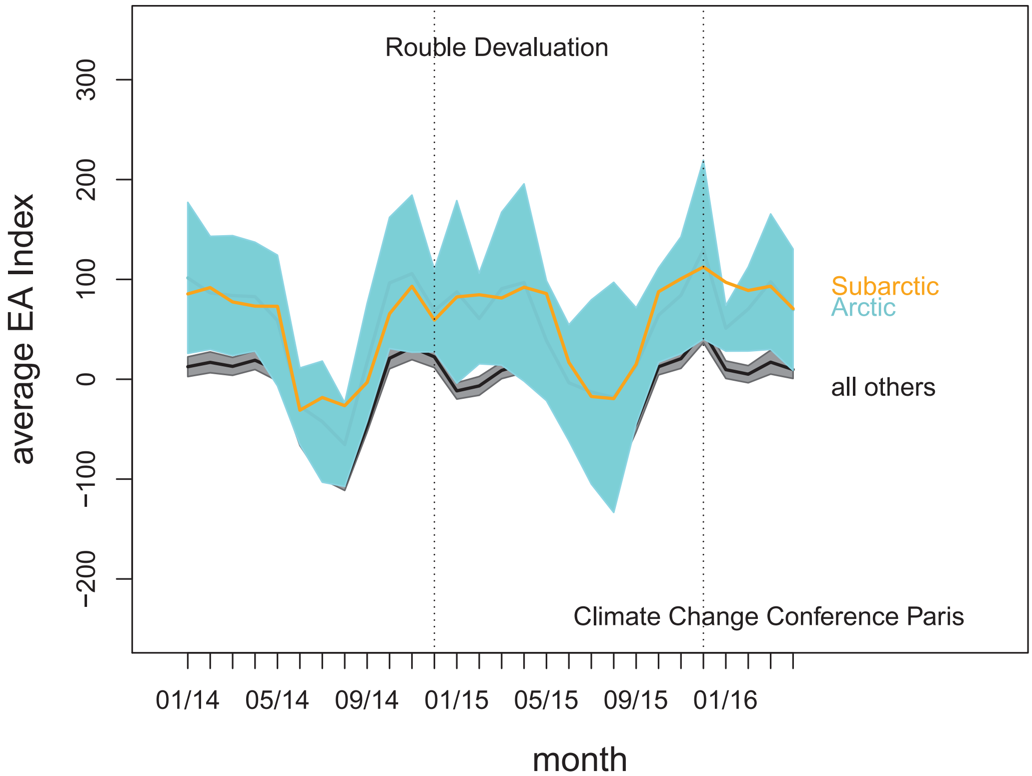

Figure 2.16: Estimated average EA index for 28 month for the Arctic (light blue), Subarctic (orange) and all other regions (black). Picture shows 95% CI of Arctic and all other regions. Model m_1 in Table 2.4 with temperature data and without seasonal dummies is used for estimating EA index.

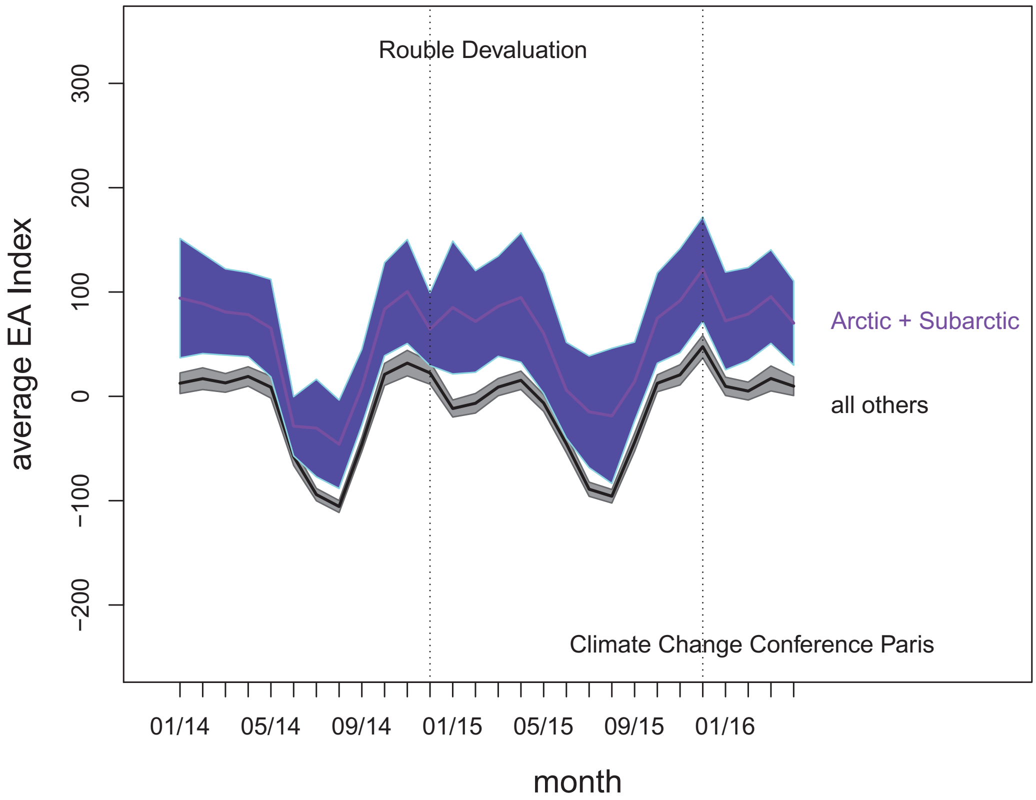

Figure 2.17: Estimated average EA index for 28 month for the joint Arctic-Subarctic Area (violet) and all other regions (black). Picture contains 95% CI of the two groups. Model m_1 in Table 2.4 with temperature data and without seasonal dummies is used for estimating EA index.

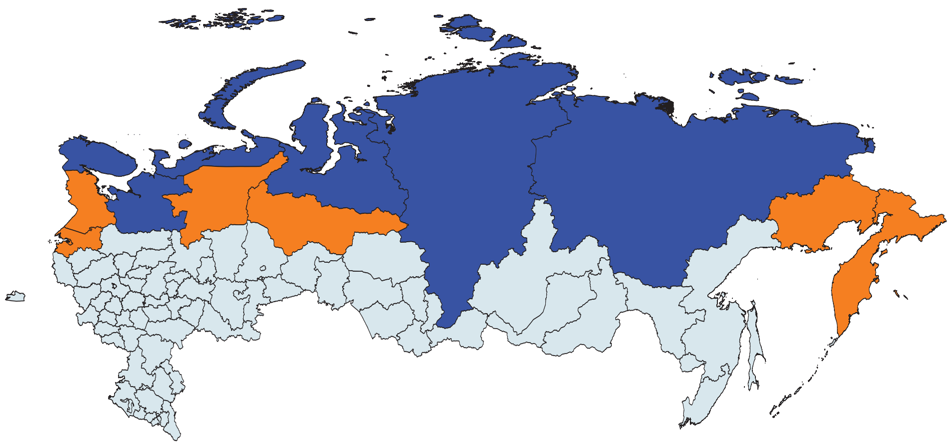

Figure 2.18: Map of Russia with defined Arctic (blue) and Subarctic (orange) regions without Chukotka Autonomous Obrug.

2.8 Concluding Remarks

To sum up, the current chapter reveals a less time-consuming and less-expensive method for estimating regional environmental awareness. After an introduction and literature overview about EA, the structure of the applied MIMIC model is described and discussed. The idea is to estimate the not-directly observable regional EA Index of 81 Russian regions through some endogenous indicators, which are here the number of Internet queries of certain environmental phrases from the search engine ©Yandex. These Internet queries are less subjective, and a general problem of EA surveys, the so-called "yes-bias" can be avoided Hiramatsu, Kurisu, and Hanaki (2015). In addition, the regional EA depends on some exogenous causes, such as GRP per capita, industry characteristics, age structure of the citizens, education level, but also the environmental attributes and regional climate condition, which are included into the model.

In Section 2.6, the results of the MIMIC model for the aggregated year data of the Russian regions are revealed and discussed. The regions are ranked concerning their EA index. In particular the relationship between regional EA and GRP per capita is pointed out, but also further correlations with regional characteristics are discussed.

In the following Section 2.7, the results of the extended MIMIC model, which considers seasonal variation, are reveals and mooted. Therefore the MIMIC model is estimated for each month through stacking the monthly data and capturing seasonal by using binary variables for quarter and years. In addition, because seasonal dummies do not allow to consider variation between the regions, temperature data are instead included into the model. This allows to taken regional climate circumstances and the involvement of the population in the climate change process into account.

As a result of both sections, a non-linear positive relationship between GRP per capita and regional EA can be found. In contrast, other cause variables, such as the share of employees in manufacturing sector, unemployment rate or greenhouse gas emissions hardly affect the consciousness of the population. The highest index levels can be measured in the eastern part of Russia, particular in Chukotka Autonomous Okrug, Oblast Magadan and Kamchatka. Large parts of these regions belong to the UNESCO natural heritage, which might be a reason for this result. Other high-awareness regions are the rich Nenets Autonomous Okrug and Republic Sakha, which live from oil-intensive industry sectors. The most of the regions with high awareness indices lie in the cold North of the country. The fear that the permafrost areas defrost will amplify climate change process might be explain this results, see Witman (2017). Moreover, the interest in EA topics decreases from the eastern to the western part of Russia. The smallest EA index values can be measured in the European part, especially in the Caucasus region, which belongs to the poorest as well as warmest region in the country.

Furthermore, there are seasonal variability in the temporal development of the EA index, which can be partly explained by the regional climate circumstances. The interest in environmental topics is lower in warmer month (summer) and warmer regions than in colder month (winter) and colder regions. There are peaks between October and December as well as in March and April of 2014 and 2015. The RUB devaluation at the end of 2014 seems not influence the regional EA. In contrast, a positive effect of the COP on the EA index is measurable. However, the analysis cannot answer conclusively the question, whether this positive effect is driven by the COP event or conditioned by the season. A longer observation period may here lead to clarification.

Notwithstanding, the approach, which is here applied, is only limited by the data and can easily be used for further (environmental or economic) issues and/or for other regions. It would be interesting to extend the analysis by using data from further countries and/or further search engines. Furthermore, it could be possible to relate the regional interest in environmental topics to consumption of electrical energy from renewable sources, instead of using Internet queries. In addition, the assumption concerning the random variables, which were here made, might not hold. Future work could deal with these issues.

Chapter 3

Labour Market Tightness and Individual Wage Growth

This section arises from a collaboration with the senior labour market researcher Stephan Brunow, formerly research associate at the Institute of Employment Research (IAB) in Nürnberg and currently Professor at University of Applied Labour Studies in Schwerin.

3.1 Introduction

Globalisation and technological progress have challenged many developed countries for a long time, especially in economic and social terms. Technological progress has shaped the labour market pattern. Recently, Acemoglu and Restrepo (2018) show that inequality is increased by faster automation and the establishment of new tasks. At the same time, robots and machines can potentially replace simpler routine tasks performed by low-skilled workers, leading to an oversupply of low-skilled labour in the non-service sector, which potentially pushes down their already rather low wages. Digitalisation has brought many new and complex tasks, which requires the control, maintenance and the extension of machines. Although the workers who undertake these jobs might profit from digitalisation, they require specialised knowledge in advanced science-related subjects. Moreover, the productivity growth in these technologically complex production processes can potentially accelerate, which increases the already rather high wages of these skilled-workers. Therefore, those industries that mainly employ workers who perform routine task might experience enormous employment restructuring by adaption of computer systems and machines. This effect accelerates wage inequality, which is also so called Skilled-Bias Technological Change (Goldin and Katz 2009, Acemoglu and Autor 2011). Acemoglu and Restrepo (2018) present mechanisms under which human work is substituted by, or is complementary to capital intensive production. This depends on production costs, and whether it leads to higher unemployment and stagnant wages for specific groups. According to Autor, Levy, and Murnane (2003), it might not (yet) be possible to replace workers by machines in non-routine tasks, such as housekeeping, hotel and personal care, which are mostly low-paid jobs. These effects could be a reason for the job polarisation at the margins of the wage distribution: low-paid non-routine tasks are still in demand and high-paid technologically demanding jobs have become increasingly important, but middle wage jobs are increasingly substituted by machines (for the United States: Autor and Dorn 2009; Autor, Katz, and Kearney 2006 and Autor, Katz, and Kearney 2008; the UK: Goos, Manning, and Salomons 2014 and West-Germany: Spitz-Oener 2006 and Dustmann, Ludsteck, and Schönberg 2009).

In Germany, specialists in medical subjects and food services are urgently needed. Specialists in nature science and technology (here called STEM for Science, Technology, Engineering and Mathematics) are also in demand because of technological change (Federal Employment Agency Germany 2017). Beside technological change, the so-called demographic change is expected to become more important over the coming decades. The retirement of the 'baby boomer' generation and population decline leads to another expected reorganisation at the labour market. The first signs of this change can already be seen: the (relative) numbers of young people leaving school is growing smaller and almost half of them decide for further training in universities. Therefore, companies have started to compete for these youngsters to fill their apprenticeship workplaces, see Statistisches Bundesamt Deutschland (2017). Consequently, the apprenticeship market becomes tighter and wage increases are expected. Technological change place higher labour demand on specific occupations (e.g. for STEM workers), whereas other occupations are crowded and do not enjoy such a wage growth. Therefore, it has to be asked if there is a shortage of young labour in occupations? If so, then it should be associated with relatively higher wage growth for a given or even growing stock of labour and a reduction in unemployment. This is the departure of this research: *when the labour market becomes tighter, we expect higher wages.* We raise two questions: first, can we identify higher individual wage growth for apprenticeship leavers depending on the relative shortage in separate occupations? And second, is this effect more pronounced for specific occupations (e.g. STEM)? For this purpose, we construct a labour market tightness measure following the Wage-Curve-literature (e.g. Blanchflower and Oswald 2008) and we estimate an individual 10 year wage growth equation in the setting of Mincer (1974). We make use of an extensive German administrative database of young individuals for the time period 1995–2014.

Believing that the effect of occupational labour market tightness is driven by techno-

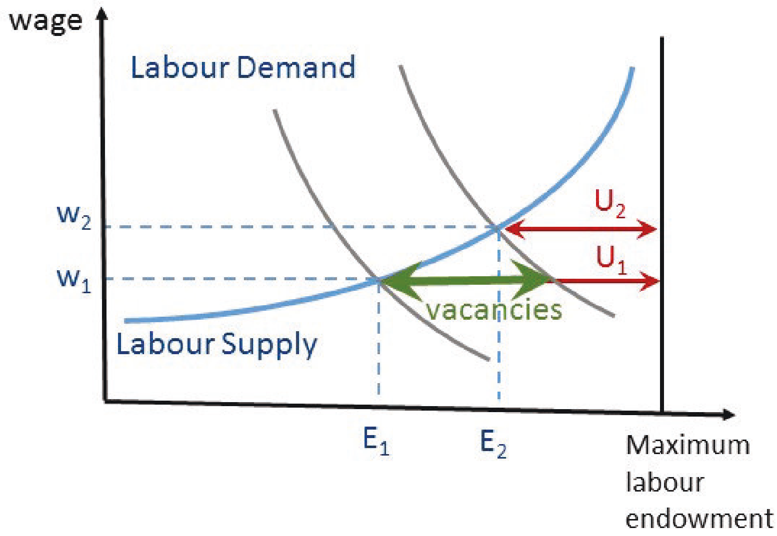

Figure 3.1: Wages in a labour demand and supply equilibrium.

logical change or by demographic change, we set up arguments of the wage curve literature in a basic macroeconomic setting, see Blanchflower and Oswald (2008). Consider labour demand D and labour supply S curves of a specific occupation as outlined in Figure 3.1. There is a maximum endowment of labour in a specific occupation and region, either employed E or unemployed U. In equilibrium, E_1 people are employed and U_1 people are unemployed, leading to wage w_1. If labour demand in an occupation increases, then the demand curve shifts outwards. Wages are expected to rise from w_1 to w_2. However, in the short run, employers usually pay w_1 and, therefore, an excess demand results. We interpret the excess demand as open vacancies. At the same time, U_1 unemployed workers are available but are unwilling to take a job for wage w_1. Within a labour market matching framework the number of open vacancies and the number of unemployed people are usually employed to construct a labour market tightness measure. However, as discussed later on, data on open vacancies is only very rarely available at the regional-occupational level and, therefore, we use the employment level E_1 instead, similar to Moscarini and Postel-Vinay (2016). Consequently, our tightness measure relates to the unemployed-to-employed in a region and occupation. In the medium and long-run, wages will adjust and the $E_2 - w_2 - U_2$ combination will describe the new equilibrium. As can be seen, when the labour market becomes tighter, this implies higher employment and lower unemployment levels. If the unemployment level decreases, then wages are expected to rise. Therefore, our measure of regional, occupation-specific labour market tightness is the ratio of unemployed to employed (ue), which has already been employed by Moscarini and Postel-Vinay (2016). If this ratio decreases, then wages are expected to rise. Moreover the question arises of whether or not there are different effects on the wage growth between certain income groups due to labour market tightness within an occupational field (for example IT, salesman, social occupation). We assume that low-paid employees within a certain occupation, who are hypothetically low-skilled, can be more easily replaced by capital (machines) than well-paid (high-skilled) employees in the same professional field. Low-skilled people might be aware of their weak bargaining power and they are much more ready to work for lower wages than high-skilled people if the unemployment rate rises. Meanwhile, the stronger bargaining power of the well-paid high-skilled employees leads to an increase of their wages when the labour supply is low. Both of these situations lead to a rising wage inequality between these both income groups. Furthermore this wage dispersion seems to be task driven and, thus, expedited by technological progress; see Kim and Sakamoto (2008), Fortin and Lemieux (2016), Firpo, Fortin, and Lemieux (2011).

Inspired by Mincer (1974) wage equation, we estimate monthly log-wages 10 years after the apprenticeship training of employees through initial characteristics at the first job and the changes of these attributes. Having this in mind, we attempt to answering following questions:

Q1: Does labour market tightness affects individual wages?

Q2: If Q1 holds, then does this effect influence all occupational/gender groups in the same fashion?

Q3: If Q1 holds, then does labour market tightness differently affect individuals from different income groups?

This chapter is structured as follows. Our data and variable constructions are introduced in Section 3.2. The econometric methods, which are used in this chapter, are presented in Section 3.3. Afterwards, the results are shown and discussed. Section 3.7 summarises and concludes the chapter.

3.2 Data and Sample Selection

We use the individual data of the Integrated Employment Biographies (IEB) provided by the Institute for Employment Research (IAB) for the years 1995–2014. This is administrative data and it includes individuals working who are subject to social security contributions and individuals who are unemployed. This allows us to construct the entire (un-)employment history of about 90% of the entire German labour market. From the whole universe of all of the employment entries of all individuals (so-called spells) contained in the IEB, we aggregate data at the level of the firm and all higher levels of hierarchy, such as the industry, occupation and region, and we derive a linked-employer-employee dataset. A summary of the constructed variables is shown in Table 4.11 in the Appendix.

Card and Lemieux (2001) argue that different cohorts are imperfect substitutes because the older ones have already attained experience and, thus, are potentially more productive. In contrast, the newest cohorts hold the newest knowledge but have almost no experience. They start with similar conditions at the labour market and have a more or less similar productivity level. Furthermore, they might profit from the latest technological changes. Thus, they are easier to compare and do not have a long and very complex employment biography, which can affect their individual wages. Consequently, we are interested in the newest cohort and we aim to get rid of this productivity gap. We consider a 10% sample of employees under 30 years of age when they just completed vocational training leading to a homogeneous sample of 350000 individuals.

We consider their wage growth within the next 10 years. Although the data is highly reliable, we drop few outlier cases with exceptional low or high wages. Occupations are grouped into 54 distinct occupational fields on the basis of the KldB88 at the three digit level. This classification takes within-occupation mobility into account, such that low mobility rates between occupations occur. Because we consider wage growth of the individuals between the first job and 10 years after, we only consider individuals that are employed at both moments in time. We are aware of the potential selection bias, especially for not controlling of being unemployed or self-employed 10 years after. Our final sample contains 316711 individuals. Because we only focus on young individuals, their wages are below a specific social security contributions threshold and we do not have to impute truncated wages. Our focus variables describe the economic conditions of the occupation-specific regional labour market. We suppose stronger individual wage growth in occupations and regions, where relatively few potential workers are available for starting a new job relative to the number of employees. Therefore, the focus variable, ue, is computed as the number of unemployed people divided by the number of employees within an occupational field and labour market region, where an observed individual started to work after finishing vocational training. In addition, we also compute the change, Δue, between the first job and 10 years after. To account for alternative explanations, we control for the occupation-specific regional labour market size to account for Marshallian externalities, Marshall (1920). We also consider the direct competition through the proportion of academics employees (holding an university degree) within an occupation and region to account for potential substitutional or complementary relationship between apprenticeship and university-degree holders. If there are, for example, many employed university degree holders and only a few workers who have passed vocational training, then the labour market is relatively tight from a vocational training perspective because this qualification level becomes scarcer. Furthermore, we include firm specific characteristics at the entry time point and its change, such as firm size, firm age, share of human capital, proportion of women and foreigner within firms, while taking different productivity levels into account.

3.3 Methodological Approach

Let us define and formalise the setup with the model. Having an initial log wage $w_{i,t}$ at time $t \in [1995, 2004]$ of an individual $i \in 1, \ldots, N$ with $N = 316711$ depending on occupational o, firm f, regional r and also industrial s environment, we intend to model its change ten years after entering the labour market $w_{i,t+10}$. Computing a wage growth measure as the dependent variable requests for controlling not just initial conditions (as e.g. in the specification of the growth model from Solow (1956) but also for the situation in the future). We estimate

$$\begin{aligned} w_{i,t+10} = \alpha_0 + w_{i,t}\alpha_1 + x_{i,t}^T\beta_1 + z_{r_i,o_i,t}^T\beta_2 + ue_{r_i,o_i,t}^T\beta_3 \\ + v_{f_i,t}^T\beta_4 + \Delta x_{i,t}^T\gamma_1 + \Delta z_{r_i,o_i,t}^T\gamma_2 + \Delta ue_{r_i,o_i,t}^T\gamma_3 \\ + \Delta v_{f_i,t}^T\gamma_4 + \mu_{o_i} + \mu_{r_i} + \mu_{s_i} + \mu_{u_i} + \varepsilon_{i,t+10}. \end{aligned} \tag{3.1}$$

According to the wage equation from Mincer (1974), the individual log wage is explained by individual characteristics $x_{i,t}$ at the entry time point, as well as their change $\Delta x_{i,t}$. Furthermore, the labour market situation (labour market size; number of other employees in the same occupation field within the same labour market region) $z_{r_i,o_i,t}$ at t and its change $\Delta z_{r_i,o_i,t}$ depending on the occupation(s) o of i and labour market region(s) r of i are considered. We also include the firm's characteristics $v_{f_i,t}$ and its change $\Delta v_{f_i,t}$ depend on the firm(s) where i works. The vectors $ue_{r_i,o_i,t}$ and $\Delta ue_{r_i,o_i,t}$ contain the labour market tightness variables, which include the number of unemployed related to the employed individuals with the same profession within a specific labour regions and the interaction of these variables with occupational dummies. In addition, we control for occupational μ_{o_i}, regional μ_{r_i}, industrial μ_{s_i} and time μ_{u_i} fixed effects. For simplification purposes, (3.1) is aggregated to

$$w_{t+10} = z_t^T\eta + \varepsilon_{t+10}, \tag{3.2}$$

with $w_{t+10} = w_{i,t+10}$. Vector $z_t = (1, w_{i,t}, x_{i,1,t}, \ldots x_{i,J,t}, \Delta x_{i,1,t}, \ldots \Delta x_{i,K,t}, z_{r_i,1_i,t}, \ldots z_{r_i,O_i,t}, \Delta z_{r_i,1_i,t}, \ldots \Delta z_{r_i,O_i,t}, ue_{r_i,1_i,t}, \ldots ue_{r_i,O_i,t}, \Delta ue_{r_i,o_i,t}, \ldots \Delta ue_{r_i,o_i,t}, v_{1_i,t}, \Delta v_{F_i,t}, d_{1_i} \ldots d_{O_i}, g_{1_i} \ldots g_{R_i}, h_{1_i}, \ldots h_{S_i}, l_{1_i}, \ldots l_{U_i})^T$ includes the intercept and the log wage $w_{i,t}$ at the begin of the first job; $x_{i,j,t}$ denotes a set of $j = 1, \ldots J$ individual characteristics at the first job to time t, $\Delta x_{i,k,t}$ is their change from t to $t + 10$, with $k = 1, \ldots, K$ and $K \subseteq J$. Capturing possible unobservable heterogeneity in the occupational $o = 1, \ldots, 53$, regional $r = 1, \ldots, 140$, industry $s = 1, \ldots, 58$, and year $u = 1, \ldots, 9$ information corresponding binary variables d_{o_i}, g_{r_i}, h_{s_i} and l_{u_i} are included in the model. The associated weights are aggregated in the coefficient vector
$\eta = (\alpha_0, \alpha_1, \beta_{11}, \ldots, \beta_{1J}, \beta_{21}, \ldots, \beta_{2R}, \beta_{31}, \ldots, \beta_{3R}, \beta_{41}, \ldots, \beta_{4F}, \gamma_{11}, \ldots, \gamma_{1K}, \gamma_{21}, \ldots, \gamma_{2R}, \gamma_{31}, \ldots, \gamma_{3R}, \gamma_{41}, \ldots, \gamma_{4F}, \mu_{1_i}, \ldots, \mu_{O_i}, \mu_{1_i}, \ldots, \mu_{R_i}, \mu_{1_i}, \ldots, \mu_{S_i}, \mu_{1_i}, \ldots, \mu_{U_i})^T$.
The error term vector is $\varepsilon_{t+10} = \varepsilon_{i,t+10}$. All of the right hand side variables (RHS) that might affect individual log wages in general are considered at the initial time point t and their changes within t and $t + 10$. Furthermore, we assume there is no unobservable correlation between ε_{t+10} and w_{t+10}. If this assumption holds, then we can consistently estimate the parameter vector η by minimising the sum of error squares, which leads to $\hat{\eta} = (z_t^T z_t)^{-1} z_t w_{t+10}$. To overcome difficulties with the different levels (f.e. occupational, regional, industrial specific, etc.) in the data, the variance of η is estimated by using a clustering method $\widehat{\text{Var}}(\eta) = \left(z_t^T z_t\right)^{-1} \left(\sum_{p=1}^{P} z_{r,t}^T \hat{\varepsilon}_{r,t} \hat{\varepsilon}_{r,t}^T z_{r,t}\right) \left(z_t^T z_t\right)^{-1}$ after the OLS estimation, see Cameron and Miller (2015). Therefore the RHS variables $z_{r,t}$ and estimated residuals, $\hat{\varepsilon}_{r,t} = w_{r,t+10} - z_{r,t}\hat{\eta}$, are clustered at the regional level. However, the variable set is quite large containing 628 variables with 291 binary variables, which leads to typical collinearity problems during the OLS estimation of the coefficients. Overfitting of the model is probable. Some of the estimated coefficients become statistically significant by coincidence, but they perform rather poorly, see Ahrens, Schaffer, and Hansen (2018). Consequently, we use Least Absolute Shrinkage and Selection Operator (LASSO) algorithms to shrink the dimension and avoid collinearity in the structural parameter. The idea is to penalise the parameters of RHS-variables η that hardly influence the LHS-variable w_{t+10} and push them down to zero. The aim is to obtain a simpler model, which does not include rather irrelevant variables. It should help to improve the forecast accuracy and the interpretability of the models. Here, the default LASSO approach by Tibshirani (1996) with the resulted LASSO coefficient vector

$$\hat{\eta}^* = \text{argmin}_{\hat{\eta}^*} \, ||w_{t+10} - \sum_{p=1}^{P} z_p \eta_p||^2 + \lambda \sum_{p=1}^{P} |\eta_p|, \tag{3.3}$$

is used. Thereby, $\lambda \geq 0$ is the penalty coefficient, which is found by cross validation, and

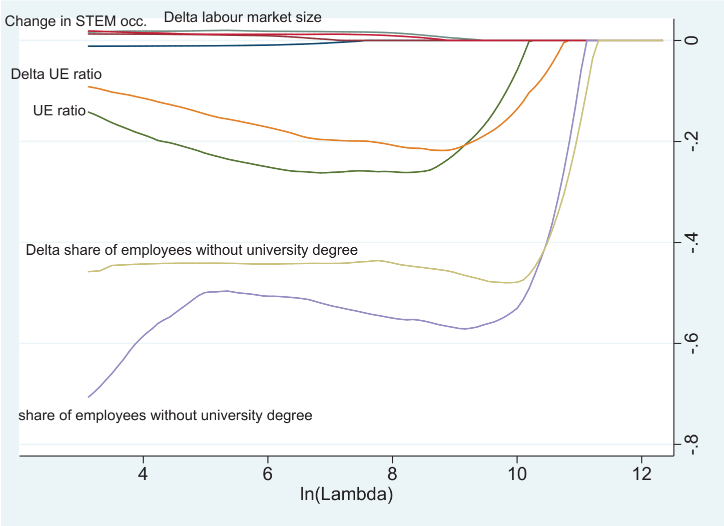

Figure 3.2: An extract of the LASSO solution path. Variables are being removed from the model as λ increases.

p denotes a specific RHS variable in the model. However, the LASSO variables selection can be inconsistent under certain conditions, see Meinshausen and Bühlmann (2006), Fan and Li (2001). Consequently, Zou (2006) augmented the penalty term through different weights for different coefficients. This yields to

$$\hat{\eta}^{*\mathrm{AD}} = \mathrm{argmin}_{\hat{\eta}^{*\mathrm{AD}}} ||w_{t+10} - \sum_{p=1}^{P} z_p \eta_p||^2 + \lambda \sum_{p=1}^{P} \frac{1}{|\eta_p|^{\gamma^{*\mathrm{AD}}}} |\eta_p|, \tag{3.4}$$

with $\gamma^{*\mathrm{AD}} > 0$ and $\lambda \geq 0$. The data-driven weights should ensure the so-called oracle-properties, see Zou (2006). The LASSO coefficients path in Figure 3.2 illustrates the variable selection process, which corresponds to a piecewise linear and continuous function of λ. Variables enter or leave the active set because of changes in the slope of the function. If $\lambda = 0$, then the full set (here called *Full Model*) is achieved.

The optimal λ, λ_{opt}, that minimises the mean-square prediction error is chosen by *k-fold cross-validation algorithm* see Ahrens, Schaffer, and Hansen (2018):

STEP 1: The dataset is divided into (approximately) five equal folds.

STEP 2: Fold 1 is treat as the validation dataset. Folds 2–5 are the training data.

STEP 3: The model is estimated by using the training data. The predictive performance of a range of λ is determined by using the validation data.

STEP 4: STEPS 2 and 3 are repeated using folds $2, \ldots 5$ as validation dataset.

STEP 5: Find the λ that reveals the best out-of-sample predictive performance.

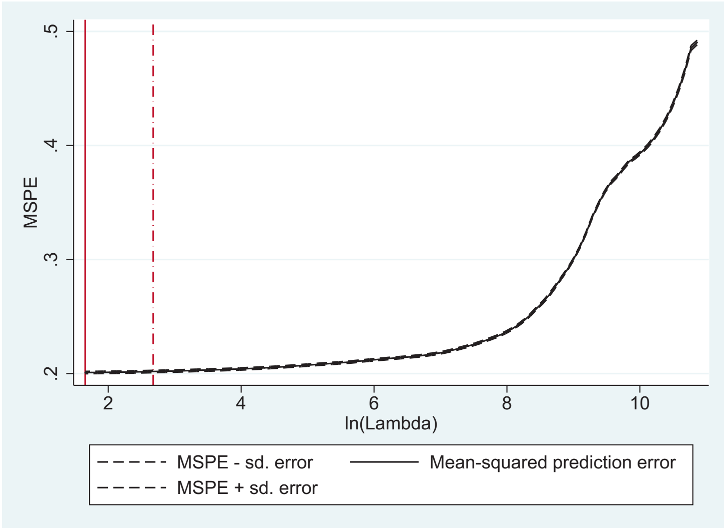

Figure 3.3: Choose λ that leads to the smallest mean-square prediction error (MSPE). The red-solid vertical line corresponds to λ_{opt} that minimises the MSPE. The red-vertical dashed line shows the largest λ_{LSE} at which the MSPE lies within one standard error of the minimal MSPE.

Figure 3.3 illustrates the λ selection process by using the observed data. Because the resulting LASSO coefficient vectors $\hat{\eta}^*, \hat{\eta}^{*\text{AD}}$ are biased, we estimate the reduced models after the variable selection by OLS and compare the estimated coefficients of the main variables, this is in line with Chernozhukov, Hansen, and Spindler (2015).

Furthermore, we expect that the levels of the individual, occupational, firm, regional, industry characteristics affect the log wage distribution $F_{w_{t+10}}$ differently. For example, a changing labour market situation might influence the demand of low-paid jobs in other ways than better-paid-jobs and thus their wage growth. Consequently, we also estimate the distribution $F_{w_{t+10}}$ of w_{t+10} through the unconditional quantile regression method by Firpo, Fortin, and Lemieux (2009). This approach is based on the concept of sample quantile q_τ added to the influence function IF $\left(w_{t+10}; q_\tau, F_{w_{t+10}}\right)$ of this quantile of the wage growth for a specific level $\tau = \{0.25, 0.50, 0.75\}$:

$$\tilde{w}_{\tau,t+10} := q_\tau + \frac{(\tau - \mathrm{I}\, w_{t+10} \leq q_\tau)}{f_{w_{t+10}}(q_\tau)} = z_t^T \eta(\tau) + \tilde{\varepsilon}_{t+10}, \tag{3.5}$$

where $\tilde{w}_{\tau,t+10}$ is the unconditional τ quantile, which is called the recentered influence function and $\mathrm{I}\left\{w_{t+10} \leq q_\tau\right\}$ is the indicator function. The density $f_{w_{t+10}}(q_\tau)$ depends on q_τ and is estimated in our study using KDE with Gaussian kernel. We regress a set of variables z_t against $\tilde{w}_{\tau,t+10}$. The coefficient vector η depending on τ can be simply estimated via OLS. In comparison with the more traditional conditional quantile regression method by Koenker (2005), the estimation takes place in one step. Given that the model is estimated

	Full Model (1)	LASSO (2)	Ada LASSO (3)
ue effect without interaction	-0.078	-0.100$^{\#}$ (0.030)	-0.195$^{\#}$ (0.019)
Δue effect without interaction	-0.059	-0.067$^{\#}$ (0.025)	-0.151$^{\#}$ (0.015)
N	316711	316711	316711
adj. R^2	0.593	0.591	0.592
AIC	3.90E+05	3.90E+05	3.90E+05
BIC	3.90E+05	4.00E+05	3.90E+05

Table 3.1: (Post-) OLS estimation of the ue coefficients: by using the full model with 628 RHS variables (2nd column), LASSO (3rd column) and adaptive LASSO with 224 RHS variables (4th column). Heteroscedasticity robust clustered standard errors of the coefficients are in parentheses; significance level $^{\#}p < 0.01$, $^{+}p < 0.05$, $*p < 0.1$.

via OLS, no particular assumptions for the error term vector $\tilde{\varepsilon}_{t+10}$ are needed. There are no convergence problems. In addition, the marginal effect of a change in the distribution of z_t on the marginal quantiles of w_{t+10} is directly readable through the coefficients η. For further information, see Firpo, Fortin, and Lemieux (2009) and Borah and Basu (2013).

3.4 Does Labour Market Tightness Affect Individual Wages?

Let us start with the relationship between log wages and unemployment variable ue. In the case of a labour force shortage rising log wages are expected. A shortage occurs if the unemployment rate shrinks but the labour demand rises. The relative number of potential freely available staff decrease, which should rise the wage power of the employees.

Table 3.1 shows the estimated coefficients for ue and Δue, by using the full model with 628 RHS variables (column 1), the selected model by using LASSO from Tibshirani (1996) (column 2), as well as the selected model by using adaptive LASSO from Zou (2006) (column 3). The latter model contains only 224 RHS variables. This corresponds to a two-third dimension reduction without a loss in the model fit. The reduced model has less collinearity problems, in particular by considering interaction variables. The estimated coefficients are more precise. In addition, the dimension reduction serves as a kind of robustness check for the coefficient's significance. All of the models can almost explain approximately 60% of the variance of the 10-years log wages.

The coefficients of the unemployment variables in all three model variations are negative and significant as expected. The elasticity of $ue = -0.195$ (column 3) can be interpreted as follow: 1% decrease in unemployment at the first job would lead to an increase in wages of about 0.195 % 10 years after. A 1% change of this ue - ratio at the first 10 years after entering the labour market would to a further wage growth of approximately 0.151%. Nijkamp and Poot (2005) find a mean wage curve elasticity of -0.07 in their meta-analysis, which is weaker than our findings and close to our result in the full model. This could be related to the very young samples, which we observe. Young people are more likely to be employed within one year unemployment than older unemployed people; see Axelrad, Malul, and Luski (2018). Firms rather decide upon younger people, thus they have a greater bargaining power, especially if there is a shortage in the labour force. This result is quite robust through all our specifications and answers Q1: Labour market tightness positively affects the individual's wages.

3.5 Differences Between Occupations And Genders

To answer the second question Q2, we interact ue and Δue with occupation dummies. The results can be found in Table 3.2. Thereby, columns 2–4 reveal the coefficients of the estimated ue variables interoperated with the occupations. Columns 5–7 correspond to the estimated Δue interaction variables. Some ue - occupation interaction variables, which are selected by adaptive LASSO, were not inevitably chosen as a Δue interaction variable with the same occupation, and vice versa. For example, the interaction variable ue for metal and plant engineers is selected by adaptive LASSO but not the corresponding Δue for the same occupation, which measures its change within 10 years. We suppose that there is no (robust) significant difference of the Δue interaction variable to the average Δue effect for this occupation. For example, the coefficient of Δue for metal and plant engineering becomes weaker from the largest to the sparingly model as Table 3.2, columns 5–7 show. The cells for the coefficients of the not-selected variables are empty. The estimated coefficients hardly change between the large model (columns 2 and 5) and the smaller models, but the respective standard errors in this cases are significantly reduced. Most of them are negative significant, which means that the log wages of these occupations were especially negatively affected by a changing in ue. This means that the log wages of the individuals who were working in these occupations react particularly strong positively through labour market tightness. The more relaxed the labour market is, the more the log wages decrease. Meanwhile, the expected log wages are higher when there is a smaller number of potential employees in an occupation.

	Interaction between ue and Occupation			Interaction between Δue and Occupation		
	Full model	LASSO	Ada LASSO	Full model	LASSO	Ada LASSO
mineworkers	-0.135 (0.156)			-0.134* (0.081)	-0.092 (0.070)	
stone workers, ceramics, glass	-0.284 (0.191)	-0.170 (0.171)		-0.321+ (0.143)	-0.306+ (0.145)	
chemical & plastics work.	-0.421# (0.160)	-0.400+ (0.168)		-0.185+ (0.080)	-0.167+ (0.081)	
paper processing	-0.353+ (0.164)	-0.453# (0.155)		-0.318# (0.101)	-0.300# (0.099)	
metalworking	-0.263# (0.091)	-0.263# (0.089)		-0.002 (0.060)	0.004 (0.060)	
metal and plant engin.	-0.378# (0.064)	-0.339# (0.063)	-0.270# (0.056)	-0.104# (0.031)	-0.097# (0.031)	
Industrial, tool mech.	-0.400# (0.068)	-0.374# (0.064)	-0.342# (0.054)	0.006 (0.037)	0.010 (0.037)	
car, aircraft construction, maint.	-0.682# (0.103)	-0.678# (0.101)	-0.620# (0.101)	-0.045 (0.042)	-0.037 (0.043)	
precision mechanic	0.029 (0.156)	0.033 (0.158)		-0.384# (0.109)	-0.381# (0.109)	-0.295# (0.099)
electrical & electronic engin.	-0.612# (0.084)	-0.690# (0.078)	-0.669# (0.074)	-0.184# (0.048)	-0.197# (0.048)	
silkworm moth, textile finish.	0.106 (0.195)	0.109 (0.197)		-0.063 (0.078)	-0.076 (0.078)	
textile & leather process.	-0.171 (0.211)	-0.130 (0.208)		-0.152 (0.209)	-0.126 (0.208)	
bakers & confectioners	-0.071 (0.159)	-0.061 (0.130)		-0.240# (0.071)	-0.233# (0.071)	-0.149+ (0.066)
butcher	-0.175 (0.149)	-0.086 (0.080)		-0.149+ (0.064)	-0.136+ (0.062)	
cook	-0.130* (0.073)	-0.079 (0.068)		-0.164# (0.048)	-0.150# (0.048)	
beverages, drink & tobacco proc.	-0.307* (0.183)	-0.285 (0.186)		-0.110 (0.228)	-0.123 (0.233)	
building professions	-0.106# (0.040)	-0.107# (0.034)	-0.051+ (0.025)	-0.025 (0.029)	-0.019 (0.030)	0.055+ (0.022)
product tester	-0.085 (0.095)	-0.030 (0.084)		-0.145+ (0.066)	-0.153+ (0.065)	
labourers	-0.135# (0.042)	-0.091+ (0.039)	-0.024 (0.034)	-0.136# (0.032)	-0.124# (0.032)	-0.044* (0.025)
engineers	-0.808+ (0.329)	-0.842# (0.315)		-0.745# (0.146)	-0.810# (0.155)	
chemists	-0.761 (1.111)	-0.603 (1.148)		-0.348 (0.725)	-0.330 (0.685)	
technicians	-0.903# (0.216)	-1.047# (0.194)	-0.777# (0.136)	-0.540+ (0.209)	-0.539# (0.207)	
technical drawer	-0.504# (0.194)	-0.434+ (0.196)		-0.565# (0.174)	-0.557# (0.172)	-0.322+ (0.140)
surveyor	-0.367 (0.268)	-0.343 (0.255)		-0.216 (0.195)	-0.159 (0.189)	
technical special forces	0.001 (0.301)			-0.462* (0.258)	-0.456+ (0.231)	
salesman	0.503# (0.066)	0.647# (0.057)	0.598# (0.049)	-0.205# (0.044)	-0.207# (0.046)	-0.144# (0.044)
retailers and distributors	0.081 (0.113)	0.134 (0.108)		-0.337# (0.077)	-0.318# (0.075)	-0.204# (0.064)
bank empl. & insurance	0.124 (0.299)	0.296 (0.301)		-0.734+ (0.364)	-0.712+ (0.362)	-0.670* (0.365)
other salesmen	0.111 (0.161)	0.132 (0.159)		-0.448# (0.139)	-0.461# (0.136)	-0.424# (0.132)
advertising experts	-0.252 (0.264)	-0.236 (0.236)		-0.017 (0.134)		
transport-related voc.	-0.426# (0.150)	-0.388+ (0.160)		-0.075 (0.069)	-0.062 (0.069)	
aviation & shipping	-0.654* (0.394)	-0.676+ (0.320)		-0.800# (0.217)	-0.843# (0.223)	
packer & warehouse	-0.196# (0.073)	-0.138+ (0.068)		-0.172# (0.042)	-0.159# (0.042)	-0.067+ (0.033)
management	-0.458* (0.265)	-0.604+ (0.273)		-0.412# (0.154)	-0.377+ (0.149)	
administration	-0.11 (0.356)	-0.033 (0.378)		-0.594* (0.355)	-0.610* (0.357)	
finance & accounting	0.122 (0.245)	0.041 (0.256)		-0.864# (0.266)	-0.824# (0.263)	-0.471+ (0.195)
IT- core occupations	-1.034# (0.355)	-1.064# (0.361)	-0.637+ (0.249)	-0.758# (0.224)	-0.761# (0.235)	
business office occ.	-0.238# (0.072)	-0.097 (0.063)	-0.162# (0.059)	-0.387# (0.056)	-0.384# (0.055)	-0.303# (0.053)
temporary office stuff	-0.143 (0.138)	-0.101 (0.124)		-0.140 (0.150)	-0.135 (0.146)	
personal security occ.	-0.171 (0.176)	-0.073 (0.156)		-0.280+ (0.110)	-0.310# (0.107)	
janitor	-0.265 (0.368)	0.115 (0.335)		0.015 (0.055)	-0.014 (0.060)	
security occupations	0.501 (0.345)	0.696* (0.356)		-0.172* (0.099)	-0.166 (0.106)	
legal professions	-0.201 (1.034)	-0.365 (1.110)		-0.532 (0.471)	-0.486 (0.475)	
artists	-0.129 (0.200)	-0.065 (0.207)		-0.108+ (0.046)	-0.099+ (0.047)	
designer	-0.407# (0.192)	-0.282 (0.200)		-0.285# (0.047)	-0.249# (0.049)	
health care occ. w license	-0.669 (0.865)	-0.733 (0.887)		0.212 (0.626)	0.228 (0.641)	
health care occ. w/o license	2.224# (0.289)	2.407# (0.291)	2.151# (0.283)	-0.678# (0.119)	-0.674# (0.121)	-0.602# (0.119)
social occupations	0.185 (0.135)	0.265* (0.151)		-0.847# (0.148)	-0.833# (0.160)	-0.784# (0.153)
teacher	1.336# (0.512)	1.815# (0.522)		-0.107 (0.128)	-0.078 (0.122)	
publicist, more scientific	-0.281 (0.293)	-0.239 (0.291)		-0.279 (0.241)	-0.255 (0.238)	
occupations in body care	0.877+ (0.413)	0.758* (0.400)		-0.373# (0.079)	-0.375# (0.079)	-0.271# (0.076)
occupations in hotel and catering	0.069 (0.073)	0.148+ (0.066)		-0.213# (0.075)	-0.206# (0.076)	-0.134* (0.074)
occup. in room clean., waste coll.	-0.124 (0.081)	-0.003 (0.068)		-0.269# (0.077)	-0.254# (0.074)	-0.216# (0.068)

Table 3.2: (Post-) OLS estimation of the ue and Δue interaction coefficients: the full model with over 628 RHS variables (2nd column), LASSO (3rd column) and adaptive LASSO with 224 RHS variables (4th column). Heteroscedasticity robust clustered standard errors of the coefficients are in parentheses; significance level $^{\#}p < 0.01$, $^{+}p < 0.05$, $*p < 0.1$.

Health care occupations without license to practice medicine (e.g. nurses, geriatric nurses, etc.) are exceptions. The *ue* interaction coefficient is positively significant and very robust. Interestingly, the effect of the Δue interaction variable is negatively significant but it is similar to the other occupations. The unusual positive effect is measured for the time when the employees enter the labour market. The reason for this could be a political intervention, so that the market mechanisms are levered out. The log wages of the employees who have worked in the occupations from Table 3.2 were particularly strongly affected by the relative number of unemployed people in their labour market region. The number of vacancies is another proxy for labour market tightness, see Hershbein and Kahn (2018). By using the IAB-Job-Vacancy Survey, we additionally employ the number of vacancies but are, unfortunately, limited to the years 2007 and 2008, in which occupation-specific data can be matched uniquely to our individual data. The data limitation does not allow to construct the change of vacancies within the 10 years. Individuals, who have worked in occupations that are not listed in the vacancy-survey are not considered in either model.

As the results in Table 3.3 show, there are no difference between the estimated (Δ) *ue* coefficients, meaning that the *ue* variables capture the labour market tightness situation very well. The (log) vacancy coefficient is positively significant but with an effect of 0.009, which is very small. This confirms that higher wages are expected if there are less unemployed people available and the number of vacancies increases. As shown in Table 3.3, the log wages of typical STEM occupations (such as metal and plant engineering, industrial and tool mechanic, car, aircraft construction, maintenance specialists precision mechanic professions) are especially strongly effected by a shortage or a flood of labour force, respectively. For these occupations, the entry situation seems to play an important role for the wages after 10 years, as the interactions of *ue* with the occupations show. The changing labour market situation becomes more important for commercial and transportation professions, which can be explained by the rapid development of online trading.

We now perform substantial robustness checks. We find that the main *ue* and Δue effects are relatively stable over the observation time, as Figure 3.4 shows. One exception is the year 2012 (entry cohort 2002). The Δue coefficient becomes insignificant but *ue* is still the same as in the other years. The economic crisis could be levered by the usual market mechanism in the short run. Furthermore, there are collectively agreed wages in Germany, which cannot easily be reduced—even the if market situation becomes poorer. Consequently, some employers may suspend or not extend fixed-term employment contracts, which increases the unemployment

	Ada LASSO	Ada 0708	Ada 0708vac	FullModel0708vac
ue without interaction	-0.195 $^{\#}$ (0.019)	-0.196 $^{\#}$ (0.059)	-0.195 $^{\#}$ (0.059)	-0.112 (0.995)
Δue without interaction	-0.151 $^{\#}$ (0.015)	-0.346 $^{\#}$ (0.041)	-0.348 $^{\#}$ (0.041)	-0.263 $^{\#}$ (0.065)
Log Vacancies			0.009 $^{\#}$ (0.003)	0.008 $^{\#}$ (0.002)
Interaction between ue and Occupation				
metal and plant engin.	-0.270 $^{\#}$ (0.056)	-0.323 $^{+}$ (0.132)	-0.335 $^{+}$ (0.133)	-0.643 $^{\#}$ (0.158)
ind. and tool mecha.	-0.342 $^{\#}$ (0.054)	-0.672 $^{\#}$ (0.214)	-0.665 $^{\#}$ (0.215)	-0.839 $^{\#}$ (0.254)
car, aircraft constr., maint. spec.	-0.620 $^{\#}$ (0.101)	-0.655 * (0.345)	-0.647 * (0.346)	-0.936 $^{\#}$ (0.362)
electrical and electronic eng.	-0.669 $^{\#}$ (0.074)	-0.869 $^{\#}$ (0.230)	-0.886 $^{\#}$ (0.231)	-0.897 $^{\#}$ (0.256)
building professions	-0.051 $^{+}$ (0.025)	-0.142 (0.090)	-0.146 (0.089)	-0.301 $^{+}$ (0.131)
labourers	-0.024 (0.034)	-0.060 (0.115)	-0.064 (0.115)	-0.295 $^{+}$ (0.134)
technicians	-0.777 $^{\#}$ (0.136)	-1.014 * (0.523)	-1.014 * (0.523)	-0.433 (0.730)
salesman	0.598 $^{\#}$ (0.049)	0.645 $^{\#}$ (0.195)	0.650 $^{\#}$ (0.196)	0.345 (0.212)
IT- core occu.	-0.637 $^{+}$ (0.249)	0.223 (0.620)	0.242 (0.614)	-0.133 (0.885)
commercial occu.	-0.162 $^{\#}$ (0.059)	-0.344 * (0.179)	-0.334 * (0.179)	-0.602 $^{\#}$ (0.224)
health care occ. w/o license	2.151 $^{\#}$ (0.283)	1.837 $^{\#}$ (0.577)	1.771 $^{\#}$ (0.560)	1.407 $^{+}$ (0.623)
Interaction between Δue and Occupation				
precision mechanic	-0.295 $^{\#}$ (0.099)	-0.490 * (0.283)	-0.479 * (0.284)	-0.343 (0.283)
bakers and confectioners	-0.149 $^{+}$ (0.066)	-0.196 (0.363)	-0.190 (0.364)	-0.194 (0.337)
building professions	0.055 $^{+}$ (0.022)	0.149 $^{\#}$ (0.050)	0.149 $^{\#}$ (0.050)	0.057 (0.071)
labourers	-0.044 * (0.025)	0.060 (0.102)	0.058 (0.102)	-0.083 (0.117)
technical drawer	-0.322 $^{+}$ (0.140)	-0.142 (0.188)	-0.135 (0.187)	-0.413 (0.256)
salesman	-0.144 $^{\#}$ (0.044)	-0.189 (0.118)	-0.180 (0.120)	-0.300 $^{+}$ (0.148)
retailers and distributors	-0.204 $^{\#}$ (0.064)	0.044 (0.128)	0.043 (0.129)	-0.044 (0.145)
bank empl. and ins. salesman	-0.670 * (0.365)	-0.912 $^{\#}$ (0.295)	-0.893 $^{\#}$ (0.294)	-1.157 $^{\#}$ (0.314)
other salesmen	-0.424 $^{\#}$ (0.132)	-0.367 (0.253)	-0.366 (0.255)	-0.630 $^{+}$ (0.290)
packer and warehouse	-0.067 $^{+}$ (0.033)	-0.114 (0.097)	-0.114 (0.097)	-0.251 $^{+}$ (0.102)
finance and accounting	-0.471 $^{+}$ (0.195)	-0.308 (0.469)	-0.307 (0.474)	-1.977 $^{\#}$ (0.770)
commercial occupation	-0.303 $^{\#}$ (0.053)	-0.285 (0.206)	-0.293 (0.209)	-0.405 * (0.208)
health care occ. w/o license	-0.602 $^{\#}$ (0.119)	-1.035 $^{\#}$ (0.374)	-1.044 $^{\#}$ (0.375)	-1.038 $^{\#}$ (0.339)
social occupations	-0.784 $^{\#}$ (0.153)	-0.145 (0.224)	-0.144 (0.226)	-0.211 (0.247)
occu. in body care	-0.271 $^{\#}$ (0.076)	0.159 (0.270)	0.166 (0.273)	-0.011 (0.244)
occu. in hotel and catering	-0.134 * (0.074)	-0.204 (0.134)	-0.209 (0.134)	-0.292 $^{+}$ (0.137)
occu. in room clean., waste coll.	-0.216 $^{\#}$ (0.068)	-0.177 (0.125)	-0.166 (0.128)	-0.354 $^{+}$ (0.169)
Contant	3.328 $^{\#}$ (0.074)	3.416 $^{\#}$ (0.175)	3.420 $^{\#}$ (0.175)	4.513 $^{\#}$ (0.691)
N	316711	30147	30147	35124
adj. R2	0.592	0.589	0.589	0.591
AIC	3.90E+05	35000	35000	35000
BIC	3.90E+05	37000	37000	37000

Table 3.3: Post-OLS estimation of the ue and Δue (interaction) coefficients by using adaptive LASSO for all observation years without (log) vacancies (2nd column), for 2007 and 2008 without (log) vacancies (3rd column) and for 2007 and 2008 with (log) vacancies (4th column) as well as the full model with all 628 RHS variables plus (log) vacancies (5th column). Heteroscedasticity robust clustered standard errors of the coefficients are in parentheses; significance level $^{\#}p < 0.01$, $^{+}p < 0.05$, $*p < 0.1$.

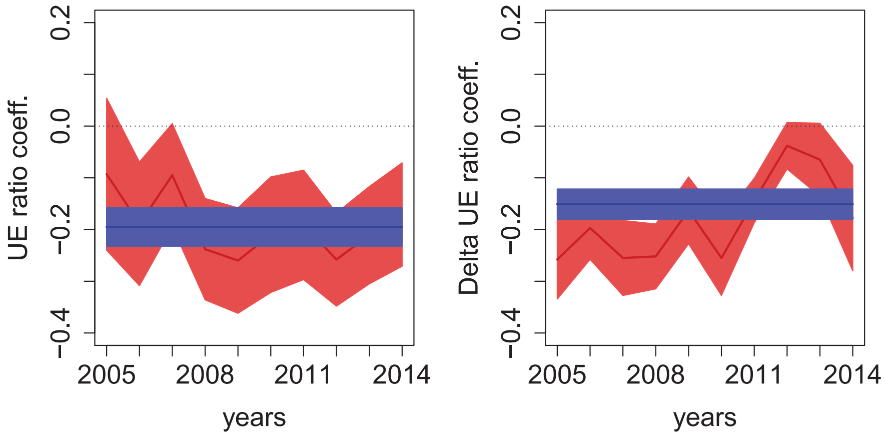

Figure 3.4: Year-by-year (red) and the overall (blue) mean estimation of the *ue* and Δue coefficients by using one model selected by adaptive Lasso including the 95% CI

rate without affecting wages. Let us consider various sub-groups. The results can be found in Table 3.4. First, we estimate the adaptive model for men and women separately (columns 3 and 4). Interestingly, the *ue* effect seems to be stronger for men than for women. Because the interaction terms of the *ue* ratio with the occupations show typically male and female occupations, the difference is driven by the selectivity of women and men into gender-specific occupations, with lower wage perspectives for the females. A list of the proportion of men and women within the occupational fields can be found in Table 4.13 in the Appendix. However, the labour market development is identical for both genders because there is no significant difference in Δue. Thus, men and women benefit to the same extent when the labour market becomes tighter. Furthermore, for comparison purposes, we estimate the model for other subsamples. The decision of an individual having an additional training and/or changing a job or/and region can be driven by insufficient wages, which leads to a potential endogeneity problem. Consequently, we only consider individuals with reduced mobility and look to find whether the previous results change. For the following subsamples, we come to the same conclusion: we limit the men sample to a more homogenous group (column 5; *A (men;edu.)*) consisting of men with a valid school-leaving certificate and who have completed vocational training. We consider individuals who did not change from a typical STEM to non-STEM occupation, and vice versa (column 6; *B (cl1)*), to account for potential income growth perspectives and between job mobility.

	Ada (mean)	men	women	A (men;edu.)	B (cl1)	C (cl2)	D (cl3)	E (cl4)	F (firm)	G (foreigner)	H (firm & foreig)
ue	-0.195 # (0.019)	-0.276 # (0.022)	-0.084 + (0.037)	-0.280 # (0.024)	-0.185 # (0.020)	-0.104 # (0.024)	-0.086 # (0.028)	-0.178 # (0.019)	-0.012 (0.018)	-0.169 # (0.020)	-0.033 * (0.018)
Δue	-0.151 # (0.015)	-0.168 # (0.016)	-0.179 # (0.048)	-0.172 # (0.017)	-0.156 # (0.019)	-0.113 # (0.022)	-0.177 # (0.033)	-0.137 # (0.014)	-0.080 # (0.011)	-0.144 # (0.015)	-0.082 # (0.012)
Interaction ue and Occupation											
MP	-0.270 # (0.056)	-0.155 # (0.055)	0.013 (0.193)	-0.163 # (0.058)	-0.288 # (0.063)	-0.382 # (0.076)	-0.524 # (0.086)	-0.274 # (0.056)	-0.100 + (0.050)	-0.266 # (0.056)	-0.121 + (0.048)
IM	-0.342 # (0.054)	-0.215 # (0.059)	0.086 (0.410)	-0.231 # (0.056)	-0.438 # (0.054)	-0.552 # (0.069)	-0.581 # (0.118)	-0.324 # (0.055)	-0.262 # (0.053)	-0.331 # (0.056)	-0.281 # (0.053)
CA	-0.620 # (0.101)	-0.486 # (0.103)	0.282 (0.587)	-0.484 # (0.104)	-0.718 # (0.108)	-0.864 # (0.116)	-0.947 # (0.132)	-0.600 # (0.100)	-0.293 # (0.088)	-0.599 # (0.105)	-0.336 # (0.080)
EE	-0.669 # (0.074)	-0.530 # (0.082)	0.604 (0.500)	-0.576 # (0.085)	-0.817 # (0.074)	-0.885 # (0.094)	-0.934 # (0.115)	-0.656 # (0.075)	-0.381 # (0.070)	-0.668 # (0.075)	-0.439 # (0.064)
Bui	-0.051 + (0.025)	0.095 # (0.028)	-0.095 (0.134)	0.086 # (0.029)	-0.056 + (0.027)	-0.150 # (0.031)	-0.167 # (0.033)	-0.060 + (0.025)	-0.031 (0.023)	-0.061 + (0.026)	-0.031 (0.023)
La	-0.024 (0.034)	0.047 (0.031)	-0.050 (0.096)	0.066 * (0.036)	-0.040 (0.036)	-0.107 + (0.043)	0.004 (0.057)	-0.036 (0.035)	-0.056 * (0.032)	-0.035 (0.034)	-0.036 (0.033)
Te	-0.777 # (0.136)	-0.635 # (0.119)	-0.403 (0.440)	-0.616 # (0.127)	-1.035 # (0.145)	-1.145 # (0.167)	-0.971 # (0.214)	-0.891 # (0.155)	-0.431 # (0.131)	-0.778 # (0.137)	-0.517 # (0.130)
Sa	0.598 # (0.049)	0.100 (0.076)	0.464 # (0.064)	0.025 (0.079)	0.591 # (0.050)	0.690 # (0.066)	0.654 # (0.075)	0.622 # (0.048)	0.605 # (0.043)	0.588 # (0.050)	0.570 # (0.043)
IT	-0.637 + (0.249)	-0.518 + (0.232)	-0.281 (0.581)	-0.506 + (0.240)	-0.713 # (0.245)	-0.840 + (0.404)	-0.529 + (0.255)	-0.790 + (0.307)	-0.528 + (0.213)	-0.678 # (0.250)	-0.567 # (0.209)
BO	-0.162 # (0.059)	-0.208 + (0.100)	-0.319 # (0.079)	-0.177 * (0.099)	-0.169 # (0.059)	-0.162 + (0.077)	-0.159 * (0.090)	-0.136 + (0.064)	0.069 (0.059)	-0.165 # (0.062)	-0.040 (0.058)
HWo	2.151 # (0.283)	-0.040 (0.355)	1.931 # (0.288)	-0.242 (0.386)	2.136 # (0.278)	2.650 # (0.320)	1.834 # (0.380)	2.206 # (0.298)	2.555 # (0.270)	2.215 # (0.285)	2.258 # (0.272)
Interaction Δue and Occupation											
PM	-0.295 # (0.099)	-0.357 + (0.147)	-0.095 (0.129)	-0.358 + (0.157)	-0.163 (0.179)	-0.497 + (0.223)	-0.290 (0.280)	-0.273 # (0.098)	-0.261 # (0.092)	-0.297 # (0.098)	-0.270 # (0.092)
Ba	-0.149 + (0.066)	-0.060 (0.073)	-0.200 (0.127)	-0.025 (0.093)	-0.131 * (0.069)	-0.037 (0.074)	0.142 (0.115)	-0.154 + (0.066)	-0.056 (0.056)		
Bui	0.055 + (0.022)	0.090 # (0.022)	-0.032 (0.118)	0.087 # (0.022)	0.057 + (0.026)	0.029 (0.029)	0.029 (0.038)	0.048 + (0.022)	0.029 * (0.015)	0.052 + (0.022)	0.025 (0.016)
La	-0.044 * (0.025)	-0.007 (0.024)	-0.023 (0.088)	0.020 (0.034)	-0.044 (0.029)	-0.081 + (0.035)	0.153 + (0.061)	-0.058 + (0.026)	-0.022 (0.022)	-0.048 * (0.025)	-0.020 (0.022)
TD	-0.322 + (0.140)	-0.194 (0.159)	-0.381 * (0.211)	-0.211 (0.164)	0.196 (0.132)	0.159 (0.197)	0.175 (0.236)	-0.274 * (0.152)	-0.384 # (0.139)	-0.318 + (0.140)	-0.383 # (0.139)
Sa	-0.144 # (0.044)	-0.020 (0.036)	-0.225 # (0.069)	-0.033 (0.035)	-0.146 # (0.046)	-0.202 # (0.053)	0.063 (0.080)	-0.154 # (0.054)	-0.133 # (0.033)	-0.148 # (0.043)	-0.135 # (0.034)
RD	-0.204 # (0.064)	-0.160 # (0.061)	-0.211 + (0.106)	-0.165 + (0.066)	-0.192 # (0.070)	-0.196 # (0.072)	0.160 (0.239)	-0.261 # (0.063)	-0.151 + (0.061)	-0.211 # (0.064)	-0.160 # (0.061)
BJ	-0.67 * (0.365)	-0.325 (0.330)	-1.016 # (0.181)	-0.297 (0.320)	-0.637 * (0.357)	-1.335 # (0.281)	-0.715 * (0.432)	-0.547 (0.399)	-0.485 * (0.269)	-0.673 * (0.363)	-0.488 * (0.269)
OS	-0.424 # (0.132)	-0.501 # (0.100)	-0.306 (0.199)	-0.405 # (0.108)	-0.413 # (0.135)	-0.399 + (0.175)	0.268 (0.248)	-0.442 # (0.149)	-0.294 + (0.121)	-0.432 # (0.132)	-0.296 + (0.121)
PW	-0.067 + (0.033)	-0.020 (0.033)	-0.170 (0.111)	0.018 (0.047)	-0.056 (0.036)	-0.087 * (0.051)	-0.043 (0.100)	-0.068 + (0.031)	-0.040 (0.024)	-0.071 + (0.033)	-0.040 (0.025)
FA	-0.471 + (0.195)	-0.146 (0.169)	-0.572 * (0.302)	-0.400 (0.246)	-0.487 + (0.199)	-0.575 + (0.270)	1.545 # (0.414)	-0.689 # (0.238)	-0.540 # (0.201)		
BO	-0.303 # (0.053)	-0.147 # (0.046)	-0.275 # (0.081)	-0.142 # (0.052)	-0.299 # (0.054)	-0.377 # (0.092)	-0.051 (0.172)	-0.324 # (0.060)	-0.281 # (0.049)	-0.311 # (0.053)	-0.287 # (0.049)
HWo	-0.602 # (0.119)	-0.406 + (0.163)	-0.387 # (0.115)	-0.424 + (0.194)	-0.582 # (0.118)	-0.813 # (0.117)	0.305 (0.202)	-0.579 # (0.119)	-0.555 # (0.099)	-0.610 # (0.119)	-0.559 # (0.099)
Soc	-0.784 # (0.153)	-0.109 (0.093)	-0.837 # (0.181)	-0.120 (0.108)	-0.776 # (0.154)	-0.842 # (0.173)	-0.269 (0.235)	-0.818 # (0.136)	-0.705 # (0.146)	-0.780 # (0.152)	-0.710 # (0.143)
BC	-0.271 # (0.076)	-0.068 (0.169)	-0.350 # (0.085)	0.002 (0.179)	-0.255 # (0.076)	-0.227 # (0.088)	-0.192 (0.119)	-0.272 # (0.074)	0.054 (0.070)	-0.265 # (0.076)	0.042 (0.071)
HC	-0.134 * (0.074)	-0.128 (0.083)	-0.084 (0.095)	-0.141 (0.099)	-0.124 * (0.075)	-0.238 + (0.096)	-0.010 (0.148)	-0.125 * (0.073)	-0.022 (0.043)	-0.138 * (0.074)	-0.022 (0.043)
Wa	-0.216 # (0.068)	-0.076 (0.051)	-0.320 + (0.136)	-0.040 (0.048)	-0.197 # (0.071)	-0.268 # (0.086)	-0.048 (0.131)	-0.212 # (0.069)	-0.128 # (0.050)	-0.225 # (0.069)	-0.136 # (0.051)
Constant	3.328 # (0.074)	3.312 # (0.091)	2.685 # (0.070)	3.285 # (0.097)	3.133 # (0.040)	3.359 # (0.072)	2.734 # (0.062)	3.233 # (0.090)	0.691 # (0.133)	3.328 # (0.075)	0.702 # (0.135)
N	316711	177008	139703	142608	285613	198176	124752	273207	316711	308670	308670
adj. R2	0.592	0.564	0.556	0.554	0.593	0.605	0.594	0.583	0.636	0.593	0.637
AIC	3.90E+05	1.30E+05	2.20E+05	9.70E+04	3.60E+05	2.30E+05	1.20E+05	3.40E+05	3.50E+05	3.80E+05	3.40E+05

Table 3.4: Post-OLS estimation of the ue and Δue and their interaction coefficients. 'Ada' in column 2 is the default model chosen by adaptive LASSO. Columns 3 – 12 include the estimated coefficients of modified models through changing sample composition or including further variables. The abbreviation of the occupations can be found in Table 4.12 in the Appendix

We restrict the sample to employees who do not change the employer as another source of endogeneity driven by mobility during their life course (column 7; *C (cl2)*). The main (Δ) *ue* coefficients become slightly weaker. A few occupations (e.g. bakers and confectioners, and building professions) become insignificant. The results hardly change if we more homogenise the sample *C (cl2)* and exclude individuals who have switched between occupations (column 8, *D (cl3)*). Most effects become less significant, especially *ue*, because it concerns a rather inflexible or inelastic sample group. However, there is no other conspicuousness.

By considering all of the employees holding up to a secondary school-leaving qualification (column 9, *E (cl4)*), differences to the initial adaptive model are hardly visible. These individuals have no option to study and are therefore less flexible. The results hold even if we augment the model by potential endogenous variables, such as the average firm wages (and its change) and the (change in the) proportion of foreigners who worked in the firm of the observed employees (Table 3.4, columns 10 - 12). The main (Δ) *ue* coefficients become solely weaker. In summary, the strength of *ue* depends on the occupations. Wages of STEM occupations react usually more sensitively to labour market tightness, indicating that technological progress associated with higher labour demand puts additional pressure in that field.

3.6 Differences Between the Quantiles of the Wage Distribution

So far, we have presented evidence of tightness effects on the average log wage. However, there is a variation in the wage growth within each occupation, as shown in Figure 3.5, for some selected occupations. Bar plots of the wage growth for different quantile levels of the occupation fields are illustrated in Figure 3.7. This variation happens not just because of within-variation arising from sub-summation of three-digit occupations to the employed 54 occupational fields but also because of uncontrolled individual heterogeneity. For instance, there is a much higher wage growth for employees in IT occupations at the first 10 years in the labour market compared with employees within the health care professions (without license to practice medicine). In addition, we observe a much stronger variation between the different quantile levels for the latter, but also social occupations in general. That means that the wage inequality rises much more between the health care workers than between the employees in the IT professions. A question arises here: are there different effects on the log wage - quantiles through the (changing) labour market

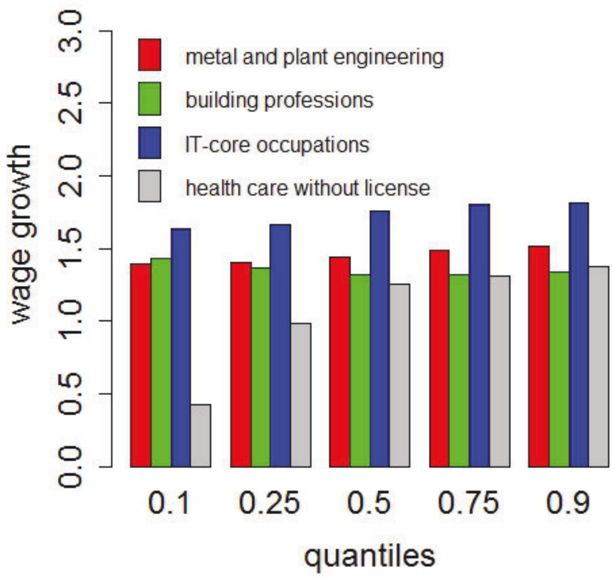

Figure 3.5: The quantiles of the ten-years wage growth for four occupations.

situation (Q3)? That is, do potentially more or less productive workers benefit more from tighter labour markets? To answer these questions, we estimate different quantile levels of the log wages by using the unconditional quantile regression by Firpo (2009). The RHS variables for each quantile level $\tau = \{0.25, 0.5, 0.75\}$ are separately selected by using adaptive LASSO, see Table 4.9 in the Appendix. For comparison purposes, we also estimate the different quantile levels of the log wages through all 628 RHS variables, see Table 4.10 in Appendix. Again, the model fit values remain the same, although almost two third of the variables are dropped through the adaptive LASSO algorithm. Considering the results of the estimations, the coefficients for (Δ) ue variables vary strongly between the different quantile regressions. The (Δ) ue coefficient is negative or not significant. In particular, the 0.25 quantile and the median regressions show a strong negative ue effect, which is rather weak for the upper quantile level of the log wages.

Figure 3.7 shows the estimated ue (left-hand side) and Δue (right-hand side) coefficients by using mean regression (blue line) and unconditional quantile regression (red line) for these different models. After choosing and estimating the models separately for each quantile level, there are hardly any deviations from the mean coefficients. In the reduced model, which is selected by adaptive LASSO, the effects for the upper quantiles become weaker. In addition, the ue effect on the median log wages is more strongly negatively than on the mean. This means that 25% of the

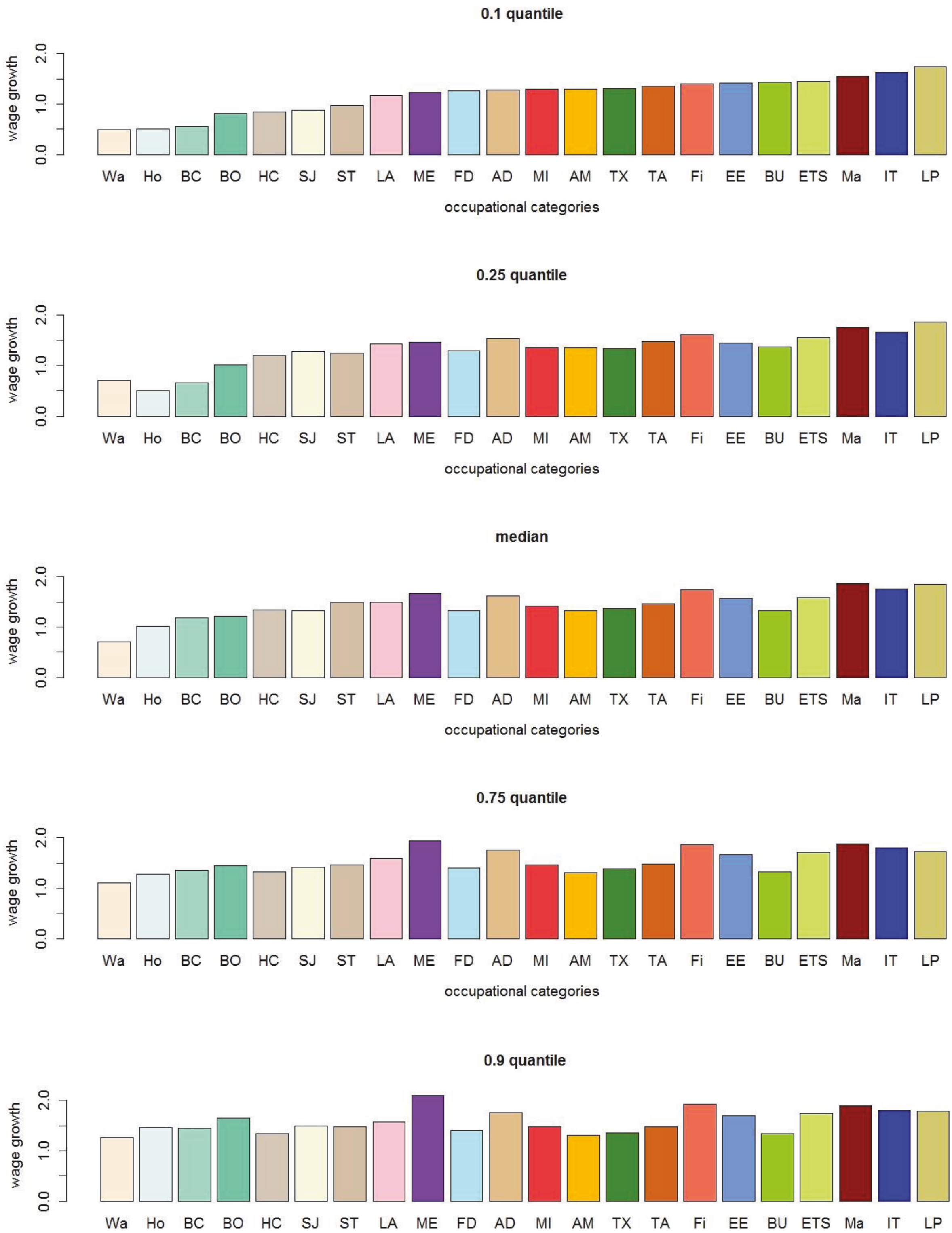

Figure 3.6: Quantiles of the occupational categories. For reasons of clarity 54 observed professions were summarised into 22 categories. A list of the keys is given in Table 4.12 in the Appendix.

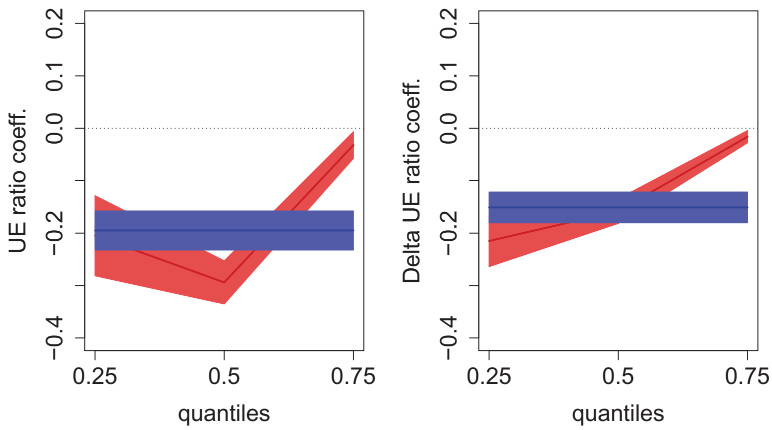

Figure 3.7: Estimated coefficients of ue (lhs) and Δ ue (rhs) by using post-OLS estimation of adaptive LASSO selected different models for mean (blue) and 0.25, 0.5 and 0.75 quantiles including 95% CI.

lowest wages in our sample are especially strong affected by a high unemployment ratio at the first job. If the unemployment ratio was 1% above the average at the first job, then 25% of the lowest wages shrinks by approximately 0.2% ten years after the first job. This reduction amounts only 0.032% for 25% of the highest wages. Interestingly, some (Δ) ue occupation interaction variables are chosen from the adaptive LASSO for a specific quantile, but not for the mean regression. The different wage levels between the occupations could be a reason for these results. As mentioned earlier, IT-specialists earn much more than nurses. The average wage of nurses lies far below the average wage of all individuals within the sample. In some cases, the interaction effects show positive signs, such as nursing or social occupations. However, this is observed here for the lower quantiles – the effect revolves around. The positive effect remains over the quantile levels for salesmen. This curious result occurs more frequently for variables that are not selected by adaptive LASSO for the mean regression. This confirms that the variable selection is important to find the relevant drivers of the (mean) log wages, although it makes a detached interpretation of quantile regression results difficult. In summary, we can now answer our three questions from the beginning in the affirmative. We find a significant negative correlation between the unemployment ratio and log wages. The elasticities of -0.195 (reduced model) or -0.078 (full model), respectively, at the entry time point, and -0.151 (reduced model) or -0.059 (full model), for the change in the first 10 years in the labour market varies between different occupations and, thus, between men and women. Furthermore, there are also differences in the quantile levels of the log wages, which are empirically measurable.

3.7 Concluding Remarks

This chapter investigates the ten-year log wage development of employees who finish a vocational training and start their first job. The focus lies on the measuring of the effect of lower labour force availability on the individual wages. The underlying economic idea is that a workforce shortage affects wages positively if the labour demand remains constant or rises. To verify this hypothesis, we measure labour market tightness by using the ratio of unemployed to employed people on occupational and regional level. Then, we look for a tightness effect on the individual wages for 54 different occupational fields. Therefore, the individual employment biographies (IEB) data are collected, which include information of individuals before they enter the labour market (e.g. during the vocational training) and during the first 10 years after starting their first job. In addition, we match data with firm and labour market characteristics to the individual data and receive a linked employee-employer dataset. First, a regression model with over 628 RHS variables is run. Avoiding collinearity problems, the set of variables is reduced by applying a variable selection by the adaptive LASSO approach from Zou (2006). The selected model is quite robust and the aim variables, unemployment–employees (ue) ratio at the first job and its change during the first 10 years after entering the labour market, show the expected negative signs. The results indicate rising log wages if the labour supply becomes tighter. This effect is stronger for some occupations, such as several engineer groups, technicians, IT professions and commercial occupations. Interestingly, the effect seems to reverse for employees in health care occupations, such as nursing. Moreover, it is possible to differentiate between typical male and female occupations. It seems then that the log wages of men are mainly affected by labour market tightness in technical occupations. By running an unconditional quantile regression, different effects in the lower (poor) than in the upper (rich) quantiles become visible. In addition, separate year-by-year estimations reveal differing effects during the observation time. Interestingly, some typical female occupations (related to our observation data), such as saleswomen and health care occupation without licence, reveal a reverse picture. The unemployment effects are positive and very robust. The labour market mechanisms seems to be completely levered. In this chapter, we cannot solve this puzzle but it would be interesting to know whether the market mechanism differ between the genders and what the reasons are for this effect.

Chapter 4

Conclusion and Prospect

In summary, this thesis gives an insight into two specific issues of empirical research in regional science. The first part proposes a method for measuring an individual attitude, such as *environmental awareness* by using Internet queries and estimating a Multiple-Indicator Multiple-Causes model. The advantages of this approach are comparability between regions and the rather less time-consuming and less-expensive access to the data. However, the 200 different words and phrases for 81 Russian regions in this study are selected by hand, which could bias the findings. Consequently, we need to clarify whether various words and phrases have a similar meaning and weight for the population of different regions. Thus, a stronger link between empirical environmental research and linguistics is required for future investigations. In addition, only the number of Internet entries from the search machine ©Yandex is considered. ©Yandex has currently a 50% market share in Russia. Approximately one half belongs to ©Google. It is assumed the distributions of the search entries from both search engines are similar, but it has been not possible to confirm this assumption yet. The empirical analysis for 81 Russian regions for a period from January 2014 until April 2016 reveals a strong negative correlation between temperature and regional *environmental awareness*, as well as a rather positive non-linear relationship with regional wealth. Furthermore, it seems that *environmental awareness* decreases from the Eastern to the Western part of Russia and is higher in the Arctic areas than in the warmer South. In addition, the Russian population might be more strongly interested in environmental topics in the colder winter than in the warmer months. Therefore, seasonal variation can be found. It would be interesting to investigate longer time-series and to examine whether there are any trends or extraordinary peaks. Furthermore, this study can be extended to other countries and regions. The data access to general regional characteristics of OECD countries, as well as sub-regions is often open, only the Internet queries need to be requested from the search engine companies. Furthermore, the proposed method can be also applied to measure other individual attitudes or further research question (i.e. estimating a kind of political attitude index).

A further issue of empirical research deals with the extraction of one aspect at regional level, here labour market tightness, and its effect on individual wages. Labour market tightness is measured through the number of unemployed people divided by the number of employees within an occupational field in a specific region. Inspired by Mincer (1974), a linear regression model for the log-wages of very young employees is estimated. The *least absolute shrinkage and selection operator* from Zou (2006) reduces the number of RHS variables by two-thirds without impairing the model fit. The estimated effects of rising wages due to labour market tightness is especially strong for technical professions, such as several engineer groups, IT occupations, technicians, and also some commercial occupations. This confirms the presumption that labour market tightness is especially driven by technological progress. Interestingly, health care occupations (e.g. nursing) reveal a complete reverse relationship. Am increase in the labour supply seems to be correlated with higher wages. The reason for this unusual effect could be political interventions and the increasing demand for more health care due to expected demographic developments. Therefore, the market mechanism might be levered out due to specific regulations in terms of the funding of the health care system. Furthermore an unconditional quantile regression by Firpo, Fortin, and Lemieux (2009) is performed to investigate the influence of occupational-regional specific labour market tightness on different quantile levels of the log wage distribution. The labour market tightness relationship is stronger for the lower and the median than for the upper quantiles. This effect becomes more distinctive for social occupations, professions in body care, hotel and catering as well as room cleaning. Wages react more elastically to changes in the relative number of unemployed people. Technical and commercial professions, who earn above average, seem to profit rather from a relative shrinking labour supply. Because of the German contribution assessment threshold, academics are hardly considered in this analysis. The potential substitute or complementary effects of university graduates on the wages of non-academics are controlled, but it would be interesting to measure the effect of labour market tightness on the log-wages of academics in general. Furthermore, the data that we have used are provided for the youngest cohort up to 2014. It would be very interesting to replay the analysis for more current data.

Bibliography

Acemoglu, D., and D. Autor. (2011). “Skills, tasks and technologies: Implications for employment and earnings.” *Handbook of Labor Economics* 4:1043–1171.

Acemoglu, D., and P. Restrepo. (2018). “The race between man and machine: Implications of technology for growth, factor shares, and employment.” *American Economic Review* 108 (6): 1488–1542.

Ahrens, A., M. Schaffer, and C. Hansen. (2018). *Prediction, model selection, and causal inference with regularized regression.* Technical report. London Stata Conference 2018 at the Cass Business School.

Ajzen, I. (1991). “The theory of planned behavior.” *Organizational Behavior and Human Decision Processes* 50 (2): 179–211.

Autor, D. H., and D. Dorn. (2009). “Inequality and specialization: the growth of low-skill service jobs in the United States.” *NBER Working Paper Series* 15150.

Autor, D. H., L. F. Katz, and M. S. Kearney. (2006). “The polarization of the US labor market.” *American Economic Review* 96 (2): 189–194.

———. (2008). “Trends in US wage inequality: Revising the revisionists.” *The Review of Economics and Statistics* 90 (2): 300–323.

Autor, D. H., F. Levy, and R. J. Murnane. (2003). “The skill content of recent technological change: An empirical exploration.” *The Quarterly Journal of Economics* 118 (4): 1279–1333.

Axelrad, H., M. Malul, and I. Luski. (2018). “Unemployment among younger and older individuals: Does conventional data about unemployment tell us the whole story?” *Journal for Labour Market Research* 52 (1): 3.

Blanchflower, D. G., and A. J. Oswald. (2008). *Wage curve.* Springer.

Borah, B. J., and A. Basu. (2013). “Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence.” *Health economics* 22 (9): 1052–1070.

Brajer, V., R. W. Mead, and F. Xiao. (2011). “Searching for an Environmental Kuznets Curve in China’s air pollution.” *China Economic Review* 22 (3): 383–397.

Brosseau-Liard, P. E., V. Savalei, and L. Li. (2012). “An investigation of the sample performance of two nonnormality corrections for RMSEA.” *Multivariate Behavioral Research* 47 (6): 904–930.

Buehn, A., and M. R. Farzanegan. (2013). “Hold your breath: A new index of air pollution.” *Energy Economics* 37:104–113.

Cameron, A. C., and D. L. Miller. (2015). “A practitioner’s guide to cluster - robust inference.” *Journal of Human Resources* 50 (2): 317–372.

Card, D., and T. Lemieux. (2001). “Can falling supply explain the rising return to college for younger men? A cohort-based analysis.” *The Quarterly Journal of Economics* 116 (2): 705–746.

Chernozhukov, V., C. Hansen, and M. Spindler. (2015). “Valid post–selection and post–regularization inference: An elementary, general approach.” *Annual Review of Economics* 7 (1): 649–688.

Diederich, J., and T. Goeschl. (2014). “Willingness to pay for voluntary climate action and its determinants: Field-experimental evidence.” *Environmental and Resource Economics* 57 (3): 405–429.

Dinda, S. (2004). “Environmental Kuznets curve hypothesis: A survey.” *Ecological Economics* 49 (4): 431–455.

Dustmann, C., J. Ludsteck, and U. Schönberg. (2009). “Revisiting the German wage structure.” *Quarterly Journal of Economics* 124 (2).

European Commission. (2008). *Europeans attitudes towards climate change. Special Eurobarometer 300 [online].* Technical report. European Parliament / European Commission. Accessed October 2015, http://ec.europa.eu/public_opinion/archives/ebs/ebs_300_full_en.pdf.

Expert RA. (2018). *Rating of regions in Russia [online].* Accessed August 2018, https://raexpert.ru/rankingtable/region_climat/2015/tab02.

Fan, J., and R. Li. (2001). "Variable selection via nonconcave penalized likelihood and its oracle properties." *Journal of the American Statistical Association* 96 (456): 1348–1360.

Federal Employment Agency Germany. (2017). *Fachkräfteengpassanalyse.* Technical report. Statistik/Arbeitsmarktberichtserstattung, Nürnberg.

Federal Statistics Service of Russia. (2016). *RFSSS public database [online].* Federal Statistics Service of Russia (2016). Accessed January 2016, `http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/`.

Firpo, S., N. M. Fortin, and T. Lemieux. (2009). "Unconditional quantile regressions." *Econometrica* 77 (3): 953–973.

———. (2011). "Occupational tasks and changes in the wage structure." *IZA Discussion Paper.*

Fortin, N., and T. Lemieux. (2016). "Inequality and Changes in Task Prices: Within and between Occupation Effects." In *Inequality: Causes and consequences,* 195–226. Emerald Group Publishing Limited.

Fosten, J., B. Morley, and T. Taylor. (2012). "Dynamic misspecification in the environmental Kuznets curve: evidence from CO_2 and SO_2 emissions in the United Kingdom." *Ecological Economics* 76:25–33.

Goldberger, A. S., and R. Hauser. (1971). "The treatment of unobservable variables in path analysis." *Sociological Methodology* 3 (8): 1–8.

Goldin, C. D., and L. F. Katz. (2009). *The race between education and technology.* Harvard University Press.

Goos, M., A. Manning, and A. Salomons. (2014). "Explaining job polarization: Routine-biased technological change and offshoring." *American Economic Review* 104 (8): 2509–26.

Grossman, G. M., and A. B. Krueger. (1995). "Economic growth and the environment." *The Quarterly Journal of Economics* 110 (2): 353–377.

Hartigan, J. A., and M. A. Wong. (1979). "Algorithm AS 136: A k-means clustering algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1): 100–108.

He, J., and P. Richard. (2010). "Environmental Kuznets curve for CO 2 in Canada." *Ecological Economics* 69 (5): 1083–1093.

Hershbein, B., and L. B. Kahn. (2018). “Do recessions accelerate routine-biased technological change? Evidence from vacancy postings.” *American Economic Review* 108 (7): 1737–72.

Hiramatsu, A., K. Kurisu, and K. Hanaki. (2015). “Environmental consciousness in daily activities measured by negative prompts.” *Sustainability* 8 (1): 24.

Hsu, A., D. C. Esty, M. Levy, and A. de Sherbinin. (2016). *2016 Environmental Performance Index (EPI).* Technical report. `https://doi.org/10.13140/RG.2.2.19868.90249`.

Huang, W. M., G. W. Lee, and C. C. Wu. (2008). “GHG emissions, GDP growth and the Kyoto Protocol: A revisit of Environmental Kuznets Curve hypothesis.” *Energy Policy* 36 (1): 239–247.

IEA Statistics. (2017). *CO_2 Emissions from Fuel Combustion [online].* Technical report. International Energy Agency, Paris, France. Accessed August 2018, `https://www.iea.org/publications/freepublications/publication/CO2EmissionsfromFuelCombustionHighlights2017.pdf`.

Inglehart, R. (1990). *Culture Shift in Advanced Industrial Society.* 1st ed. Princeton University Press, Princeton.

Jöreskog, K. G., and A. S. Goldberger. (1975). “Estimation of a model with multiple indicators and multiple causes of a single latent variable.” *Journal of the American Statistical Association* 70 (351a): 631–639.

Kerneck, B., and B. Oertel. (2016). *Russland: Seine Städte und Regionen.* 1st ed. Komet, Köln.

Ketchen Jr, D. J., and C. L. Shook. (1996). “The application of cluster analysis in strategic management research: an analysis and critique.” *Strategic Management Journal:* 441–458.

Khakimova, D., S. Lösch, D. Wende, H. Wiesmeth, and O. Okhrin. (2019). “Index of environmental awareness through the MIMIC approach.” *Papers in Regional Science* 98 (3): 1419–1441.

Kim, C., and A. Sakamoto. (2008). “The rise of intra-occupational wage inequality in the United States, 1983 to 2002.” *American Sociological Review* 73 (1): 129–157.

Koenker, R. (2005). *Quantile regression.* 38. Cambridge University Press.

Kozeltsev, M. L., E. Strukova, A. A. Golub, and A. Martusevich. (2013). "The challenge of reforming environmental regulation in Russia." In *The Oxford Book of the Russian Economy,* 426–450. Oxford University Press.

Lisciandra, M., and C. Migliardo. (2017). "An empirical study of the impact of corruption on environmental performance: evidence from panel data." *Environmental and Resource Economics* 68 (2): 297–318.

Lorenzoni, I., and N. F. Pidgeon. (2006). "Public views on climate change: European and USA perspectives." *Climatic Change* 77 (1): 73–95.

Lösch, S., O. Okhrin, and H. Wiesmeth. (2017). "Diffusion of environmental awareness." *diffusion-fundamentals.org* 30 (2). Accessed March 2018, `http://diffusion.uni-leipzig.de/pdf/volume30/diff_fund_30(2017)02.pdf`.

———. (2018a). "Awareness of climate change: Differences among Russian regions." *Area Development and Policy* 4 (3): 284–307. Accessed August 2019, `https://doi.org/10.1080/23792949.2018.1514982`.

———. (2018b). "Awareness of Climate Change - Focus on the Russian Arctic Zone." *Proceedings of the International Research Workshop on Information Technologies and Mathematical Modeling for Efficient Development of Arctic Zone* 2109:38–42. Accessed June 2018, `http://ceur-ws.org/Vol-2109/paper-07.pdf`.

Maddison, D. (2006). "Environmental Kuznets curves: A spatial econometric approach." *Journal of Environmental Economics and Management* 51 (2): 218–230.

Marshall, A. (1920). *Principles of economics (1920: 1st.*

Meinshausen, N., and P. Bühlmann. (2006). "Variable selection and high-dimensional graphs with the lasso." *Annals of Statistics* 34:1436–1462.

Meteo. (2018). *Weather Data for Russia [online].* Accessed March 2018, `http://aisori.meteo.ru/climateR`.

Mincer, J. (1974). "Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2."

Moscarini, G., and F. Postel-Vinay. (2016). "Wage posting and business cycles." *American Economic Review* 106 (5): 208–13.

Nijkamp, P., and J. Poot. (2005). "The last word on the wage curve?" *Journal of Economic Surveys* 19 (3): 421–450.

Organisation for Economic Co-operation and Development. (2016). *OECD.Stat public database [online].* Organisation for Economic Co-operation and Development. Accessed January 2016, `http://stats.oecd.org/Index.aspx?datasetcode=REG_DEMO_TL2#`.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. `https://www.R-project.org/`.

Rabe, B. G., C. P. Borick, and E. Lachapelle. (2011). "Climate compared: Public opinion on climate change in the United States & Canada."

Rosseel, Y. (2012). "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48 (2): 1–36. `http://www.jstatsoft.org/v48/i02/`.

Rousseeuw, P. J. (1987). "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." *Journal of Computational and Applied Mathematics* 20:53–65.

Rupasingha, A., S. J. Goetz, D. L. Debertin, and A. Pagoulatos. (2004). "The environmental Kuznets curve for US counties: A spatial econometric analysis with extensions." *Papers in Regional Science* 83 (2): 407–424. doi:`10.1007/s10110-004-0199-x`. `https://doi.org/10.1007/s10110-004-0199-x`.

Satorra, A., and P. M. Bentler. (1994). "Corrections to test statistics and standard errors in covariance structure analysis." In *Latent Variables Analysis: Applications for Developmental Research Sage,* edited by A. von Eye and C. C. Clogg, 399–419. Thousand Oaks, CA.

Solow, R. M. (1956). "A contribution to the theory of economic growth." *The Quarterly Journal of Economics:* 65–94.

Soyez, K. (2012). "How national cultural values affect pro-environmental consumer behavior." *International Marketing Review* 29 (6): 623–646.

Soyez, K., S. Hoffmann, S. Wünschmann, and K. Gelbrich. (2009). "Proenvironmental value orientation across cultures: Development of a German and Russian scale." *Social Psychology* 40 (4): 222–233.

Spitz-Oener, A. (2006). “Technical change, job tasks, and rising educational demands: looking outside the wage structure.” *Journal of Labor Economics* 24 (2): 235–270.

StatCounter Global Stats. (2019). *StatCounter Global Stats database [online].* Accessed June 2019, `http://gs.statcounter.com/search-engine-market-share/all/russian-federation`.

Statistisches Bundesamt Deutschland. (2017). *Bevölkerungvorausberechnung - Vorausberechnung bis 2060 Ergebnisse der 13. koordinierten Bevölkerungsvorausberechnung.* Technical report. Statistische Bundesamt Deutschland, Wiesbaden.

Stern, D. I. (2004). “The rise and fall of the environmental Kuznets curve.” *World Development* 32 (8): 1419–1439.

Tanaka, J. S. (1993). “Multifaceted conceptions of fit in structural equation models.” *Sage focus editions* 154:10–10.

Thorndike, R. L. (1953). “Who belongs in the family?” *Psychometrika* 18 (4): 267–276.

Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso.” *Journal of the Royal Statistical Society. Series B (Methodological):* 267–288.

Wang, Y.-C. (2013). “Functional sensitivity of testing the environmental Kuznets curve hypothesis.” *Resource and Energy Economics* 35 (4): 451–466.

Wiesmeth, H., and D. Häckl. (2017). “Integrated environmental policy: A review of economic analysis.” *Waste Management & Research* 35 (4): 332–345.

Witman, S. (2017). “Arctic permafrost thaw would amplify climate change.” *Eos* 98. Accessed September 2018, `https://doi.org/10.1029/2017EO068489`.

World Bank. (2018). *Worldbank public database [online].* Accessed January 2019, `http://search.worldbank.org/`.

Yang, H., J. He, and S. Chen. (2015). “The fragility of the Environmental Kuznets Curve: Revisiting the hypothesis with Chinese data via an “Extreme Bound Analysis”.” *Ecological Economics* 109:41–58.

Yuan, K.-H., and P. M. Bentler. (2000). “Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data.” *Sociological Methodology* 30 (1): 165–200.

Zou, H. (2006). "The adaptive lasso and its oracle properties." *Journal of the American Statistical Association* 101 (476): 1418–1429.

Appendix

Appendix for Chapter 2

Table 4.1: Descriptive Statistics of the observed variables.

Variable	Description	Min.	Mean	Max.	Source
Climate Change Queries (2014)	Share of the number of queries on all queries per year	1.00E-04	1.70E-04	2.99E-04	©Yandex
Climate Change Queries (2015)	Share of the number of queries on all queries per year	1.16E-04	1.72E-04	3.33E-04	©Yandex
Endangered Environment Queries (2014)	Share of the number of queries on all queries per year	2.80E-04	4.07E-04	1.05E-03	©Yandex
Endangered Environment Queries (2015)	Share of the number of queries on all queries per year	2.77E-04	4.01E-04	5.11E-04	©Yandex
Politic Queries (2014)	Share of the number of queries on all queries per year	2.74E-06	6.69E-06	1.68E-05	©Yandex
Politic Queries (2015)	Share of the number of queries on all queries per year	2.15E-06	6.86E-06	2.47E-05	©Yandex
Science Queries (2014)	Share of the number of queries on all queries per year	4.34E-06	7.68E-06	1.81E-05	©Yandex
Science Queries (2015)	Share of the number of queries on all queries per year	3.80E-06	6.59E-06	1.40E-05	©Yandex
Renewable Energies /Technologies Queries (2014)	Share of the number of queries on all queries per year	6.48E-06	1.40E-05	3.38E-05	©Yandex
Renewable Energies/ Technologies Queries (2015)	Share of the number of queries on all queries per year	5.38E-06	1.20E-05	3.76E-05	©Yandex
Regional GDP per capita per year (2014)	in USD Dollar, constant prices, constant ppp, based on year 2010	4590	19200	187700	OECD.stat
Regional GDP per capita per year (2015)	in USD Dollar, constant prices, constant ppp, based on year 2010	4973	19600	191100	own estimations
Internet access (2014)	Percent on regional households	26.0	59.1	87.0	OECD.stat
Internet access (2015)	Percent on regional households	25.4	63.3	83.5	OECD.stat
Proportion of labour force in manufacturing sector (2014)	Percent on regional labour force	1.0	13.8	27.2	OECD.stat
Proportion of labour force in mining sector (2014)	Percent on regional labour force	0.0	2.5	22.3	OECD.stat
Air pollution per capita (2014)	in tonnes per head	0.1	16.4	199.4	RSSSS
Air pollution per capita (2015)	in tonnes per head	0.1	16.8	234.2	RSSSS
Private Vehicle Rate (2014)	Percent on regional population	8.0	27.5	48.7	
Agglomeration (2014)	Population per km^2	0.1	106.7	4674.0	OECD.stat
Share of people older than 65 years (2014)		9.2	23.0	29.5	OECD.stat
Proportion of labour force working in tertiary sector	Percent on regional labour force	26.6	53.7	71.2	OECD.stat
Environmental protection costs in 2007 prices (land and water protection)	Tsd. Rubel per km^2 regional surface (for 2014)	0.089	25.270	328.200	Ministry of Natural Resources of Russia
Environmental protection costs in 2007 prices (land and water protection)	Tsd. Rubel per km^2 regional surface (for 2015)	0.000	25.100	334.700	own estimations
Dumped contaminated (no cleaned) and insufficiently purified water	Tsd. cubic meter per km^2 surface (for 2014)	0.000	2.275	17.860	Ministry of Natural Resources of Russia
Dumped contaminated (no cleaned) and insufficiently purified water	Tsd. cubic meter per km^2 surface (for 2015)	0.000	2.532	41.080	own estimations

Table 4.2: Estimated parameters for 2014 and 2015 using the MIMIC approach with reduced cause variable set.

	2014 (E)	2015 (E)	pooled (E)	2014 (F)	2015 (F)	pooled (F)	2014 (G)	2015 (G)	pooled (G)	2014 (H)	2015 (H)	pooled (H)
λ												
Climate Change	0.273	0.334	0.27	0.266	0.334	0.265	0.266	0.344	0.267	0.26	0.343	0.259
Endangered Environment	0.137 **	0.118	0.143 **	0.136 **	0.117	0.141 **	0.137 **	0.117	0.146 **	0.133 **	0.116	0.144 **
	(0.247)	(0.196)	(0.246)	(0.245)	(0.185)	(0.243)	(0.236)	(0.199)	(0.24)	(0.239)	(0.184)	(0.236)
Politic	0.116 **	0.055	0.089 **	0.115 **	0.054	0.089 **	0.11 **	0.05	0.085 **	0.104 **	0.05	0.082 **
	(0.163)	(0.124)	(0.122)	(0.164)	(0.122)	(0.121)	(0.157)	(0.11)	(0.113)	(0.165)	(0.108)	(0.111)
Science	0.294 ***	0.25 **	0.28 ***	0.302 ***	0.251 **	0.285 ***	0.314 ***	0.25 **	0.297 **	0.336 ***	0.251 **	0.311 **
	(0.182)	(0.269)	(0.277)	(0.205)	(0.265)	(0.283)	(0.282)	(0.271)	(0.344)	(0.339)	(0.269)	(0.391)
Renewable Energies/Technologies	0.181 ***	0.243 ***	0.218 ***	0.181 ***	0.244 ***	0.22 ***	0.173 **	0.239 ***	0.205 ***	0.168 **	0.24 ***	0.203 ***
	(0.19)	(0.2)	(0.165)	(0.193)	(0.202)	(0.172)	(0.199)	(0.169)	(0.159)	(0.225)	(0.175)	(0.174)
β												
GRP per capita in ppp							0.336 **	0.269 **	0.262 **	0.377 ***	0.302 **	0.296 **
							(0.117)	(0.104)	(0.089)	(0.085)	(0.131)	(0.092)
Environmental Protection Costs	0.127	-0.059	-0.027	0.15	-0.021	0.012				0.088	0.012	0.018
	(0.098)	(0.132)	(0.069)	(0.081)	(0.156)	(0.067)				(0.056)	(0.152)	(0.057)
Contaminated Water	-0.367 **	-0.144	-0.17	-0.331 **	-0.16	-0.169	-0.212 **	-0.19 **	-0.17 **	-0.197 **	-0.175	-0.138
	(0.129)	(0.155)	(0.116)	(0.115)	(0.166)	(0.107)	(0.078)	(0.095)	(0.083)	(0.098)	(0.149)	(0.079)
Access to Internet				-0.268	-0.149	-0.211				-0.317 **	-0.202	-0.251 **
				(0.144)	(0.194)	(0.114)				(0.126)	(0.194)	(0.107)
Year Dummy (2015==1)			-0.125			-0.093			-0.142			-0.109
			(0.076)			(0.073)			(0.073)			(0.064)
Observation	77	77	154	77	77	154	77	77	155	77	77	154
Degrees of Freedom	13	13	17	17	17	21	13	13	17	21	21	25
AIC	986	1020	2020	980	1020	2013	973	1014	2004	959	1014	1991
CFI	0.977	0.843	0.811	0.859	0.817	0.768	0.932	0.815	0.789	0.857	0.771	0.749
RMSEA	0.052	0.126	0.126	0.126	0.121	0.132	0.098	0.143	0.141	0.123	0.13	0.135

Standard errors in parentheses; significance level: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.3: Estimated parameters for 2014 and 2015 using the MIMIC approach as in Table 2.1, but with reduced indicator variable set.

	2014 (I)	2015 (I)	pooled (I)	2014 (J)	2015 (J)	pooled (J)	2014 (K)	2015 (K)	pooled (K)	2014 (L)	2015 (L)	pooled (L)
λ												
Climate Change	0.333	0.439	0.357	0.334	0.416	0.362	0.338	0.378	0.345	0.341	0.372	0.344
Science	0.465 ***	0.288 **	0.381	0.46 ***	0.303 **	0.37 **	0.445 ***	0.318 **	0.357 ***	0.436 **	0.319 **	0.344 ***
	(0.32)	(0.248)	(0.553)	(0.235)	(0.268)	(0.342)	(0.326)	(0.283)	(0.257)	(0.391)	(0.276)	(0.234)
Renewable Energies/Technologies	0.202 ***	0.273 ***	0.262 ***	0.207 ***	0.28 ***	0.268 ***	0.218 **	0.303 ***	0.298 ***	0.223 **	0.309 **	0.312 ***
	(0.178)	(0.146)	(0.144)	(0.135)	(0.15)	(0.158)	(0.23)	(0.225)	(0.198)	(0.24)	(0.264)	(0.246)
β												
GRP per capita in ppp	0.374 ***	0.327 **	0.29 **	2.031 **	3.107 ***	2.656 ***	2.068	3.549 **	2.942 ***	2.297 **	3.398 ***	2.878 ***
	(0.095)	(0.117)	(0.089)	(0.68)	(0.89)	(0.779)	(1.066)	(1.181)	(0.777)	(1.118)	(0.876)	(0.586)
GRP per capita^2 in ppp				-4.673 **	-6.072 **	-5.277 **	-6.272 **	-8.371 **	-7.305 ***	-6.985 **	-8.767 ***	-7.678 ***
				(1.611)	(2.224)	(1.753)	(2.705)	(2.935)	(1.944)	(2.973)	(2.45)	(1.643)
GRP per capita^3 in ppp				3.147 **	3.358 **	3.036 **	4.705 **	5.448 **	4.945 ***	5.206 **	5.955 ***	5.348 ***
				(1.056)	(1.433)	(1.09)	(1.86)	(2.005)	(1.338)	(2.076)	(1.744)	(1.182)
Access to Internet	-0.269 **	-0.218		-0.239 **	-0.292 **	-0.309 **	-0.058	-0.061	-0.07	-0.068	-0.086	-0.101
	(0.13)	(0.184)		(0.101)	(0.142)	(0.101)	(0.079)	(0.107)	(0.071)	(0.08)	(0.1)	(0.064)
Mining							0.523 **	0.373	0.398 **	0.515 **	0.415	0.403 **
							(0.2)	(0.253)	(0.161)	(0.233)	(0.274)	(0.187)
Manufacturing							-0.104	-0.305 **	-0.259 **	-0.065	-0.135	-0.155 **
							(0.095)	(0.109)	(0.084)	(0.078)	(0.091)	(0.062)
Air pollution							-0.481 **	-0.739 **	-0.66 ***	-0.512	-0.801 ***	-0.707 ***
							(0.244)	(0.246)	(0.177)	(0.299)	(0.232)	(0.168)
Environmental Protection Costs							-0.006	-0.12	-0.127	-0.037	-0.158	-0.135
							(0.053)	(0.114)	(0.065)	(0.074)	(0.139)	(0.075)
Contaminated Water							-0.182 **	-0.17	-0.111	-0.212 **	-0.278	-0.114
							(0.072)	(0.111)	(0.066)	(0.089)	(0.222)	(0.098)
Population Density										0.006	0.146	-0.005
										(0.067)	(0.171)	(0.074)
People with Age +65										-0.086	-0.311	-0.257
										(0.172)	(0.204)	(0.137)
Labour Force with Tertiary Education										0.108	0.108	0.105 **
										(0.06)	(0.079)	(0.05)
Vehicle Rate										0.03	0.236	0.186
										(0.145)	(0.155)	(0.095)
Year Dummy (2015==1)			-0.141			-0.077			-0.127 **			-0.115 **
			(0.119)			(0.082)			(0.046)			(0.043)
Observation	81	81	162	81	81	162	75	75	150	75	75	150
Degrees of Freedom	4	4	4	8	8	10	18	18	20	26	26	28
AIC	570	606	1219	556	588	1172	485	522	1024	489	517	1011
CFI	0.874	0.965	0.8	0.8	0.963	0.793	0.7	0.879	0.742	0.636	0.824	0.708
RMSEA	0.239	0.102	0.254	0.235	0.081	0.187	0.231	0.116	0.177	0.222	0.124	0.167

Standard errors in parentheses; significance level: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

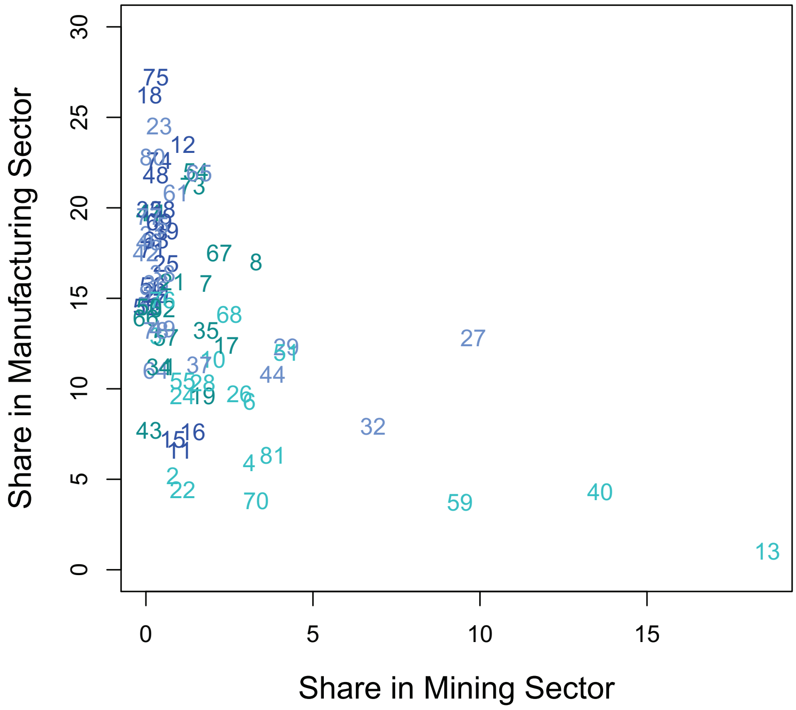

Figure 4.1: Scatter plot illustrating the relation between the share of employees in manufacturing and mining sector for the 81 regions in 2014. High EA ranked regions are light blue and low ranked regions dark blue. The numbers correspond to the regions in Table 6.

Table 4.4: Goodness-of-Fit Measures

Measure	Formula
Akaike Information Criterion	$AIC = \chi^2_{model} + g(g+1) - 2\,df_{model}$ $\chi^2_{model} \ldots$ χ^2-value of the full model $g \ldots$ number of variables in the full model $df_{model} \ldots$ degrees of freedom of the full model Source: Tanaka (1993)
Robust Root Mean Square Error of Approximation	$RMSEA = \sqrt{\max\left(0, \frac{\hat{c}(\chi^2_{SB,n} - df_{model})}{(n-1)df_{model}}\right)}$ $\chi^2_{SB,n} \ldots$ *Satorra-Bentler* χ^2-value of the full model $\hat{c} \ldots$ scaling constant Source: Brosseau-Liard, Savalei, and Li (2012)

Table 4.5: Region IDs, which are used in Chapter 2.

ID	Region	ID	Region	ID	Region	ID	Region
1	Adygea	21	Kaliningrad	41	Mari El	61	Samara
2	Altai (Republic)	22	Kalmykia	42	Mordovia	62	Saratov
3	Altai (Krai)	23	Kaluga	43	Moscow	63	Smolensk
4	Amur	24	Kamchatka	44	Murmansk	64	Stavropol
5	Arkhangelsk	25	Karachay-Cherkessia	45	Nenets AO	65	Sverdlovsk
6	Astrakhan	26	Karelia	46	Nizhny Novgorod	66	Tambovsk
7	Bashkortostan	27	Kemerovo	47	North Ossetia-Alania	67	Tatarstan
8	Belgorod	28	Khabarovsk	48	Novgorod	68	Tomsk
9	Bryansk	29	Khakassia	49	Novosibirsk	69	Tula
10	Buryatia	30	Khanty-Mansi AO	50	Omsk	70	Tuva
11	Chechnya	31	Kirov	51	Orenburg	71	Tver
12	Chelyabinsk	32	Komi	52	Oryol	72	Tyumen
13	Chukotka AO	33	Kostroma	53	Penza	73	Udmurtia
14	Chuvashia	34	Krasnodar	54	Perm	74	Ulyanovsk
15	Dagestan	35	Krasnoyarsk	55	Primorsky	75	Vladimir
16	Ingushetia	36	Kurgan	56	Pskov	76	Volgograd
17	Irkutsk	37	Kursk	57	Rostov	77	Vologda
18	Ivanovo	38	Leningrad + St.Petersburg	58	Ryazan	78	Voronezh
19	Jewish	39	Lipetsk	59	Sakha	79	Yamalo-Nenets AO
20	Kabardino-Balkaria	40	Magadan	60	Sakhalin	80	Yaroslavl
						81	Zabaykalsky

Table 4.6: Generated cluster solution by using k-means algorithm on the estimated regional EA indices from 01/14 to 04/16.

EA Index Cluster	Region
Cluster 1	Chechnya, Dagestan, Ingushetia, Kabardino-Balkaria, Karachay-Cherkessia, North Ossetia-Alania
Cluster 2	Adygea, Arkhangelsk, Bashkortostan, Belgorod, Bryansk, Chelyabinsk, Chuvashia, Ivanovo, Kaliningrad, Kaluga, Karelia, Kemerovo, Kirov, Kostroma, Kurgan, Kursk, Leningrad + St.Petersburg, Lipetsk, Mari El, Mordovia, Nizhny Novgorod, Novgorod, Oryol, Penza, Pskov, Rostov, Ryazan, Samara, Saratov, Smolensk, Stavropol, Sverdlovsk, Tambovsk, Tula, Tver, Ulyanovsk, Vladimir, Voronezh
Cluster 3	Altai (Krais), Amur, Astrakhan, Buryatia, Irkutsk, Jewish, Kalmykia, Khakassia, Khanty-Mansi AO, Komi, Krasnodar, Krasnoyarsk, Moscow, Murmansk, Novosibirsk, Omsk, Orenburg, Perm, Primorsky, Tatarstan, Tomsk, Tyumen, Udmurtia, Volgograd, Vologda, Yamalo-Nenets AO, Yaroslavl, Zabaykalsky
Cluster 4	Altai (Republic), Chukotka AO, Kamchatka, Khabarovsk, Magadan, Nenets AO, Sakha, Sakhalin, Tuva

Table 4.7: Russian regions grouped by the authors.

Geo Cluster	Region
North – Eastern Europe	Arkhangelsk, Bryansk, Chuvashia, Ivanovo, Kaliningrad, Kaluga, Karelia, Kirov, Komi, Kostroma, Kursk, Leningrad with St.Petersburg, Lipetsk, Mari El, Mordovia, Moscow, Murmansk, Nizhny Novgorod, Novgorod, Oryol, Penza, Perm, Pskov, Ryazan, Smolensk, Tambovsk, Tatarstan, Tula, Tver, Udmurtia, Ulyanovsk, Vladimir, Vologda, Yaroslavl
South-Eastern Europe	Adygea, Astrakhan, Bashkortostan, Belgorod, Chechnya, Dagestan, Ingushetia, Kabardino-Balkaria, Kalmykia, Karachay-Cherkessia, Krasnodar, North Ossetia-Alania, Orenburg, Rostov, Samara, Saratov, Stavropol, Volgograd, Voronezh,
Western and Central Siberia	Altai, Altai (Republic), Chelyabinsk, Irkutsk, Kemerovo, Khakassia, Khanty-Mansi AO, Krasnoyarsk, Kurgan, Nenets AO, Novosibirsk, Omsk, Sverdlovsk, Tomsk, Tuva, Tyumen, Yamalo-Nenets AO,
(Far) East (ern) Siberia	Amur, Buryatia, Chukotka AO, Jewish, Kamchatka, Khabarovsk, Magadan, Primorsky, Sakha, Sakhalin, Zabaykalsky,

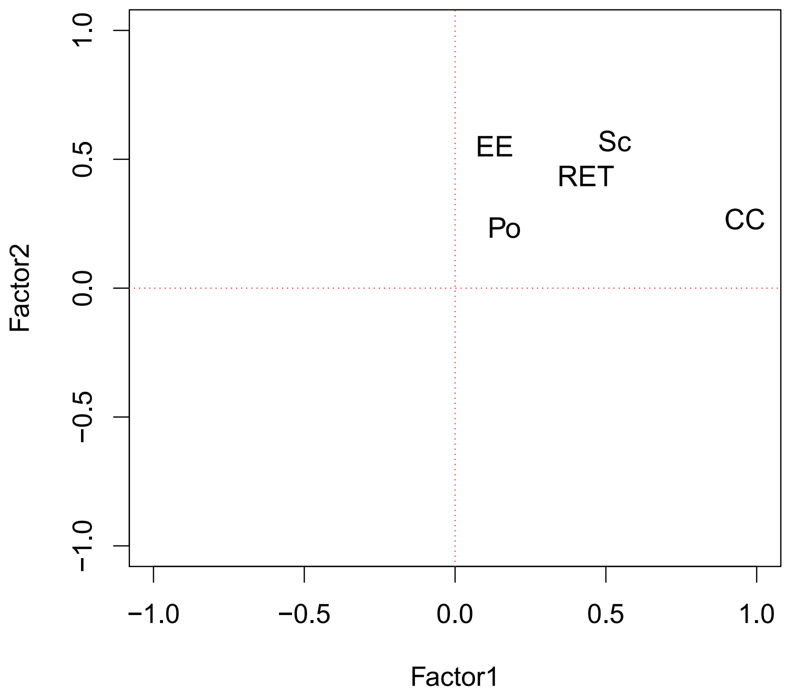

Figure 4.2: Loadings of the indicator variables after factor rotation by performing a factor analysis. Notation: Climate Change (CC), Endangered Environment (EE), Political (Po), Science (Sc) and Renewable Energies and Technologies (RET).

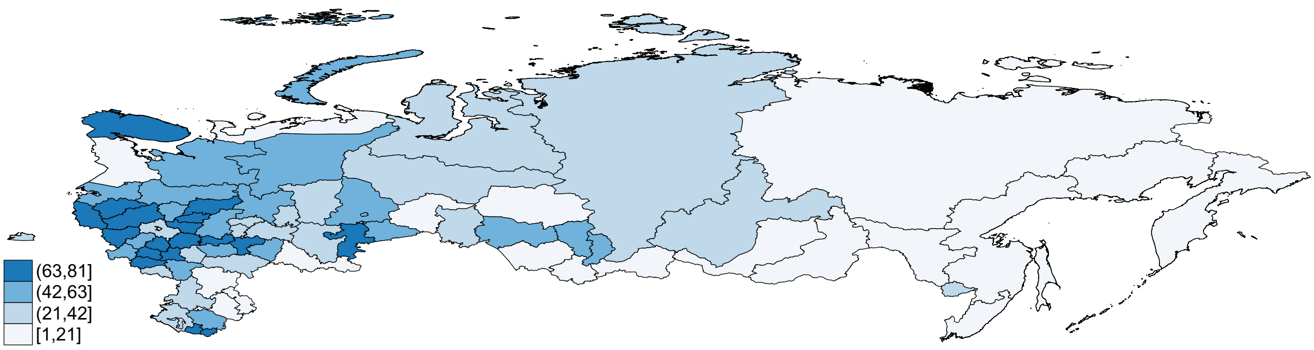

Figure 4.3: Map of 80 Russian regions illustrating the Environmental Awareness Index for 2014 (without Autonomous Okrug Chukotka). High ranked regions are light blue and low ranked regions dark blue.

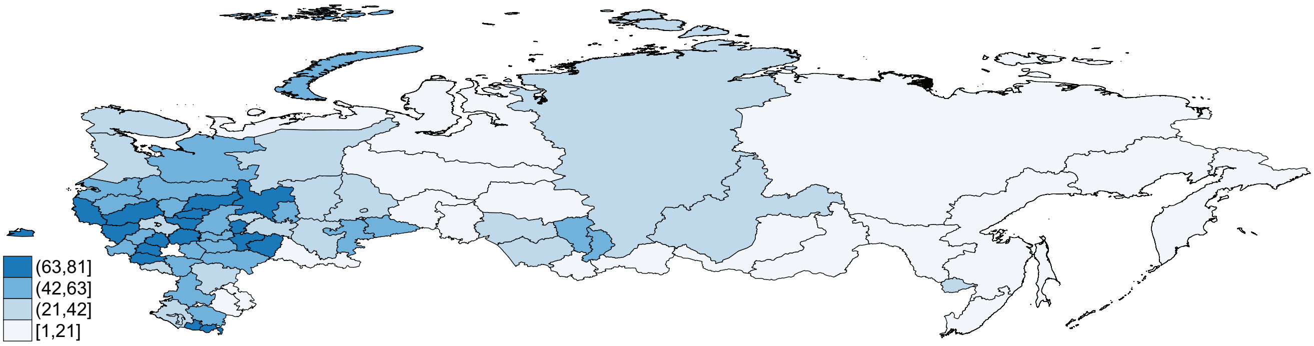

Figure 4.4: Map of 80 Russian regions illustrating the Environmental Awareness Index for 2015 (without Autonomous Okrug Chukotka). High ranked regions are light blue and low ranked regions dark blue.

	GRP per capita	Manufact-uring	Mining	Fishery	Air pollution	Carbon dioxide	Nitrogen dioxide
GRP per capita	1	-0.414	0.784	-0.393	0.179	0.133	0.109
Manufacturing	-0.414	1	-0.573	-0.051	-0.033	-0.145	0.081
Mining	0.784	-0.573	1	-0.415	0.321	0.286	0.175
Fishery	-0.393	-0.051	-0.415	1	-0.346	-0.337	-0.416
Air pollution	0.179	-0.033	0.321	-0.346	1	0.737	0.789
Carbon dioxide	0.133	-0.145	0.286	-0.337	0.737	1	0.654
Nitrogen dioxide	0.109	0.081	0.175	-0.416	0.789	0.654	1
Unemployment rate	-0.195	-0.337	-0.098	0.277	-0.129	-0.095	-0.169
Elderly people	-0.373	0.713	-0.564	0.016	-0.125	-0.098	0.005
Young people	0.160	-0.581	0.329	0.140	0.029	0.042	-0.074
Education	-0.124	-0.393	0.031	0.248	-0.117	-0.091	-0.201
Population density	0.101	-0.075	-0.087	-0.234	-0.057	-0.059	0.100
Internet access	0.196	0.173	0.166	-0.459	0.269	0.256	0.314
Vehicle rate	0.008	0.173	-0.127	-0.172	0.168	0.139	0.213

	Unemployment rate	Elderly people	Young people	Education	Population density	Internet access	Vehicle rate
GRP per capita	-0.195	-0.373	0.16	-0.124	0.101	0.196	0.008
Manufacturing	-0.337	0.713	-0.581	-0.393	-0.075	0.173	0.173
Mining	-0.098	-0.564	0.329	0.031	-0.087	0.166	-0.127
Fishery	0.277	0.016	0.14	0.248	-0.234	-0.459	-0.172
Air pollution	-0.129	-0.125	0.029	-0.117	-0.057	0.269	0.168
Carbon dioxide	-0.095	-0.098	0.042	-0.091	-0.059	0.256	0.139
Nitrogen dioxide	-0.169	0.005	-0.074	-0.201	0.100	0.314	0.213
Unemployment rate	1	-0.592	0.778	0.819	-0.151	-0.712	-0.441
Elderly people	-0.592	1	-0.896	-0.63	0.089	0.343	0.455
Young people	0.778	-0.896	1	0.835	-0.171	-0.538	-0.542
Education	0.819	-0.63	0.835	1	-0.168	-0.662	-0.523
Population density	-0.151	0.089	-0.171	-0.168	1	0.208	0.037
Internet access	-0.712	0.343	-0.538	-0.662	0.208	1	0.422
Vehicle rate	-0.441	0.455	-0.542	-0.523	0.037	0.422	1

Table 4.8: Pearson's correlation coefficients of the causes variables in 2014.

Appendix for Chapter 3

	Ada LASSO	Ada 25 sel	Ada 50 sel	Ada 75 sel
ue without interaction	-0.195 # (0.019)	-0.205 # (0.039)	-0.294 # (0.021)	-0.032 + (0.013)
Δue without interaction	-0.151 # (0.015)	-0.215 # (0.025)	-0.158 # (0.012)	-0.016 # (0.006)
Interaction between ue and Occupation				
metal and plant engineering	-0.270 # (0.056)		0.053 (0.051)	-0.299 # (0.063)
industrial and tool mechanic	-0.342 # (0.054)			-0.242 # (0.093)
car, aircraft constr., maintenance	-0.620 # (0.101)	-1.030 # (0.156)		
electrical and electronic engineering	-0.669 # (0.074)	-0.861 # (0.099)		
building professions	-0.051 + (0.025)	0.186 # (0.056)		
labourers	-0.024 (0.034)	-0.242 # (0.062)		-0.033 * (0.019)
technicians	-0.777 # (0.136)			-0.495 # (0.151)
salesman	0.598 # (0.049)		0.623 # (0.062)	0.266 # (0.050)
IT- core occupations	-0.637 + (0.249)			-0.895 # (0.263)
commercial occupation	-0.162 # (0.059)	0.736 # (0.107)	0.093 (0.073)	-0.695 # (0.061)
health care occ. w/o license	2.151 # (0.283)	2.066 # (0.392)		-0.736 # (0.161)
bakers and confectioners		-0.625 + (0.294)		
butcher		-1.106 # (0.261)		0.413 # (0.141)
cook		-0.731 # (0.144)		0.197 # (0.047)
engineers		1.185 # (0.357)		-0.895 # (0.334)
bank employees & insurance salesman		3.858 # (0.293)	1.938 # (0.243)	-1.564 # (0.387)
management		1.596 # (0.335)		-1.337 # (0.255)
finance and accounting		0.962 + (0.385)	0.085 (0.193)	
security occupations		2.149 # (0.457)		
social occupations		1.222 # (0.258)	-0.046 (0.110)	-0.448 # (0.089)
teacher		3.729 # (0.779)		
occupations in hotel & catering		-0.213 * (0.109)		0.167 # (0.047)
occupations in room clean., waste coll.		-0.369 # (0.117)		
silkworm moth, textile finishers			0.183 (0.229)	
textile and leather proc.			0.223 + (0.111)	
product tester			-0.038 (0.082)	
chemists			0.923 + (0.381)	
technical drawer			0.382 # (0.128)	
surveyor			0.825 + (0.358)	
technical special forces			0.388 * (0.229)	-0.500 # (0.193)
other salesmen			-0.303 + (0.149)	-0.756 # (0.155)
advertising experts			0.688 # (0.182)	-0.698 + (0.342)
transport-related vocations			-0.308 # (0.097)	
administration			0.254 (0.429)	-1.192 # (0.411)
temporary office stuff			0.021 (0.090)	-0.384 # (0.067)
janitor			0.237 (0.404)	
legal professions			0.714 (0.722)	
artists			0.080 (0.189)	-0.483 # (0.155)
chemical and plastics workers				-0.834 # (0.215)
paper processing				0.282 + (0.137)
metalworking				-0.436 # (0.147)
precision mechanic				0.475 # (0.096)
beverages, drink & tobacco proc.				-0.598 # (0.195)
retailers and distributors				-0.424 # (0.120)
packer and warehouse worker				-0.068 * (0.037)
health care occ. with license				-0.492 (0.538)
publicist, scientific occ.				-1.208 # (0.218)

	Ada LASSO	Ada 25 sel	Ada 50 sel	Ada 75 sel
Interaction between Δue and Occupation				
precision mechanic	-0.295 # (0.099)			
bakers and confectioners	-0.149 + (0.066)	-0.699 # (0.108)		
building professions	0.055 + (0.022)	0.082 + (0.033)		
labourers	-0.044 * (0.025)	-0.270 # (0.052)		
technical drawer	-0.322 + (0.140)			0.145 + (0.071)
salesman	-0.144 # (0.044)	-0.334 # (0.106)	-0.025 (0.036)	
retailers and distributors	-0.204 # (0.064)			-0.399 # (0.070)
bank employees and insurance salesman	-0.670 * (0.365)		-0.222 (0.168)	-1.122 # (0.406)
other salesmen	-0.424 # (0.132)			-0.421 # (0.079)
packer and warehouse worker	-0.067 + (0.033)	-0.334 # (0.073)		-0.052 (0.032)
finance and accounting system	-0.471 + (0.195)			-0.610 # (0.177)
commercial occupation	-0.303 # (0.053)	-0.276 # (0.060)	-0.025 (0.022)	-0.176 # (0.031)
health care occ. w/o license	-0.602 # (0.119)	-0.503 # (0.094)		
social occupations	-0.784 # (0.153)	-0.546 # (0.205)		-0.261 # (0.045)
occupations in body care	-0.271 # (0.076)	-0.586 # (0.133)		0.225 # (0.035)
occupations in hotel and catering	-0.134 * (0.074)	-0.399 # (0.124)		
occupations in room cleaning, waste coll.	-0.216 # (0.068)	-0.389 # (0.088)		
mineworkers		-0.407 + (0.202)		
metalworking		0.171 + (0.076)		-0.090 * (0.050)
butcher		-0.499 # (0.137)		
cook		-0.379 # (0.073)		
personal security occupations		-0.541 # (0.184)	-0.075 (0.090)	
beverages, drink and tobacco proc.			-0.068 (0.188)	-0.294 # (0.089)
technicians			-0.006 (0.121)	
technical special forces			0.210 (0.132)	
management			0.197 (0.156)	-0.541 # (0.156)
IT- core occupations			-0.251 + (0.124)	-0.542 # (0.125)
temporary office staff			0.060 (0.060)	
janitor			0.014 (0.085)	
security occupations			0.005 (0.131)	
artists			0.094 + (0.038)	
teacher			0.026 (0.063)	-0.136 (0.088)
metal and plant engineering				-0.119 # (0.022)
electrical and electronic engin.				0.081 + (0.034)
surveyor				0.271 # (0.097)
aviation and shipping				-0.858 # (0.203)
health care occ. with license				0.604 # (0.201)
publicist, more scientific occu.				-0.384 # (0.128)
Constant	3.328 # (0.074)	2.702 # (0.065)	3.888 # (0.076)	3.963 # (0.045)
adj. R^2	0.592	0.440	0.454	0.387
AIC	3.90E+05	6.90E+05	3.80E+05	4.00E+05
BIC	3.90E+05	6.90E+05	3.80E+05	4.00E+05

Table 4.9: Post-OLS estimation of the ue and Δue and their interaction coefficients by chosen different models through adaptive LASSO from Zou for the mean (2nd column), the 0.10 quantile (3rd column), the lower quantile (4th column), the median (5th column), upper quantile (6th column). Heteroscedasticity robust clustered standard errors of the coefficients are in parentheses; significance level $^{\#}p < 0.01$, $^{+}p < 0.05$, $*p < 0.1$.

	Full Model		Full q0.25		Full q0.5		Full q0.75	
ue without interaction	-0.078 +	(0.034)	-0.205 +	(0.086)	-0.119 #	(0.042)	0.036	(0.024)
Δue without interaction	-0.059 +	(0.024)	-0.207 +	(0.086)	-0.062 +	(0.029)	-0.005	(0.014)
Interaction between *ue* and and Occupation								
occupations in agricul., animal breeding	*reference*		*reference*		*reference*		*reference*	
mineworkers	-0.135	(0.156)	-0.306	(0.398)	-0.315	(0.258)	-0.012	(0.216)
stone workers, ceramics, glass workers	-0.284	(0.191)	0.172	(0.387)	0.264	(0.202)	-0.036	(0.221)
chemical and plastics workers	-0.421 #	(0.160)	0.073	(0.235)	-0.036	(0.164)	-0.880 #	(0.220)
paper processing	-0.353 +	(0.164)	-0.651 +	(0.254)	0.132	(0.172)	0.459 #	(0.162)
metalworking	-0.263 #	(0.091)	0.119	(0.146)	-0.234 +	(0.118)	-0.493 #	(0.142)
metal and plant engineering	-0.378 #	(0.064)	-0.108	(0.121)	-0.415 #	(0.083)	-0.288 #	(0.065)
industrial and tool mechanic	-0.400 #	(0.068)	-0.326 #	(0.124)	-0.215 +	(0.100)	-0.241 +	(0.103)
car, aircraft construction, maintenance	-0.682 #	(0.103)	-1.037 #	(0.194)	0.140	(0.106)	0.172 *	(0.094)
precision mechanic	0.029	(0.156)	-0.194	(0.263)	0.577 #	(0.134)	0.490 #	(0.107)
electrical and electronic engineering	-0.612 #	(0.084)	-0.775 #	(0.149)	0.092	(0.101)	-0.092	(0.092)
silkworm moth, textile finishers process	0.106	(0.195)	0.122	(0.376)	0.009	(0.232)	0.156	(0.183)
textile and leather processing	-0.171	(0.211)	-0.206	(0.238)	0.114	(0.121)	0.138	(0.102)
bakers and confectioners	-0.071	(0.159)	-0.665 +	(0.303)	0.124	(0.164)	0.077	(0.123)
butcher	-0.175	(0.149)	-1.054 #	(0.289)	-0.195	(0.172)	0.436 #	(0.158)
cook	-0.130 *	(0.073)	-0.787 #	(0.165)	0.252 #	(0.078)	0.176 #	(0.055)
beverages, drink and tobacco processing	-0.307 *	(0.183)	-0.139	(0.257)	-0.168	(0.170)	-0.525 #	(0.200)
building professions	-0.106 #	(0.040)	0.150	(0.097)	-0.256 #	(0.055)	-0.044	(0.030)
product tester	-0.085	(0.095)	0.094	(0.187)	-0.242 +	(0.106)	-0.152	(0.114)
labourers	-0.135 #	(0.042)	-0.267 #	(0.099)	-0.021	(0.049)	-0.074 +	(0.033)
engineers	-0.808 +	(0.329)	1.383 #	(0.394)	0.784 +	(0.341)	-1.115 #	(0.410)
chemists	-0.761	(1.111)	-0.727	(1.212)	1.475 *	(0.790)	-0.050	(0.880)
technicians	-0.903 #	(0.216)	-0.265	(0.321)	-0.040	(0.230)	-0.455 +	(0.188)
technical drawer	-0.504 #	(0.194)	-0.192	(0.256)	0.322 +	(0.144)	0.108	(0.128)
surveyor	-0.367	(0.268)	-1.011 *	(0.594)	0.866 +	(0.376)	0.496 *	(0.263)
technical special forces	0.001	(0.301)	0.698 +	(0.289)	0.322	(0.248)	-0.697 +	(0.283)
salesman	0.503 #	(0.066)	0.155	(0.132)	0.400 #	(0.079)	0.261 #	(0.058)
retailers and distributors	0.081	(0.113)	0.400 +	(0.192)	0.106	(0.124)	-0.394 #	(0.140)
bank employees and insurance salesman	0.124	(0.299)	2.979 #	(0.393)	1.343 #	(0.278)	-1.366 #	(0.391)
other salesmen	0.111	(0.161)	0.651 +	(0.302)	-0.185	(0.174)	-0.687 #	(0.182)
advertising experts	-0.252	(0.264)	0.551	(0.380)	-0.090	(0.281)	-0.609 *	(0.342)
transport-related vocations	-0.426 #	(0.150)	-0.117	(0.394)	-0.587 #	(0.197)	-0.008	(0.138)
aviation and shipping	-0.654 *	(0.394)	0.636	(0.462)	0.252	(0.248)	-0.721 *	(0.415)
packer and warehouse worker	-0.196 #	(0.073)	-0.119	(0.169)	-0.063	(0.072)	-0.173 +	(0.068)
management	-0.458 *	(0.265)	1.828 #	(0.342)	0.822 +	(0.349)	-1.258 #	(0.269)
administration	-0.110	(0.356)	-0.146	(0.707)	0.207	(0.421)	-1.013 #	(0.383)
finance and accounting system	0.122	(0.245)	1.072 #	(0.394)	0.503	(0.319)	0.026	(0.301)
IT- core occupations	-1.034 #	(0.355)	0.300	(0.233)	0.100	(0.199)	-0.853 #	(0.254)
business office occupation	-0.238 #	(0.072)	0.644 #	(0.155)	-0.042	(0.089)	-0.685 #	(0.093)
temporary office stuff	-0.143	(0.138)	-0.052	(0.210)	-0.099	(0.112)	-0.304 #	(0.098)
personal security occupations	-0.171	(0.176)	-0.545	(0.375)	-0.235	(0.180)	0.109	(0.157)
janitor	-0.265	(0.368)	0.961	(0.620)	0.150	(0.423)	-0.250	(0.314)
security occupations	0.501	(0.345)	2.037 #	(0.457)	0.637	(0.511)	0.404	(0.554)
legal professions	-0.201	(1.034)	0.925	(1.228)	0.637	(0.751)	0.324	(0.950)
artists	-0.129	(0.200)	0.147	(0.438)	-0.127	(0.189)	-0.491 #	(0.149)
designer	-0.407 #	(0.192)	-0.068	(0.361)	-0.094	(0.204)	-0.228	(0.160)
health care occ. with license	-0.669	(0.865)	-1.400	(1.534)	-0.957	(0.858)	-1.112 *	(0.671)
health care occ. without license	2.224 #	(0.289)	1.941 #	(0.437)	1.148 #	(0.252)	-0.480 +	(0.190)
social occupations	0.185	(0.135)	1.174 #	(0.263)	-0.267 #	(0.103)	-0.423 #	(0.091)
teacher	1.336 #	(0.512)	4.016 #	(0.824)	1.893 #	(0.690)	0.048	(0.517)
publicist, more scientific occupation	-0.281	(0.293)	0.565	(0.396)	-0.411	(0.253)	-1.122 #	(0.226)
occupations in body care	0.877 +	(0.413)	-0.342	(0.626)	0.864 #	(0.255)	0.241	(0.199)
occupations in hotel and catering	0.069	(0.073)	-0.291 +	(0.146)	0.296 #	(0.081)	0.142 +	(0.068)
occupations in room cleaning, waste col.	-0.124	(0.081)	-0.374 +	(0.160)	-0.107	(0.069)	0.020	(0.054)

	Full Model	Full q0.25	Full q0.5	Full q0.75
Interaction between Δue and occupation				
occupations in agricul., animal breeding	*reference*	*reference*	*reference*	*reference*
mineworkers	-0.134* (0.081)	-0.512* (0.289)	-0.330+ (0.143)	-0.056 (0.098)
stone workers, ceramics, glass workers	-0.321+ (0.143)	-0.453* (0.232)	-0.138 (0.131)	-0.203+ (0.103)
chemical and plastics workers	-0.185+ (0.080)	0.082 (0.131)	-0.167+ (0.071)	-0.080 (0.066)
paper processing	-0.318# (0.101)	-0.393* (0.202)	-0.192+ (0.098)	0.046 (0.101)
metalworking	-0.002 (0.060)	0.165 (0.114)	-0.003 (0.079)	-0.105* (0.055)
metal and plant engineering	-0.104# (0.031)	-0.054 (0.097)	-0.158# (0.041)	-0.124# (0.026)
industrial and tool mechanic	0.006 (0.037)	0.071 (0.096)	-0.028 (0.048)	-0.038 (0.041)
car, aircraft construction, maintenance	-0.045 (0.042)	0.093 (0.103)	0.009 (0.038)	0.000 (0.023)
precision mechanic	-0.384# (0.109)	-0.029 (0.129)	0.082 (0.067)	0.056 (0.054)
electrical and electronic engineering	-0.184# (0.048)	-0.030 (0.095)	0.014 (0.052)	0.062* (0.038)
silkworm moth, textile finishers process	-0.063 (0.078)	-0.323 (0.198)	-0.117 (0.093)	-0.031 (0.082)
textile and leather processing	-0.152 (0.209)	0.028 (0.158)	0.021 (0.091)	0.102 (0.090)
bakers and confectioners	-0.240# (0.071)	-0.733# (0.140)	-0.077 (0.062)	0.065 (0.047)
butcher	-0.149+ (0.064)	-0.531# (0.167)	-0.183+ (0.072)	-0.083* (0.049)
cook	-0.164# (0.048)	-0.353# (0.116)	-0.070 (0.050)	0.031 (0.032)
beverages, drink and tobacco processing	-0.110 (0.228)	-0.140 (0.329)	-0.185 (0.196)	-0.285# (0.090)
building professions	-0.025 (0.029)	0.069 (0.089)	-0.083+ (0.038)	-0.012 (0.018)
product tester	-0.145+ (0.066)	-0.167 (0.141)	-0.150# (0.057)	-0.067 (0.057)
labourers	-0.136# (0.032)	-0.272# (0.098)	-0.053 (0.037)	0.002 (0.026)
engineers	-0.745# (0.146)	0.170 (0.201)	-0.026 (0.146)	-0.304* (0.182)
chemists	-0.348 (0.725)	-0.954 (1.151)	0.426 (0.790)	-0.640 (0.731)
technicians	-0.540+ (0.209)	-0.190 (0.225)	-0.119 (0.131)	-0.050 (0.100)
technical drawer	-0.565# (0.174)	0.312+ (0.156)	0.208+ (0.095)	0.230# (0.083)
surveyor	-0.216 (0.195)	-0.669+ (0.288)	0.349+ (0.149)	0.328# (0.107)
technical special forces	-0.462* (0.258)	0.511# (0.188)	0.017 (0.148)	-0.201 (0.149)
salesman	-0.205# (0.044)	-0.316+ (0.139)	-0.057 (0.039)	-0.023 (0.024)
retailers and distributors	-0.337# (0.077)	-0.175 (0.145)	-0.397# (0.075)	-0.406# (0.075)
bank employees and insurance salesman	-0.734+ (0.364)	0.063 (0.157)	-0.295* (0.173)	-1.143# (0.413)
other salesmen	-0.448# (0.139)	0.029 (0.169)	-0.456# (0.096)	-0.441# (0.082)
advertising experts	-0.017 (0.134)	0.376 (0.302)	-0.010 (0.130)	0.152 (0.161)
transport-related vocations	-0.075 (0.069)	-0.049 (0.164)	-0.071 (0.095)	-0.030 (0.035)
aviation and shipping	-0.800# (0.217)	0.137 (0.402)	-0.217 (0.172)	-1.160# (0.239)
packer and warehouse worker	-0.172# (0.042)	-0.360# (0.113)	-0.156# (0.053)	-0.079+ (0.035)
management	-0.412# (0.154)	0.347 (0.360)	-0.365# (0.127)	-0.562# (0.162)
administration	-0.594* (0.355)	-0.563+ (0.284)	0.093 (0.143)	0.107 (0.141)
finance and accounting system	-0.864# (0.266)	-0.612* (0.319)	-0.418+ (0.180)	-0.661# (0.199)
IT- core occupations	-0.758# (0.224)	-0.146 (0.185)	-0.313+ (0.157)	-0.535# (0.126)
business office occupation	-0.387# (0.056)	-0.283# (0.103)	-0.165# (0.039)	-0.191# (0.034)
temporary office stuff	-0.140 (0.150)	-0.091 (0.157)	-0.019 (0.062)	-0.011 (0.036)
personal security occupations	-0.280+ (0.110)	-0.782# (0.294)	-0.151 (0.119)	0.017 (0.063)
janitor	0.015 (0.055)	0.173 (0.127)	-0.072 (0.084)	-0.026 (0.026)
security occupations	-0.172* (0.099)	0.155 (0.184)	-0.088 (0.140)	-0.136 (0.100)
legal professions	-0.532 (0.471)	-0.020 (0.528)	-0.156 (0.142)	-0.348* (0.191)
artists	-0.108+ (0.046)	0.027 (0.125)	-0.027 (0.056)	0.005 (0.048)
designer	-0.285# (0.047)	-0.001 (0.154)	-0.003 (0.080)	-0.015 (0.049)
health care occ. with license to practice	0.212 (0.626)	0.704 (0.704)	0.279 (0.320)	0.457* (0.272)
health care occ. w/o license to practice	-0.678# (0.119)	-0.501# (0.127)	0.066 (0.042)	0.000 (0.030)
social occupations	-0.847# (0.148)	-0.577# (0.214)	-0.434# (0.089)	-0.273# (0.049)
teacher	-0.107 (0.128)	-0.102 (0.158)	-0.119* (0.070)	-0.170* (0.102)
publicist, more scientific occupation	-0.279 (0.241)	0.024 (0.293)	-0.333+ (0.157)	-0.375+ (0.158)
occupations in body care	-0.373# (0.079)	-0.503# (0.167)	0.157+ (0.066)	0.174# (0.044)
occupations in hotel and catering	-0.213# (0.075)	-0.406# (0.149)	-0.074 (0.049)	-0.001 (0.030)
occupations in room cleaning, waste col.	-0.269# (0.077)	-0.406# (0.124)	-0.107+ (0.051)	0.044 (0.034)
Constant	4.039# (0.290)	4.783# (0.627)	4.471# (0.341)	3.756# (0.233)
adj. R^2	0.593	0.441	0.458	0.389
AIC	3.90E+05	6.90E+05	3.80E+05	4.00E+05
BIC	3.90E+05	6.90E+05	3.80E+05	4.10E+05

Table 4.10: OLS estimation of the ue and Δue and their interaction coefficients by chosen the full model (628 RHS variables) for the 0.10 quantile (2nd column), the lower quantile (3rd column), the median (4th column), upper quantile (5th column). Heteroscedasticity robust clustered standard errors of the coefficients are in parentheses; significance level $\#p < 0.01$, $^{+}p < 0.05$, $*p < 0.1$.

Level	RHS Variables	Range	Sources
Individual	age	binary; classification: 15 to <18 years, 18 to <21 years, 21 to <24 years, 24 to <27 years and 27 to 30 years	IEB
	gender	binary; male and female	IEB
	mother	binary; whether woman interrupts her job for child care within the ten years observation time	IEB
	foreigner	binary; individual did not have a German passport	IEB
	school-leaving qualification	binary; no degree, lower secondary education (Hauptschule), secondary education (Mittlere Reife), higher education (Abitur)	IEB
	additional qualification	binary; Master craftman or university degree; interacted with school-leaving qualification	IEB
	duration of the vocational training	binary; 3-4 years, 4 and more years	IEB
	duration between the finishing of the occupational training and the start of the first job	binary; $0 >$ duration > 3 month, 3 month $\leq$ duration < 1 year; duration ≥ 1 year	IEB
	STEM and non-STEM	binary; individual works in a (non-) STEM occupation at the first job, change from STEM to non-STEM and vice versa	IEB
	log (wage)	metric; log (average net-wage per day) at the first job	IEB
	full and part time	binary; individual works in full (part) time at the first job and 10 years later, changes from full to part time and vice versa	IEB
Mobility (individual level)	switching from STEM to non-STEM and vice versa	binary; whether individual switch from a STEM to a non-STEM occupation and vice versa, classification from the German Federal Employment Agency (KldB 88-3 digit)	IEB
	upgrade/downgrade	two discrete, ordinate variables; occupations are ranked in respect to their wages at the first job and 10 years after	IEB
	change the employer	binary; changing the employer after the vocational training for the first job, changing the employer between the first job and ten years later	IEB
	enforced firm change due to firm closure	binary; whether individual becomes unemployed due to firm closure of the first employer	IEB
	regional migration	binary; whether individual moves to another labour market within the ten years; additional interacted with STEM dummy	IEB
	(log) distance of regional migration	metric; (log) distance of the labour market migration between the first job and ten years after in km (interacted with regional migration dummy)	IEB
	migration from east to west	binary, migration from former East Germany to West Germany	IEB
	unemployment periods ¿ 3 month	discrete, number of unemployment periods with > 3 month duration within the ten years	IEB
	additional firms	binary, additional number of firms (from 1, 2, 3, 4 and 5+), where individual worked within the ten years	IEB
	duration in employment	metric, discrete; total number of days in employment (independent from employer) of individual i within the 10 years	IEB
Firm level	firm age	binary; whether firm is of up to <5 years old	IAB Empl. Stat.
	log (firm size)	metric; log (number of employees in firm) at the first job	IAB Empl. Stat.
	share of women in firm	metric; at the first job	IAB Empl. Stat.
	share of foreigner in firm	metric; additional for robustness check; not in baseline model	IAB Empl. Stat.
	share of human capital	metric; proportion of experts and specialists on all employees	IAB Empl. Stat.
	average firm wage	metric; additional for robustness check	IAB Empl. Stat.
Occupational and regional level	labour market size	metric; log(number of employees in the occupational field and region from individual i) and interacted with an occupation dummy; for both times t and the growth between t and $t+10$ of individual i	IAB Empl. Stat.
	academic competitors	metric; number of employees with university degree related to the number of employees without university degree in the occupational field and labour market region from individual i and interacted with an occupation dummy; for both times t and the growth between t and $t+10$ of individual i	IAB Empl. Stat.
	ue	metric, the number of unemployed people related to the number of employed people in the occupation field and labour market region from individual and interacted with an occupation dummy; for both t and the growth $t+10$	IAB Empl. Stat.
	Labour market dummies	Binary variables	
Industry level	Industry -Dummies	Binary variables	
Year level	Time- Dummies	Binary variables	

Changing or rather growth rates between the first job and ten years after for all time-variant variables as additional variables.

Table 4.11: Description of the RHS variables and data source. *Empl. Stat.* means Employment Statistics.

Key	Key category	abbrev.	occupation (engl.)
AD	Artist, designer	Ar	artists
AD	Artist, designer	De	designer
AM	occu. in agriculture and animal breeding; mining	AC	occupations in agriculture and animal breeding
AM	occu. in agriculture and animal breeding; mining	Mi	mineworkers
BC	occupations in body care	BC	occupations in body care
BU	building professions	Bui	building professions
EE	electrical and electronic engineering	EE	electrical and electronic engineering
ETS	engineers, techn. and natur. Science	En	engineers
ETS	engineers, techn. and natur. Science	Ch	chemists
ETS	engineers, techn. and natur. Science	Te	technicians
ETS	engineers, techn. and natur. Science	TD	technical drawer
ETS	engineers, techn. and natur. Science	Su	surveyor
ETS	engineers, techn. and natur. Science	TS	technical special forces
Fi	finance	Ad	administration
Fi	finance	FA	finance and accounting system
FD	food professions	Ba	bakers and confectioners
FD	food professions	Bu	butcher
FD	food professions	Co	cook
FD	food professions	BD	beverages, drink and tobacco processing
HC	health care	He	health care occ. with license to practice medicine
HC	health care	HWo	health care occ. without license to practice medicine
IT	IT and workers	IT	IT- core occupations
IT	IT and workers	BO	business office occupation
IT	IT and workers	TO	temporary office stuff
LA	labourer	PT	product tester
LA	labourer	La	labourers
LP	legal professions	LP	legal professions
Ma	management	Ma	management
MI	manufacturing industry	SW	stone workers, ceramics, glass workers
MI	manufacturing industry	CW	chemical and plastics workers
MI	manufacturing industry	PP	paper processing
MI	manufacturing industry	MW	metalworking
MI	manufacturing industry	MP	metal and plant engineering
MI	manufacturing industry	IM	industrial and tool mechanic
MI	manufacturing industry	CA	car, aircraft construction, maintenance specialists
MI	manufacturing industry	PM	precision mechanic
ME	merchant	Sa	salesman
ME	merchant	RD	retailers and distributors
ME	merchant	FI	bank employees and insurance salesman
ME	merchant	OS	other salesmen
ME	merchant	AE	advertising experts
Ho	occupations in hotel and catering	HC	occupations in hotel and catering
SJ	security and janitor	PS	personal security occupations
SJ	security and janitor	Ja	janitor
SJ	security and janitor	SO	security occupations
ST	Social and teaching	Soc	social occupations
ST	Social and teaching	Te	teacher
ST	Social and teaching	Pu	publicist, more scientific occupation
TX	textil professions	ST	silkworm moth, textile finishers process
TX	textil professions	TE	textile and leather processing
TA	transportation	Tr	transport-related vocations
TA	transportation	AS	aviation and shipping
TA	transportation	PW	packer and warehouse worker
Wa	occupations in room cleaning, waste collection	Wa	occupations in room cleaning, waste collection

Table 4.12: Keys for the 22 occupational categories and abbreviations for the 54 observed professions.

Occupational field	men	women	Occupational field	men	women
occupations in agriculture and animal br.	0.733	0.267	retailers and distributors	0.498	0.502
mineworkers	0.901	0.099	bank empl./insurance salesman	0.458	0.542
stone workers, ceramics, glass workers	0.792	0.208	other salesmen	0.437	0.563
chemical and plastics workers	0.889	0.111	advertising experts	0.360	0.640
paper processing	0.766	0.234	transport-related vocations	0.858	0.142
metalworking	0.966	0.034	aviation and shipping	0.707	0.293
metal and plant engineering	0.950	0.050	packer and warehouse worker	0.778	0.222
industrial and tool mechanic	0.982	0.018	management	0.568	0.432
car, aircraft construction, maintenance	0.980	0.020	administration	0.399	0.601
precision mechanic	0.509	0.491	finance and accounting system	0.356	0.644
electrical and electronic engineering	0.973	0.027	IT- core occupations	0.850	0.150
silkworm moth, textile finishers process	0.783	0.217	commercial occupation	0.298	0.702
textile and leather processing	0.350	0.650	temporary office stuff	0.343	0.657
bakers and confectioners	0.690	0.310	personal security occupations	0.710	0.290
butcher	0.917	0.083	janitor	0.891	0.109
cook	0.639	0.361	security occupations	0.929	0.071
beverages, drink and tobacco processing	0.807	0.193	legal professions	0.491	0.509
building professions	0.963	0.037	artists	0.574	0.426
product tester	0.668	0.332	designer	0.379	0.621
labourers	0.803	0.197	health care occ. with license	0.356	0.644
engineers	0.868	0.132	health care occ. without license	0.109	0.891
chemists	0.638	0.362	social occupations	0.167	0.833
technicians	0.844	0.156	teacher	0.482	0.518
technical drawer	0.461	0.539	publicist, more scientific occupation	0.449	0.551
surveyor	0.668	0.332	occupations in body care	0.046	0.954
technical special forces	0.539	0.461	occupations in hotel and catering	0.246	0.754
salesman	0.274	0.726	occupations in room cleaning, waste	0.480	0.520
			Total	0.551	0.449

Table 4.13: Proportion of men and women in each of the 54 occupational fields.