Similar jobs, same pay. Or?

Exploring what skills, gender structures and organizational forms do to wage differentials

Petrus Garefelt

A Thesis Submitted to the Dept. of Sociology and Work Science

Supervised by Martin Henning and Examined by Jing Wu

Abstract

I examine the labor market's demand for knowledge and skills and compile a database archive of Swedish job advertisements with official register data for occupational structures in order to map the extent to which similarity in terms of skills, gender structures and organizational forms correspond with wage differences. The essay explores one million jobs and ask; what do they have in common, and what is the difference? I will undertake various analyses and neural network models to explore the labor market's need for skills. It makes an objective approach based on demand and structure. And so, job-requested skills show strong significant correlation with wage differentials, which means, the greater skill similarity, the smaller the wage difference.

Keyword: *skill-similarity; labor structures; word-embedding*
Thesis: 30 hp
Level: Second Cycle
Semester/year: At/2021
Supervisor: Martin Henning
Examiner: Jing Wu
Report No:

Foreword

I want to thank Martin, Mom, and my partner, soon wife, Ida.

Thank you.

Table of Contents

1. Introduction

In paid work, many qualitative features can be completely different to other jobs. Researchers, therefore, sometimes explore the qualitative areas of the labor market. Often by examining quality differences between workers (see: Nedelkoska & Neffke 2019). However, there are few studies that explore these differences between jobs. When this thesis seeks to identify the extent to which wages are associated with skills, gender structures, and organizational forms, it examines just over one million jobs in the Swedish labor market.

With the care of the Swedish authorities, this essay can map some of the qualitative features which appears in the Swedish labor market. And based on a large archive of published job advertisements and official labor market statistics, a surprisingly rare assumption about wage mechanisms can develop. Various empirical studies have been conducted on occupational wage (Murnane et al. 1995; Autor, Katz & Kearney 2006; Florida & Mellander 2020) which find, among other, that skills and education are strongly associated with wage gaps in society. Though an active research area, the most scholars suppose, or at least settle with the thought that the employers pay according to the actual skills, while surprisingly few suggests that the employers pay for what they ask for (Thurow 1975).

While Sweden is usually ranked as one of the world's most equal countries (World Bank 2021), a new report (SACO 2021) shows that academic women have on average SEK 3.7 million lower living wage than academic men. With living wage, as an estimate of individuals' profitability over life, the report considers not only wages, study grants and pension but also income tax, loss of income during the study period, unemployment and repayment of student loans. These inequalities are likely due to gender structures, which can be a slow colossus in several countries (Blau & Kahn 2017). Obviously also in Sweden. Therefore, this essay aims to answer how much and to what extent the gender distribution at work can account for wage differences. And, to work in either a female or male-dominated profession; does it affect wages?

Within the framework of the thesis, there is also the purpose of identifying the extent to which workplaces' organizational forms are associated with wages. In these underlying societal structures, various studies usually show that the private sector pays more than the public sector (Giordano et al. 2011). Despite this, there are few studies that not only consider organizational forms but also gender structures and skills. Therein lies the question: What is the most important of these aspects in the labor market? And depending on the profession; how is wage affected by the supply of jobs between private and public jobs, and in occupations that lack job alternative in one sector; are wages in those jobs less affected by organizational form?

Wage issues have also come to the forefront of the issue of global technological change - a rather difficult problem that has divided researchers into different camps. According to some, employment will increase in high- and low-wage jobs, while employment will decrease in between (Acemoglu and Autor 2011). Others suggest that today's technological advances are accelerating the pace of change, and that we are the forefront of a new industrial age (Brynjolfsson & McAfee 2014). There is an imminent change in each scenario, where the latter suppose that computers and machine learning tools will replace larger chunks of the workforce than the former expected.

1.1. Aim and Research Questions

Wage formation, as a concept, could have been interpreted in different ways. Generally referring to the process that shapes wages, often in terms of productivity, inflation rate and the general health of the economy. In this essay, the aim is to take a step closer and explore this formation at a distance where the cracks start to appear. What seems to influence wage gaps? Are those gaps influenced by the labor market's demand and structure? Is similarity in job-required skills associated with the wage gaps; are similarities in the gender distributions of workplaces associated with these gaps; are wage differences somehow related to organizational forms?

Searching to understand how wage interact with the labor market's demand and structure may as well extend our current knowledge of wage mechanisms. I examine a great number of jobs, within 335 occupational categories. The essay aims to identify the extent which skills; gender structures, and organizational forms interacts with wage. In observation, official government sources are compiled, which account for different parts of the Swedish labor market, both in 2014 and in 2018.

2. Theory and Concept

In this section, I first intend to describe how this essay interprets basic concepts such as wage and wage gaps. Then I will set the table for the theoretical parts that are needed to underline how this research is connected to reality. And finally, each subheading culminates in hypotheses.

2.1. Concepts

Wages generally aim to compensate for work, often in terms and conditions according to employers and employees. In this essay, wages more specifically refer to monthly wages, which means the amount paid to the employees each month. Wage is synonymous with the terms salary and pay, which may allow for some variation in the language of the essay.

Wage gaps generally refer to wage differences between wage workers, often in terms of demographics, geography and the general divisions of industry and occupation in the labor market.

A more accurate terminology for the wage gap is wage differentials, which often refers to the very measure of wage differences. And the same goes for the language of the essay.

Occupations generally refer to employment categories, often in divisions of knowledge, experience and general sets of skills and abilities of workers. Another term for occupation is profession, which in the same way refers to the category division for employment. The essay thus uses the term profession synonymous with occupation.

2.2. Similar skills, same pay.

In this essay, skills refer to something that an individual does, and is good at - something that comes with practice, or education. This is sometimes called a competence, but to avoid ambiguity I use the term skill. It is also not too uncommon for skills to be confused with the concept of abilities (Ovens & Smith 2006), which is interpreted a bit differently as: the quality or the state of being able to perform. In the context of job advertisements, abilities are sometimes called soft-skills (Deming 2017), which are not included within the framework of this study.

As for the smorgasbord of skills that read out in vacancies for different jobs, each skill has it unique purpose for their job. In other words, employers describe their needs in terms of skills, some of which are important, and others that complement. Thus, our skills can be seen as indicators of productivity and profitability for the employer (Acemoglu & Autor 2011). And seen from a larger context in which trade in goods and services shapes the body of society, productivity factors are vital for the development of the economy. This interplay between skills and wages may then create certain incentives that enable society to grow and develop.

Generally, there is no confusion as to whether or not skills are related to wages. On the contrary, employers quite frequently state to follow skill-based-pay principles (Lawler et al. 1993). However, it would be interesting to find out to what extent this applies to jobs in the Swedish labor market. After all, every principle leave room for exception.

2.2.1. Hypothesis 1.

H1: The stronger skill similarity, the smaller wage differential

2.3. Similar gender structures, same pay.

Our ideas about gender can be one of the oldest structures in the labor market. Like other ideas that are receptive to us individuals, ideas, such as gender, are made into identities that we can relate to (Acker 1990). Researchers usually refer to the process of socialization when individuals internalize norms and values in society (Clausen 1968). Recognizing the socialization of working life creates insight into the complex interplay that unfolds in the labor market, as well as into the quandaries of social researchers.

Despite the emphasis on gender equity for many of today's employers, women have suffered from gender inequality for centuries. That is, wage differences that cannot be explained in any other way than gender. Nonetheless, our ideas about masculinity and femininity continue to carve out the basic forms of life, which include the decisions of education and career. Francine Blau and Lawrence Kahn do note (2017), however, that the gap is narrowing in the United States in observations of the recent decades. Still, the convergence slowed and became more uneven after the 1980s (Blau & Kahn 2017)

Within the framework of the essay, the relation between gender structures and wages, two outcomes are expected: Similarities in gender structures are negatively associated with the wage gap, which means that wage differences decrease, or else; with a positive effect in which the wage differences increase. A negative outcome could indicate some support for the claim: similar gender structure, same pay. There are some indications that this may be the case. Particularly, a relatively strong opinion for increased equality in the Swedish labor market, which indicate both a will for change and a room for improvement.

2.3.1. Hypothesis 2.

H2: The stronger gender structure similarity, the smaller wage differential

2.4. Similar organizations, same pay.

The thesis explores the interplay of wage differentials with organizational forms and focuses on differences in the forms of ownership. A fundamental difference in ownership often refers to organizations of either public or private ownership. These are usually referred to by the terms the public sector and the private sector. Thus, the essay examines sector distribution ratio from the data obtained for sector affiliation in workplaces by occupation. Similarly, I study the distribution ratio for organizational forms.

The interplay between wage and organizational forms illustrates some of the most characteristic features of the labor market. The clearest sign may lie in the organizational forms' different goals and objectives, which mean different requirements and conditions at work (Lucifora & Meurs 2006). An important dividing line is financial support. Public sector organizations receive financial support for the purpose of maintaining and guaranteeing certain services. In this respect, companies in the private sector are in a more precarious position, of which an important purpose will be to financially support their own cause. Given this, to some extent, independent position, there is also no requirement to guarantee certain public services.

The essay explores similarity in the sector's distribution ratio, in which two outcomes are expected in the interaction with wage differentials: The more similar sector structures, the smaller the wage

gaps, or the more dissimilar sector structures, the smaller the wage gaps. Considering the earlier research, the current outcome can be excepted lean towards the direction of similar sector structure, the same pay. For example, Ghinetti and Lucifora (2015), note that the return on sector is significant in France, Britain, and Italy. On the other hand, at least 13 years have passed between those observations and the essay.

2.4.1. Hypothesis 3.

H3: The stronger sector similarity, the smaller wage differential

3. Method

In this section, I first intend to explain where the essay's data content comes from. Then I want to describe how the data can be seen from a statistical perspective in terms of sample and population. Furthermore, I describe how data collection is handled with automation and text extraction tools. Then the focus shifts to data analysis, in which I describe how the essay is analyzed and compiled for the statistical analysis. And finally, I describe the statistical regression model used in the essay.

3.1. Sample and Population

With the care of the Swedish authorities, the Swedish Public Employment Service, AF, and Statistics Sweden, SCB, this thesis was able to map a generous share of the demand for labor and skills in the Swedish labor market. AF and SCB provide statistics that they produce themselves, and for this essay, in different parts of the research.

Initially, the essay collected data from AF's historical database archive of job advertisements, whereby I collected the total job ad data in the years 2014 and 2018. That is, two individual years in the span of five years. I then compiled this data with the data from SCB's data archive of occupational structures, whereby I collected standardized occupational classifications and annual aggregated data of the Swedish labor market.

The compiled data set contained 1 070 315 job advertisements, of which approximately one third of the advertisements were published in 2014 in AF's job advertising platform. SCB's annual statistics contain statistics on Swedish occupational structures, whereby the official statistics are declared quality at publication and evaluated annually.[1] The authority clearly points out where there may be inadequacies. This essay uses official statistics on occupational wage structures, in which the authority underlines caution, and describes that the estimation procedure differs when collecting data on occupational wages in private and public jobs.

[1] Note: For further information about Statistics Sweden's quality declaration, visit: https://www.scb.se/dokumentation/kvalitet-och-framtagning/

In order to compile job advertisement data with the aggregated labor market data, and to make them compatible with each other, I used AF's translation code to insert SCB's Standard Classification of Occupations within the data set of job ads. This standard classification forms the basis for labor market data in SCB's data archive of occupational structures. That is, SSYK 2012.[2]

In summary, data for job advertisements were categorized for 335 occupations, which in addition cover 20 industry areas. In doing so, the essay could add the official statistics for professional structures in Sweden.

3.2. Data Collection

The essay categorizes AF's historical archive of job advertisements with SCB's occupational classifications in the labor market. And so, AF's open source and developer community provided several good examples of how this could be done.[3] In addition, the authority offers various analysis tools to explore both archived and real-time ad data. I use one of their tools to extract skill concepts that appear in the ad text content.

The tools are equipped with an application programming interface, API, which anyone who registers with AF's developer community can use for purposes of both small and large scale. The API used in the framework of this study received, for example, my sent API calls. Consisting of a certain amount of job ads as well as their textual content, which were then returned (tirelessly) back to the user, containing an extracted list of terms that appeared in the individual ads, which for the model were likely to fall into the category of skills.

The essay collected and categorized data in an automated order that was adapted to AF's analysis tool for text extraction. I used the programming application Apache NiFi, which enables automated flow-based data management.[4] This allowed me to define where the data was retrieved from and how it was handled. In doing so, I adjusted the pace of the data flow to interact with AF's text extraction tool. Thus, data collection and processing were automated to meet the conditions of the external tools. Similarly, the thesis seeks to ensure reproducibility (Djurfeldt et al. 2013.)

[2] Note: For further information on the Swedish occupational register, visit https://www.scb.se/en/finding-statistics/statistics-by-subject-area/labour-market/employment-and-working-hours/the-swedish-occupational-register-with-statistics/

[3] Note: For further information on AFs open data and source code, visit https://jobtechdev.se/en/about-jobtech-development

[4] Note: For further information on the software used for data collection and processing, visit https://nifi.apache.org

3.3. Data Analysis and Compilation

Once the initial steps of data collection and processing finished, words and numbers could gradually be broken down into results. From AF's text extraction, there were a wonderful number of skills. Well, about 15 million of them. In the subsection, I intend to describe how the essay handled data analysis and compilation.

Table 1. Skill Extraction Outcome

Archivist: 16,255 terms extracted, of which 1,416 are unique. A total of 852 ads from 2014.
['Archive report', 'Information science', 'Driving license', 'Systems science', 'Long-term storage' 'IT skills', 'Process mapping', 'Coordination'],
['Computer skills', 'Archive education', 'Archive databases', 'Research', 'Participant observation', 'Photographing', 'Documentation methods', 'Ethnology', 'Fieldwork']
Demolition worker; machine operator: 4,877 terms extracted, of which 662 are unique. A total of 661 ads from 2014.
['Driver's license', 'Crusher', 'Vocational certificate', 'Swedish', 'Excavator', 'Wheel Loader', 'Driving license', 'Own car'],
['Loader', 'Professional certificate for excavator', 'Land development', 'Own car', 'Swedish', 'Driving license', 'Land development', 'Wheel Loader', 'Sorting', 'Crushing plant']

Table 1 shows a brief summary of data on extracted terms in job advertisements, which in the first case concerned vacancies for archivist jobs, and in the second case demolition jobs. In each case, two sets of extracted terms are compiled from the job advertisement data in the study. Please note: the text content in the lists has been translated by Google Translate from Swedish to English. Some words might thus read out wrong in this table.

3.3.1. Skill extraction.

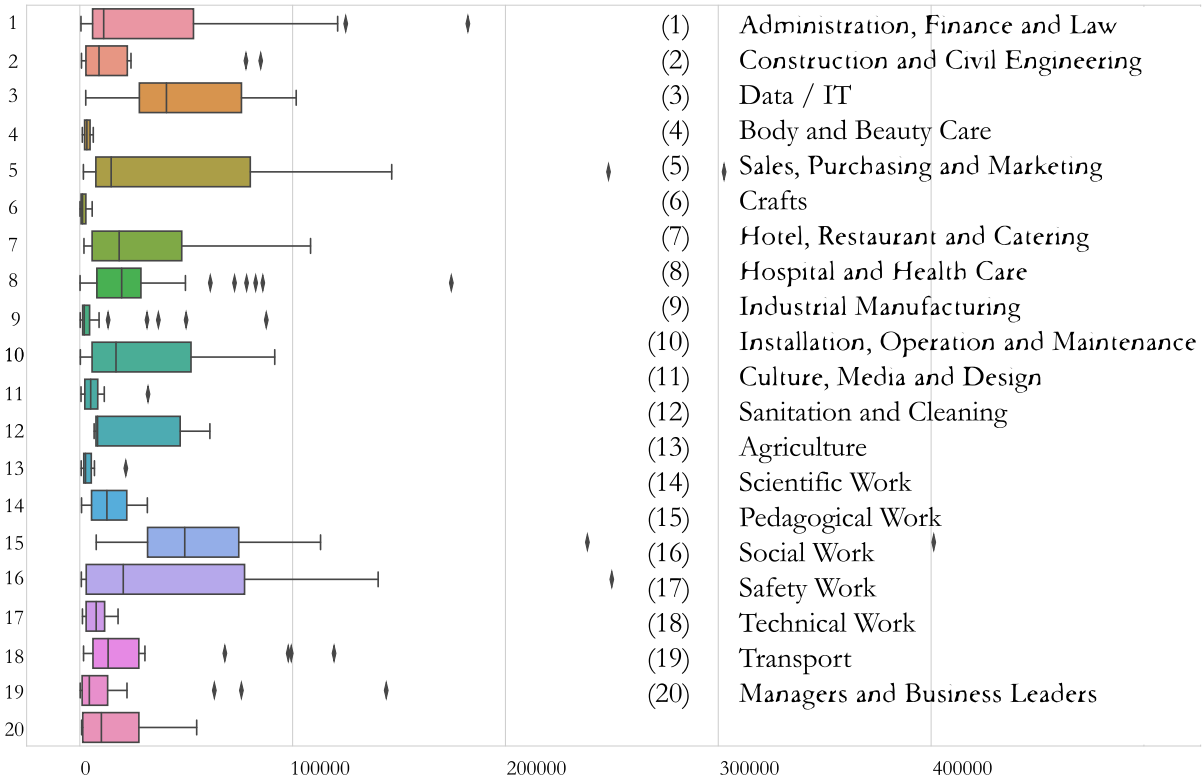
Thus, a generous collection of concept skills was acquired with AF's text extraction tool JobAd Enrichments. This tool is an AI solution that automatically retrieves relevant words and phrases in job ads while filtering out redundant information. I adjusted the tool to extract words that belong to skill concept classifications. Then I programmed the tool to return the outcome into the relevant ad in the dataset. The results are exemplified in Table 1. There I show the outcome from the text extraction, in which two professions are presented and discussed.

Table 1 shows that the number of job advertisements relating archivist jobs amounted to 852 in 2014, which is 191 more than the published demolition ads during the same year. Moreover, it is reported that just over 16 000 skill concepts had been extracted from advertisements concerning archivists, which can be compared with nearly 5 000 skill concepts for the demolitionist. Although the set of skills is greater for archivists, the proportion of unique skills within the total set suggest that they are repeated more often, i.e., about eleven times on average.

The essay explored the variation of the extracted skills for the years 2014 and 2018, where boxplots illustrated the extracted concept variation. In summary, the distribution varied between occupations, where in some outliers fell far beyond the average. Figure 1.1 presents the distribution of the extracted skills in 2018. For skill distribution of 2014, see Figure 1.2, Appendix.

In Figure 1.1, I compiled the number of extracted skills in occupations per industry in order to examine properties of the data in 2018. A closer look indicates that the supply of skills for occupations in the Data / IT industry (read: no 3) is more frequent than the skill concepts in the administrative sector (read: no 1). On the other hand, it can be seen that certain administrative professions are between 100 000 and 200 000 skills concepts. Thus, a large variation of the extracted skills between both professions and industry in 2018.

Figure 1.1. Extracted skills in occupations by industry, 2018



Boxplot in Figure 1.1. shows boxes, which hold the middle half of the proportion of professional skills by industry. The dots represent individual outliers. The calculation is based on a total of 10,198,673 skills, of which a total of 637,514 job advertisements is extracted.

3.3.2. Word embedding.

In the calculation model that I used to calculate similarity between skill sets, two conditions were met. First, the number of skills varied between professions. The archivists resulted in about 16 000 skill concepts, while the demolitionist resulted in considerably less extracted skill concepts. Thus, the calculation model needed to consider text variation. Secondly, the concepts are repeated somewhat differently between occupations, whereby unique and synonymous concepts are observed. The calculation model therefore needed to be able to associate between words.

Word embedding refers to calculation models that represent words in forms other than text characters. Embedding can be used to find words that have some kind of connection to other words - it can be that they are synonyms, or opposite, or in some way occur in the same context (Mikilov 2014). I use a kind of paragraph embedding model, which is an unsupervised method for learning distributed representations of text data in paragraph form. This tool is a neural network solution developed in open source and offered by the analysis company Gensim.[5]

The calculation model for similar skill concepts reads each occupational set in the data, where I train the model to learn to associate between words within the vocabulary. Learning is iterated to increase reliability in the outcome. In addition, I use negative sampling, which is used in research to modify weight and deal with frequent word repetition (Cosica & Neffke 2017.)

3.3.3. Skill similarity index.

The essay indexes skill similarity using the cosine similarity measurement, which measures the similarity between two vectors in an internal product space. In this way, the essay measures skill similarity by the cosine of the angle between two paragraph vectors. That is, paragraph vectors from the paragraph vector model.

To illustrate this, Figure 2 demonstrates indexed cosine similarity between industrial divisions for occupations in Sweden. In this case, the paragraph vector model has been programmed to read the individual skill sets and develop paragraph vectors by profession, and then by industry. The cosine calculation is then repeated in a sequence of operations a specific number of times, giving the indexed similarity shown below in matrix form for a heat map illustration.

In general, the cosine scales between 0 and 1. This essay scales the cosine's similarity between 0 and 100, which is shown on the right in the heat map. The map highlights similarity through color intensity, where white indicates absolute difference and dark blue indicates absolute similarity.

[5] Note: For further information on Doc2vec, visit:
https://radimrehurek.com/gensim/models/doc2vec.html

Figure 2. Predicted Skill Similarity by Swedish Industry Divisions, in 2014

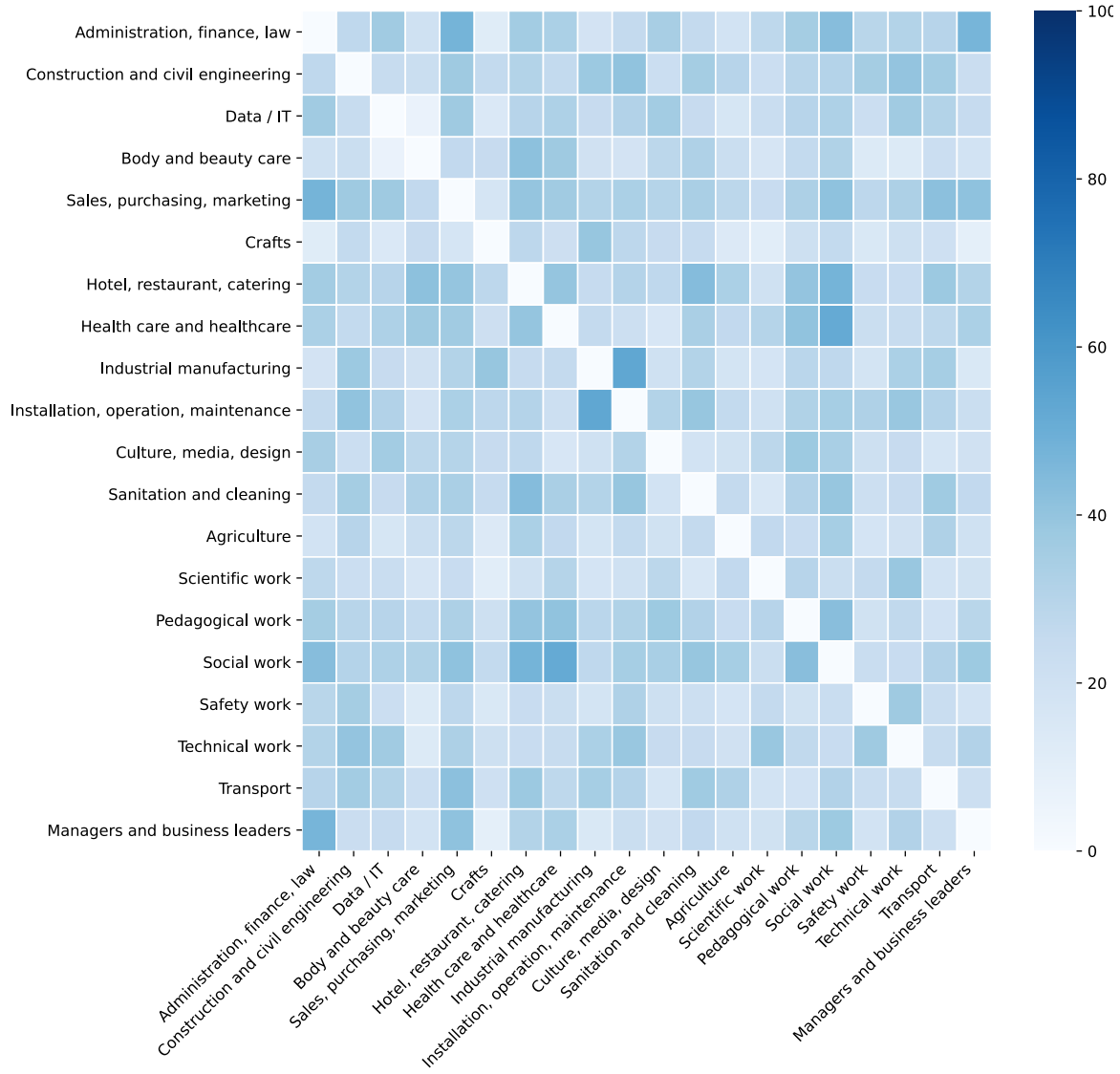


Figure 2. Predicted similarities in skills between industries in Sweden in 2014. A heat map presented by pairwise similarity distances between N-dimensional vector representations of words. Paired distances have been calculated according to a cosine measure and then converted to percentage.

3.4. Regression Model

This section describes the type of linear least squares method that the thesis used for estimating correlation. Although, I first summarize how the essay arrange the data, and then follow with a method description. In the subsection, I describe how the statistical variables are coded for the regression model in Table 2.

In the regression analysis, data are compiled according to unique combinations of occupations, where data on 335 occupations are reorganized into 55 693 unique occupational combinations. This form of combination is called permutation, which is illustrated by the arrangement of the observations in Figure 3 (see: Appendix). Thus, occupations are arranged in pairs by indexed similarity of skills, gender structures and organizational forms. In this multivariate regression model, the independent effects are controlled for aspects such as average age level and average level of education.

The method for estimating correlation in this case refers to the ordinary least squares method, which generally aims to minimize the difference of the observed responses with the responses predicted by linear approximation (Djurfeldt et al. 2013). This regression model can mainly be interpreted as: the greater similarity, the smaller wage differentials, and vice versa. In sum, I examine wage differentials along three independent variables, controlled by five control variables.

3.4.1. Statistical coding.

The dependent variable of this essay refers to wage differentials of average monthly wages between occupations. Thereby I constructed a relative wage differential measure in order stable variations of average monthly wages. Thus, the measure compares two quantities while considering sizes in relation to each other.

The independent variables of the essay refer to skills similarity, gender structure similarity and organizational similarity, in which the values are coded in uniform scale and direction. Scales read between 0 and 100, where 100 indicates absolute similarity, and 0 indicates absolute dis-similarity. Skill similarity is derived from the extent to which occupations are indexed similar in occupational sets of skill concepts. Gender structure similarity and organizational similarity are both derived from the extent to which the distribution ratio converge within the occupational combination.

The thesis controls the independent effects of three demographic control variables. Technically, this is coded by a type of dummy variable construction where the indicatives aim at pairwise conformity. This conformity is explored along industry, average age level and average level of education. Direction and scale are equal, with 1 indicating conformity, and 0 indicating deviation.

Gender structure similarity and organizational similarity are further controlled by the variables unequal gender distribution and uneven sector distribution. Similar type of dummy variable is constructed in which occupational pairs are either read as strongly unevenly distributed (75/25) or distributed closer to the middle (50/50). Correspondingly, 1 indicates conformity in the strong uneven distribution, while 0 refers to more evenly distributed occupations.

4. Result

In this section, I present the findings from the thesis regression analysis. I primarily intend to identify the extent to which wage differentials interact with skill similarity, gender structure similarity and organizational similarity. The results are presented according to the multivariate regression models in Table 2.

Table 2. Summary of Regression Models Predicting Wage Differentials in 2014 and 2018

	2014				***2018***			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
skill similarity	-0.341***	-0.247***	-0.230***	-0.249***	-0.623***	-0.434***	-0.400***	-0.437***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.016)	(0.016)	(0.016)	(0.016)
gender structure	0.075***	0.075***	0.063***	0.075***	0.161***	0.170***	0.142***	0.172***
similarity	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.006)
organizational	0.0004	0.026***	0.025***	0.024***	-0.023***	0.032***	0.033***	0.031***
similarity	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)	(0.004)	(0.004)
same industry		-0.1701	-0.0838	-0.1859		-1.542**	-1.025**	-1.602***
		(0.241)	(0.240)	(0.241)		(0.545)	(0.542)	(0.545)
same average age		-2.829***	-2.689***	-2.841***		-4.86***	-4.403***	-4.922***
		(0.161)	(0.161)	(0.118)		(0.365)	(0.364)	(0.365)
same avg. education		-8.196***	-8.036***	-8.193***		-18.71***	-18.20***	-18.72***
		(0.118)	(0.118)	(0.118)		(0.266)	(0.266)	(0.266)
unequal gender distributions			-2.843***				-7.665***	
			(0.131)				(0.307)	
uneven sectoral distributions				0.746***				2.123***
				(0.116)				(0.259)
Constant	22.30***	23.28***	24.31***	22.85***	34.06***	36.37***	38.96***	35.14***
	(0.215)	(0.213)	(0.217)	(0.222)	(0.496)	(0.489)	(0.498)	(0.512)
Adjusted R^2	.056	.139	.147	.140	.035	.121	.130	.122

Table 2. Summary of regression models predicting wage differentials in 2014 and 2018. Observations: 55684
Note: The OLS model is built using the Python model statsmodels.
*Unstandardized coefficients; standard errors in parentheses. *p < .05. **p < .01. ***p < .001.*

4.1. Regression Models of 2014

The findings from the multivariate model for 2014 can be interpreted as follows: Skill similarity had the greatest effect on wage differentials. Both direction and significance correlated according to H1, where the greater skill similarity, the smaller wage differential. Gender structure similarity indicated less association with wage differentials. This contradicts the H2 assumption, where the stronger gender structure similarity, the smaller wage differential. Organizational similarity only

interacted with wage differentials after the effect was explored in interaction of the control variables. With few signs of statistical significance, this effect deviates from the H3 assumption, where the stronger organizational similarity, the smaller wage differentials.

4.1.1. Findings from the regression models 2014.

Model 1 reports the effects from the independent variables of skill similarity, gender structure similarity and organizational similarity. Given the estimates, sufficiently strong support for the skill similarity and gender structure similarity is included in the model for statistical certainty at 95% significance level. A negative coefficient of 0.34 gradients is read for skill similarity, which means that the expected relative wage differential decreased by 34 percent when skill similarity increased, on average. A positive coefficient of 0.07 gradients for gender structure similarity means that wage differentials increased by about seven percent, when the gender structure similarity increased.

Model 2 controls the effect of independent variables with the effects of same industry, average age level and average level of education. The interaction includes sufficient support to ensure all independent variables at a 95% significance level. The effect of skill similarity drops 0.10 degrees, which means a negative coefficient at 0.25. The model shows that the same average level of education reduces the expected wage differential by eight percent, compared to wage differentials of dissimilar level of education.

Model 3 shows that when the unequal gender distribution overlaps among occupations, the expected wage differential decreases by 2,8 percent, compared with wage differentials of equally distributed occupations.

When the uneven sector distribution was kept constant in model 4, the relative wage differentials increased by 0,75 percent, compared with wage differentials of evenly distributed occupations.

The independent variables are estimated to explain about five percent of the variation in wage differentials. This estimate increased to about 14 percent when the independent effects were controlled by overlap in industry, average age level, and average education level.

4.2. Regression Models of 2018

The findings from the multivariate model for 2018 can be summarized as follows: Skill similarity had the greatest effect on wage differentials. Given the estimates, both direction and significance may seem to support the H1 assumption. The effect of gender structure similarity on wage differentials reaffirmed direction and significance. The interplay contradicts the H2 assumption, which states that the stronger gender structure similarity, the smaller wage differential. Organizational similarity had minor effects on wage differentials, which alone does not seem to substantiate the H3 claim.

4.2.1. Findings from the regression models 2018.

Model 1 displays the effects from the independent variables of skill similarity, gender structure similarity and organizational similarity for 2018. Given the estimates, sufficiently strong support for the independent variables is included in the model for statistical certainty at 95% significance level. The effect of the skill similarity interacts strongly with wage differentials, with a negative coefficient of 0.62. A weak correlation for gender structure similarity can be noted with a positive coefficient at 0.16.

Model 2 shows that the independent effect of skill similarity drops to a negative coefficient at 0.43. Thus, enough statistical support can continue to ensure all independent variables at the 95% significance level. The effect of control variables can be seen to increase compared to 2014. Given the estimate, the wage differential decreased by 18 percent when the average age level overlapped among the occupations, compared to wage differentials of different average age levels.

Model 3 keeps the unequal gender distribution constant. In doing so, the expected wage differential decreases by 7,5 percent, compared with wage differentials of equally distributed occupations. In addition, the effect of gender structure similarity decreases by about three gradients, which gives a positive coefficient at 0.14.

Model 4 shows few signs of change for the independent effect of organizational similarity. On the other hand, the wage differential of unequal distribution is expected to increase by about two per cent compared to the wage differential of evenly distributed occupations.

Independent variables are estimated to explain about three percent of the variation for wage differentials in 2018. The proportion of the variance rises to about twelve percent by including the effects of industry, average age level and average level of education.

5. Discussion

The results indicate a rather significant link between skills and wage differences that develop stronger over the years. The analysis supports the theory that similarities in skills overlap in wages, which is explained by reasoning where skills have specific economic purposes for employers. Also from a larger perspective where financial incentives for skills are effective means and market instruments. This development, where the need for skills takes a greater influence on wage formation, may also point to clearer competition between professions. The structural change of the labor market could provide an explanation in that direction. (Acemoglu & Autor 2011; Brynjolfsson & McAfee 2014)

The results show a weak relationship between gender structure and wages, which strengthens over the years. Contrary to the hypothesized association, these findings indicate that similar structures increasingly correlate with greater wage differences. This development is likely to accommodate more nuanced patterns of the labor market. Results may, for example, indicate that there are clear wage divisions within the gender structures. In other words, various wage developments, heading in different directions. This may suggest that gender differences solidify in career and educational choices, which is somewhat confirmed by observations from the report on Swedish living wages (SACO 2020).

The findings do not seem to indicate an interplay between organizational similarities and wage differences, despite support at the 95% significance level. The reason is that there is not enough variation between the variables. In that sense, it can be quite challenging to place small effects in nominal or absolute terms.

While previous research has generally focused on actual skills, and how these human qualities affect wages (Nedelkoska & Neffke 2019), this essay rather show that the demand for skills has an influence on wage formation. The results may reflect Lester Thurow's ideas on the mechanics and functioning of wage setting in the labor market. According to Thurow (1975), employers probably pay for what they ask for.

This rather unexplored data may contribute to a clearer understanding of labor market demand. The essay concerns only a fraction of the collection of "wish lists" that are made available. The data can thus be processed to capture the qualities that are in demand such as skills, soft skills, tasks, knowledge. As a result, it becomes challenging to claim that the data fills a real substance. On the other hand, it may be just as difficult to claim that wages do.

In this vacuum, questions may be raised as to why jobs are not paid the same. In this case, the results can, for example, explain a small part of a much larger context. More specifically, it is shown that the estimate of the proportion of the total variance ends up being between three and five percent with respect to the independent variables. The essay rather provides insight into a new approach to examining issues more closely.

6. Conclusion

This essay aimed to discern the extent to which skills, gender structures and organizational forms influenced wages. Three hypotheses were developed for this purpose. An extensive data archive of job advertisements was undertaken and examined in relation to aggregated labor market data. The thesis explored these interactions from data concerning one million jobs. There, at least one

hypothesis could be confirmed. Skills. The findings suggest that similar skills overlap with similar wages. This interaction grew stronger over the five-year period between 2014 and 2018. Almost double in strength, which alone may point in the direction of the statement: similar skill, same pay.

The thesis applied various analysis tools and combined unstructured data with aggregated data on labor market structures. A relatively unexplored data was designed from the approach. 335 occupations in the Swedish labor market were mapped and compiled in relation to each other. The essay looked for their common features from data obtained through official authorities. There, the second statement could be contradicted. These results indicate that similar gender structures overlap with greater wage differences. The thesis discussed whether the finding reflected a more complex relationship, in which gender structures consist of several wage divisions, or strata. The correlation grew stronger during the thesis five-year period, albeit from weak to moderate.

The intention was to look through the gaps to identify how professions overlap in other areas. Three factors were explored from the gaps. A generous collection of data was automated and processed systematically. Based on the results from two regression models, the last hypothesis could be rejected. The results indicate that organizational forms interact insufficiently with wages. This correlation was examined between the years 2014 and 2018 without any significant signs of influence.

Reference List

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In Handbook of labor economics (Vol. 4, pp. 1043-1171). Elsevier.

Acker, J. (1990). Hierarchies, jobs, bodies: A theory of gendered organizations. Gender & society, 4(2), 139-158.

Bauer, T.K., 2002. Educational mismatch and wages: a panel analysis. Economics of education

review, 21(3), pp.221-229.

Becker, G.S., 1962. Investment in human capital: A theoretical analysis. Journal of political economy, 70(5, Part 2), pp.9-49.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. Journal of economic literature, 55(3), 789-865. Giordano, R., Depalo, D., Pereira, M. C., Eugène

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company.

Colombo, E., Mercorio, F., & Mezzanzanica, M. (2019). AI meets labor market: Exploring the link between automation and skills. Information Economics and Policy, 47, 27-37.

Coscia, M., & Neffke, F. M. (2017, April). Network backboning with noisy data. In 2017 IEEE 33rd International Conference on Data Engineering (ICDE) (pp. 425-436). IEEE.

Deming, D. J. (2017). The growing importance of social skills in the labor market. The Quarterly Journal of Economics, 132(4), 1593-1640.

Djurfeldt, G., Larsson, R., & Stjärnhagen, O. (2003). Statistisk verktygslåda-samhällsvetenskaplig orsaksanalys med kvantitativa metoder.

Dessler, G. (2013). Fundamentals of human resource management. Pearson.

Elliott, R., Lucifora, C., & Meurs, D. (Eds.). (1999). Public sector pay determination in the European Union. Springer.

Eriksson-Zetterquist, U. (2012). Institutionell teori: idéer, moden, förändring. Liber AB.

Florida, R., & Mellander, C. (2020). The creative class and national economic performance. In Development Studies in Regional Science (pp. 553-575). Springer, Singapore.

Hellevik, O., & Johansson, M. (1984). Forskningsmetoder i sociologi och statsvetenskap. Natur och kultur.

Inglehart, R., Basanez, M., Diez-Medrano, J., Halman, L., & Luijkx, R. (2000). World values surveys and European values surveys, 1981-1984, 1990-1993, and 1995-1997. Ann Arbor-Michigan, Institute for Social Research, ICPSR version.

Lawler III, E. E., Ledford Jr, G. E., & Chang, L. (1993). Who uses skill-based pay, and why. Compensation & Benefits Review, 25(2), 22-26

Liang, H., Sun, X., Sun, Y., & Gao, Y. (2017). Text feature extraction based on deep learning: a review. EURASIP journal on wireless communications and networking, 2017(1), 1-12.

Lucifora, C., & Meurs, D. (2006). The public sector pay gap in France, Great Britain and Italy. Review of Income and wealth, 52(1), 43-59.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. The review of economics and statistics, 251-266.

Nedelkoska, L., & Neffke, F. (2019). Skill Mismatch and Skill Transferability: Review of Concepts and Measurements. Papers in Evolutionary Economic Geography.

Ovens, A. & Smith, W. (2006). Skill: Making sense of a complex concept. Journal of Physical Education New Zealand. 39. 72-82.

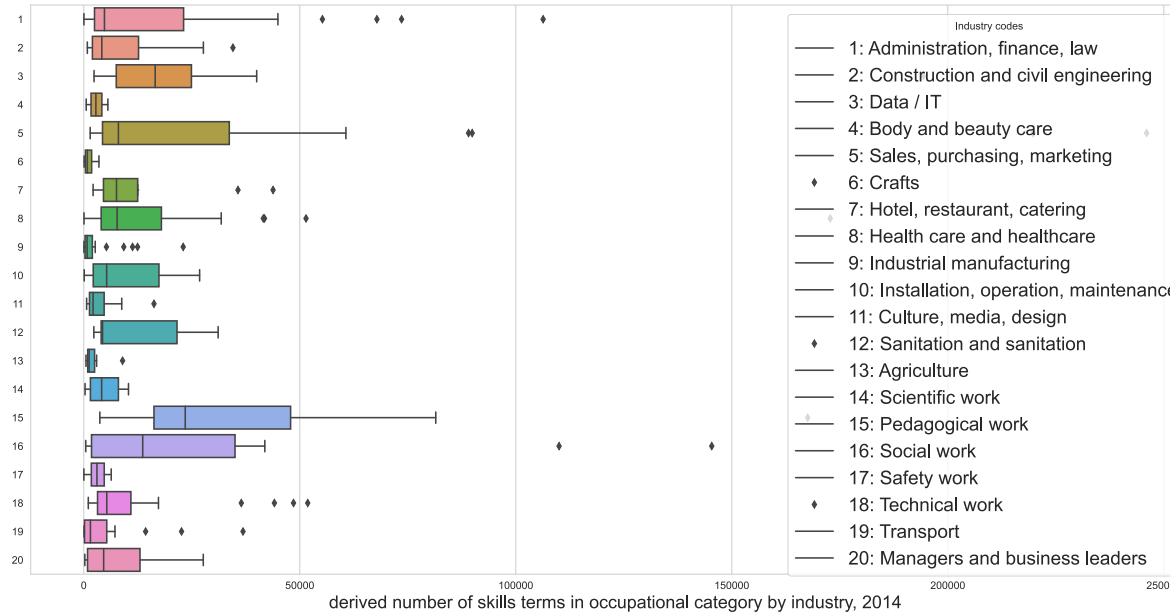
Schultz, T.W., 1961. Investment in human capital. The American economic review, 51(1)

Thurow, L. C. (1975). Generating inequality. Basic books.

World Bank (2021) Women, Business and the Law 2021. Washington, DC: World Bank.

Appendix

Figure 1.2. Derived skill terms in occupations by industry in 2014



Boxplot of Note: Total skill terms: 4803096. Total number of ads: 377906

Figure 3. Correlation of wage differentials with skill similarity, 2014

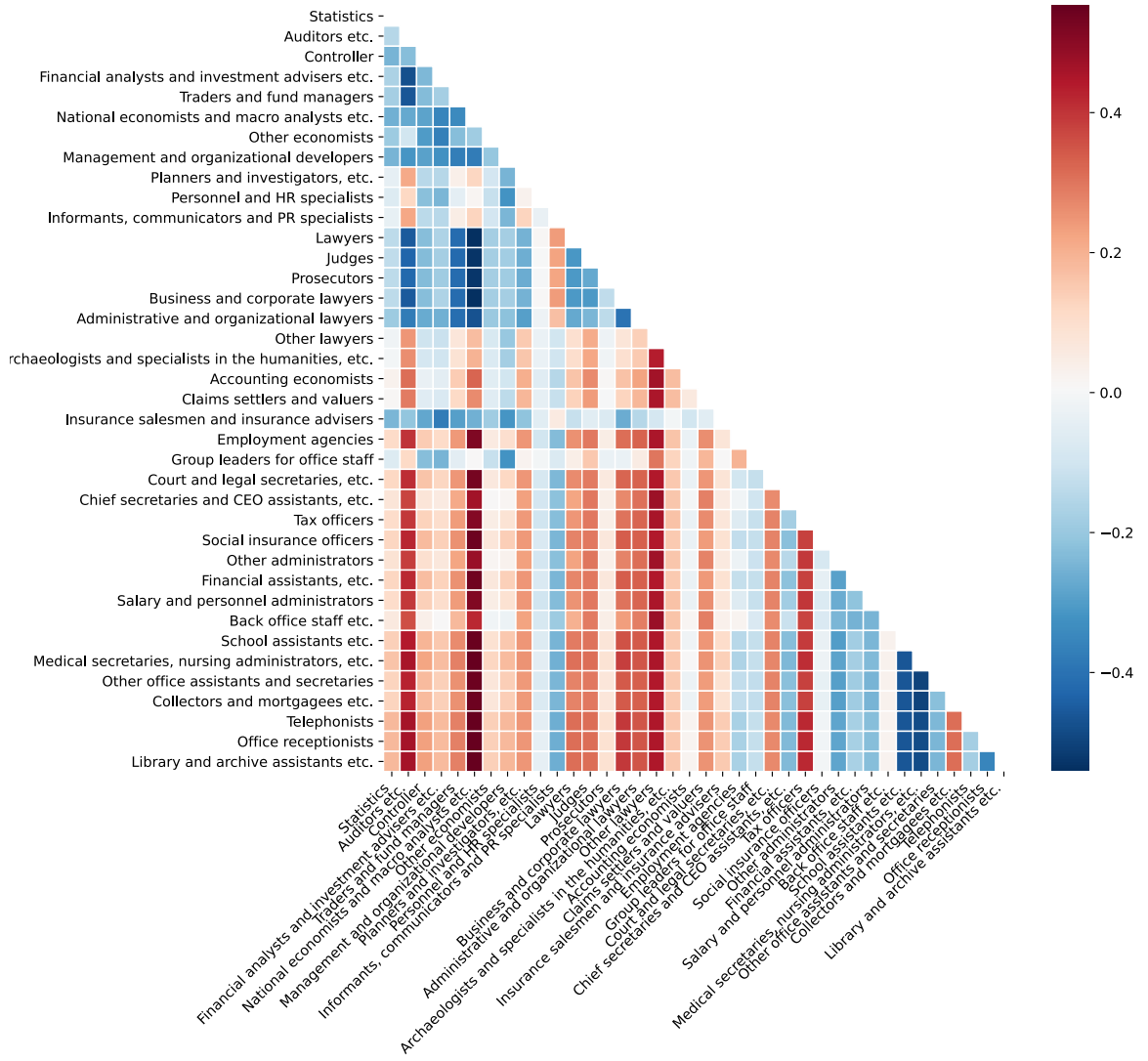


Figure 3. Correlation between wage differential and skill similarity in Sweden in 2014. The map above highlights pairwise correlation between wage differentials and skills, of which skill sets with negative dependencies are represented in red, zero correlation in white, and positive dependencies are blue. Also note: this model does not consider the numerator and denominator's position at the dependent variable, which is taken into account for the main regression models in the thesis.