Essays on Human Capital Accumulation

Susanna Sten Gahmberg

Department of Economics

Norwegian School of Economics

October 2015

Acknowledgements

Whenever I go to a PhD defense, the first thing I do is to read the acknowledgements, and dream about the day it is time to write the acknowledgements for my own thesis. For me, writing the acknowledgements has always symbolized the finish line of writing the PhD. For years it seemed so distant, but now that day is finally here!

Studying for the PhD has been quite an experience. I have learned so much about economics and about life in general, and realized that I have so much yet to learn. There have been good times and bad times, times of great despair and times when I have felt that the Nobel Prize is just around the corner (at the moment it feels quite distant, but as long as I survive the defense, I'm OK - fingers crossed!). I am truly grateful for this experience and I want to thank everybody who have taken part in it, no matter how large or small your contribution might has been. A couple of people deserve a special note.

I start by thanking Norges Handelshøyskole, and especially the Department of Economics, for accepting me into the PhD program, and for all the support along the way. I have always felt that I have been in good hands.

I am deeply grateful for the guidance of my supervisor Kjell G. Salvanes, whose feedback and support has been highly valued and appreciated. I am also very thankful to him for introducing me to his research network and for the many opportunities to travel to workshops, conferences and other research visits. I am also grateful for the help and support from Aline Bütikofer, who seems to have the answer to most of my questions.

Further, I want to thank my fellow PhD students at NHH for sharing thoughts and ideas on all aspects of life. Especially, I would like to thank Eirin Mølland and Katrine Holm Reiso for sharing the ups and downs of PhD life, and Morten Sæthre and Magnus Helliesen for their assistance with various issues, and especially with millions of LaTeX issues!

Over the years I have met a lot of inspiring people and had many nice

experiences abroad, and I want to thank the people who have made this possible. Especially, I thank my co-author Steve Machin for the opportunity to visit the Centre for Economic Performance at LSE in London several times over the years. In June 2015, I was so fortunate to visit the OECD in Paris to work on the analysis for the second chapter of this thesis, and I am sincerely grateful to Mariagrazia Squiccarini at the Directorate for Science, Technology and Industry for this opportunity. I would also like to thank Carl Gjersem and the Ministry of Trade, Industries and Fisheries for financial support for going to Paris.

My family and my friends have been very important in this process and I thank all of you for your support. I want to thank my parents for always being there for me - it means a lot to me.

Especially, I want to thank my husband Ville and my daughter Juni. Thank you for being who you are and for reminding me that there are more important things in life than work. I am really looking forward to spending more time together with you from now on! My last words go to our cat, Bagira. You have absolutely no discretion when it comes to loudly demanding your playtime - 5 a.m. in the morning, the second I sit down for dinner, when I try to focus on work and when Juni is about to fall asleep and I try to be quiet seem to be your favorite occasions. But our bedtime play sessions have been a very nice way of relaxing after a long day of work, and I thank you for that.

Bergen, October 2015
Susanna Sten Gahmberg

Contents

Introduction

The common theme of the chapters in this thesis is investments in and returns to human capital. Human capital is a very broad concept that has many dimensions, including a wide range of cognitive and non-cognitive skills. When talking about human capital, however, we often think about education. In the discussion that follows, I also mainly restrict my attention to investments in education that are made before entering the labor market.

Human capital investments have clear benefits both for the individual and for society. First and foremost, there are large private returns in terms of higher higher productivity and earnings (Becker, 1964). However, there is also much to gain for society as a whole. Investments in skills are crucial for economic growth, both in terms of directly increasing the productivity of workers, and indirectly through a higher rate of innovations in economies that have invested in skills. Another important motivation for investing in skills is its impact on wage inequality. Goldin and Katz (2009), among others, show that in periods when the demand for skills is increasing (as a consequence of skill-biased technological change, for example), wage inequality would increase in the absence of increasing supply of skilled workers.

Human capital theory, as first formulated by Becker (1964), views education as an investment that has future returns, which arise because education raises the productivity of the worker. The optimal investment in education is determined by comparing the costs of obtaining more education and the expected returns. The costs are the sum of directs costs of obtaining more education (such as tuition fees, study materials etc.), indirect costs (such as foregone earnings while studying), and possibly a physic cost of obtaining more education. For the most part, we think of the returns to education as higher earnings, but the returns can also take non-pecuniary forms, such as higher employment probabilities, more satisfactory tasks at work, better health, improved attractiveness/success in the marriage market and marriage stability, parenting skills, as well as spillover effects on

children and increased patience (see, for example, Oreopoulos and Salvanes (2011) for an overview of non-pecuniary benefits from education). Human capital theory postulates that an individual will invest in education until the point where the marginal return to education is zero.

In the first chapter, I study whether financial incentives affect the pace of human capital accumulation of Norwegian students, by evaluating the effect of a student financial aid reform. In this setting, the students have already decided on their optimal level of education, but they spend considerably longer time on obtaining their degrees than stipulated by the government. One plausible explanation for this was the fact that the direct costs of studying were very low, which resulted in the students putting in suboptimal effort. This was deemed expensive both for the students, because of the foregone labor income, and for society in general, as excess resources were invested in higher education, and because of foregone tax revenues. Therefore, the government introduced a financial incentive to make the students graduate faster. The incentive was given in the form of a reduction of the study loan for students who finished on time, and it thereby increased the cost of spending excess time in the education system.

There is an enormous body of literature that tries to estimate the returns to human capital investments, and in the spirit of the influential work of Mincer (1974), there has been a strong convergence towards interpreting human capital as education. Part of the explanation for this development is of course that the purpose of education is to increase the skills, i.e., human capital, of the student. But another explanation is that data on education is, in contrast to many other potential measures of human capital, often readily available in data sets.

One weakness of using education as a measure of human capital appears when we compare investments in and returns to education across countries. It immediately becomes clear that there are as many education systems as there are countries and it is highly likely that there are productivity differences between educational systems in different countries. In other words, one year of education in Italy does not increase the human capital of a student as much as one year of education in Finland. These differences have become more evident in the last few decades, especially after the first PISA results were published in 2000 (OECD, 2000). The PISA results show that the skills of 15-year old students, who have the same amount of education, vary considerably between countries. While the flaws of using education as a measure of human capital in comparative studies of the returns to human capital investments is increasingly appreciated, the fact

remains that there are rarely better alternatives available.

One recent exception is the Survey of Adult Skills (PIAAC),[1] which is an internationally comparable survey of cognitive skills of adults in 24 countries.[2] What makes the PIAAC survey unique is that it also contains a rich set of background information of the respondents, including earnings. Using these data, Hanushek et al. (2015) have shown that there is considerable variation in the distribution of and returns to cognitive skills between the participating countries. In chapter 2, we show that the distribution of numeracy skills varies considerably even between workers with the same level of education in different countries. Similarly, the returns to numeracy skills vary between countries, even within educational groups. These findings highlight the fact it is problematic to use education as measure of skill in comparative studies.

The theme of the third chapter is quite different from the first two, but it is still closely related to investments in human capital. In the last forty years, Norway has undergone enormous changes that have greatly affected gender equality. Norwegian women have overtaken men in educational attainment, and the gender gap in labor force participation has almost disappeared. But despite these un-disputable improvements, the Norwegian labor market remains very gender unequal along certain dimensions. For instance, the level of gender segregation is high in an international comparison, meaning that men and women to a large extent work in different occupations and industries.

In the third chapter, I study the development of industrial gender segregation in the Norwegian labor market between 1970 and 2009. The dominant feature in the labor market was the emergence of the public service sector, which has taken on a wide range of care responsibilities, such as child care are care for the elderly and disabled. The public service sector in Norway is very large compared to that in other developed countries, and it now employs almost 40 percent of the female workers. In practice, the emergence of the public service sector can be seen as a reorganization of traditional female work, which has moved from the informal to the formal labor market. Many women today get paid to perform the same tasks they would have performed at home without getting paid four decades ago. This development has had important implications for the human capital investments of women in Norway. The reorganization of care for children and the elderly has freed many women from these tasks, and allowed them to choose careers in other

[1] And its predecessor International Adult Literacy Survey (IALS).

[2] The respondents are tested in three domains of cognitive skills: literacy, numeracy and problem solving in technology-rich environments.

sectors of the economy. But even so, a large share of women choose education in health and social welfare (while only a fraction of men do so).

Chapter 1: Early bird caught the worm? The effect of a student aid reform on time-to-degree

Enormous amounts of money are spent on student aid worldwide, and in many countries expenditures are increasing along with the student population. At the same time, there are concerns that current student aid systems are inefficient since a large fraction of graduates spend excess time in education[3] and dropout rates are high.[4] Consequently, there is a large public debate on how to redesign student aid systems that i) incentivize students to graduate faster, and ii) target financial aid at those who need it while iii) keeping total costs low. Many countries are implementing new student aid programs,[5] but there is surprisingly little empirical research supporting these reforms. The empirical literature finds mixed effects of student aid on academic achievement, and many of the existing studies cannot convincingly control for confounding factors.

The aim of this study is to analyze the effects of a student aid reform in Norway in the 1990s that aimed at increasing the number of students who graduated on stipulated time and thereby reducing delays in higher education in Norway in the 1990s. The reform, which is often referred to as the turbo reform [*turboreformen*], entitled students in certain graduate programs to an extra reduction of their state funded study loan if they graduated on stipulated time. The reduction was NOK 18,000, which corresponded to about 35 percent of the financial support for one year, or 9 percent of the total loan of a student who had taken up the full amount of loan for her entire course of study. The reform created a discontinuity in the incentives to graduate on time and offers an opportunity to estimate whether

[3]For instance, Brunello and Winter-Ebmer (2003) report that the share of students who expect to delay graduation ranges from close to zero in the UK and Ireland to about 31 % in Sweden and Italy. In Norway, only 29 % of the graduates from 5-year graduate programs and 44 % of the graduated from 3-year undergraduate programs completed on expected time in 2011-2012 (Statistics Norway, 2013).

[4]On average 30 percent of all entrants into higher education do not graduate with a degree in the OECD countries (OECD, 2013a).

[5]For example, Norway implemented a progression dependent student aid system in 2002 (Ministry of Education and Research, 2002; St.meld. nr.7, 2008) and Finland is currently looking to reform their student aid system to more efficiently promote progression (Ministry of Education and Culture, 2010, 2012). In Germany and Italy higher tuition fees for those who are delayed have been introduced (Heineck et al., 2006; Garibaldi et al., 2012). In the US, several merit based student aid programs have been implemented in recent years (Goodman, 2008; Scott-Clayton, 2011; Angrist et al., 2009).

students respond to financial incentives by adjusting their study pace. This study is one of few to study a student aid reform that was directly targeted at reducing delays. Another contribution of this study is to highlight the presence of heterogeneous treatment effects by studying how the take-up varies by parental background, student ability and field of study.

Using a difference in difference strategy and detailed Norwegian panel data on study progression, I find that one additional year of treatment, defined as the number of years enrolled in higher education during the reform period, increased the probability of graduating on time by 1.5 percentage points compared to a baseline probability of timely graduation of 15.5 percent. The reform also reduced duration of delays: one year of treatment reduced delay by 0.13 semesters. There was, however, considerable heterogeneity in the take-up of treatment. The treatment effect was largely driven by high ability students and students with highly educated parents. Further, the treatment effect varied by study program.

Chapter 2: Skills, education and wage inequality (joint work with Stephen Machin and Kjell G. Salvanes)

Differences in the levels and dispersion of adult basic skills have commonly been discussed as a reason why some countries have higher levels of wage inequality. A particular concern has been a higher frequency of adults with low levels of literacy and numeracy skills in some countries, notably the US and the UK, whereas this lower tail is largely absent in other countries. One key issue concerns the source of these basic skills deficiencies, specifically how it can be traced to how well or poorly the schooling systems of different countries deliver literacy and numeracy education. A second issue concerns the wage penalty that poor basic skills impart on workers and how these can explain wage inequality differences across countries.

Much of what we know on these issues can be traced to a series of papers based on the International Adult Literacy Survey (IALS) which took place in the mid-1990s (Freeman and Schettkat, 2001; Devroye and Freeman, 2001; Leuven, Oosterbeek, and van Ophem, 2004; Blau and Kahn, 2005). These papers rely on the observation that there are considerable differences between countries in the distribution of cognitive skills in IALS, and that there is a positive correlation between the inequality of skills and the inequality of wages across countries. The conclusion of these studies is, however, that differences in skills inequality only explain a modest part of the differences in wage inequality. These studies only look at various specific summary measures of wage dispersion, such as percentile

ratios, and not at the full distribution.

We reconsider these findings, first by using a method introduced by Firpo, Fortin and Lemieux (Firpo, Fortin, and Lemieux, 2007, 2009; Fortin, Lemieux, and Firpo, 2011) that allows us to study the impact of skills on the entire distribution of wages, and second by looking at more recent data from the Survey of Adult Skills (PIAAC) (OECD, 2013b). Further, we study the effects of basic skills within educational groups to highlight the fact that the performance gap between countries varies across education groups. We place a focus on two countries, one that has high inequalities in basic skills (the US), and one that has low inequalities in basic skills (Finland). This focus permits us to consider how inequality in basic skills and in earnings can vary right across the entire distribution.

Similarly to what previous studies have found using the IALS data, we find that the level of wage and skill inequality is higher in the US than in most other countries. In our decomposition exercise, we find that changing the distribution of skills has a positive but rather small impact on the distribution of wages in the US. Contrary to what one could have expected, imposing Finland's skill distribution on the US actually increases wage dispersion in the US. This is explained by the higher returns to skills among workers in the higher end of the wage distribution. Changing the in skill prices has a larger effect on the wage distribution in the US. When Finland's skill prices are imposed on the US, US wages are on average reduced, and again, the level of inequality increases slightly.

Our descriptive analysis of skill and wage differences within education groups suggest than the basic skills of low educated workers in the US are considerably lower than those of low educated workers in any other country, but that the cross-country differences are much smaller among more educated workers. This motivated us to study the impact of skills on wages within education groups. We find that low educated workers would gain the most in terms of wages from having the skills of Finnish workers, but that this would also substantially increase the level of wage inequality in this group. The introducing the Finnish skill prices would reduce wages in all education groups, but to a varying extent. High education workers above the median in the wage distribution would be hit very hard relative to workers below the median, which indicates that there are very high returns to skills in the very top of the wage distribution on the US relative to Finland.

Our results are in line with previous work in that skill prices seem to affect the distribution of wages more than the distribution of skills. In contrast to other

studies, however, we find that neither differences in skills nor skill prices can explain the large differences in wage inequality between the US and Finland. We find that both the composition effect and the wage effect of skills would increase wage inequality expressed as the 90/10 differential in log wages. By studying the effect of skills on the entire wage distribution, we give a more nuanced picture of the importance of skills than much of the existing literature. We also show that only measuring wage inequality in terms of percentile differentials is a rather crude measure that can miss important patterns in the data. In addition, we show that the impact of skills varies within education groups, which is a dimension of heterogeneity that has been ignored in much of the existing literature.

Chapter 3: Gender Segregation in the Welfare State: Industrial Segregation in Norway 1970-2009

The Norwegian labor market is among the most gender equal in the world. For instance, Norway has repeatedly scored very well on the UN's Gender Inequality Index[6], and the female labor force participation rate is one of the highest in the world. Still, the Norwegian labor market has long been among the most gender segregated, meaning that males and females to a large extent work in different industries and occupations. One reason for the high level of equality may be the well-developed welfare state and the generous family policies that encourage female labor force participation (Blackburn, Browne, Brooks, and Jarman, 2002; Charles and Grusky, 2004; Mandel and Semyonov, 2006). However, the welfare state may also be a reason for the high level of segregation in the labor market. Advanced welfare states, such as Norway, are characterized by large public service sectors that provide health care, child care and care for the elderly. These services are to a large extent provided by women who would have performed the same tasks as unpaid work outside the labor market if it was not for the welfare state. The public sector is usually also more flexible than the private sector in offering reduced hour contracts, which are easier to combine with family responsibilities. Thus, while the welfare state encourages and facilitates female labor force participation, it channels female workers into the public service sector, which may result in a high level of gender segregation in the labor market.

The aim of this study is to provide an overview of the changes in gender segregation in Norway between 1970 and 2009, and to assess them in the light of

[6]The Gender Inequality Index measures gender inequalities along the dimensions of reproductive health, empowerment and labor market status (UNDP, 2010, 2011, 2013, 2014).

the emerging welfare state. The expansion of the public service sector started in the 1960s, and was extended to cover health care, care for the elderly, education, social security among other things. This development had an enormous impact on the Norwegian labor market, and especially on the employment of women. This paper contributes to the literature in four ways. First, I use a longer time frame than any previous Norwegian study, and can thereby identify trends in a longer perspective. Second, the study also uses more detailed data on industry than previous studies, while it still covers the whole economy. In doing so, I can identify more narrowly defined industries that have been important in driving changes in segregation than before. Third, I study whether the level and trend in segregation differ between skill groups, which also helps to understand the changes in segregation in Norway better. As in most industrialized countries, educational attainment started to increase rapidly in the 1970s, especially among women. Norwegian females have increased their formal qualifications considerably relative to males, and this has changed women's possibilities to compete with men in the labor market. I study whether this also changed the patterns in gender segregation, which is an under-researched area in the literature. Fourth, I discuss segregation in the light of the historical context. This study is purely descriptive, and the aim is not to explain why there is gender segregation in the labor market in the first place. Rather, the objective is to study the observed trends in segregation in Norway using the historical context, and to assess how institutional factors may have contributed to changes in the observed level of segregation.

I find that throughout the period, males and females became more evenly distributed within industries, although this slowed down after 1990 when the female labor force participation rate stabilized. Further, the analysis suggests that changes in the industry composition played an important role in the development of gender segregation over time. The expansion of the health and welfare sector was very important in driving segregation, as it absorbed almost 50 percent of all female labor market entrants since 1970, and employed almost 40 percent of all female workers in 2009. The expansion of child care services and care for the aged and disabled were the main the drivers of between sector segregation, especially after 1990.

The story of segregation in the Norwegian labor market is primarily a story about female workers, but men have also played an important role. In the 1980s, downsizing of male dominated industries, such as agriculture and manufacturing, counteracted the upward pressure that the expansion of female dominated service

industries put on segregation. In later years, male employment increased in male dominated industries such as business activities, mining and quarrying and construction, which lead to more segregation.

The extent of gender segregation also varied between educational groups. Workers with secondary education experienced a rapid increase in the level of segregation after 1990. This was to a large extent driven by relative employment growth in segregated industries such as child care, care for the aged and disabled, construction and transportation. Workers with tertiary degrees (both short and long) experienced a large reduction in the level of segregation between 1970 and 1998, and it was mainly driven by within industry changes in gender composition. Short tertiary degree holders were more concentrated in public sector industries than other educational groups, while public sector industries were slightly less important for the changes in segregation in the long tertiary education group. In the 2000s, there were signs of increased segregation among workers with long tertiary degrees, mainly because the female (male) employment share increased in a number of female (male) dominated industries, such as general somatic hospitals, higher education and veterinary services (software consultancy and supply).

Bibliography

Angrist, J., D. Lang, and P. Oreopoulos (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics 1*(1), 136–163.

Becker, G. (1964). *Human Capital : A Theoretical and Empirical Analysis, with Special Reference to Education.* Univeristy of Chicago Press.

Blackburn, R. M., J. Browne, B. Brooks, and J. Jarman (2002). Explaining gender segregation. *The British Journal of Sociology 53*(4), 513–536.

Blau, F. D. and L. M. Kahn (2005). Do cognitive test scores explain higher u.s. wage inequality? *The Review of Economics and Statistics 87*(1), 184–193.

Brunello, G. and R. Winter-Ebmer (2003). Why do students expect to stay longer in college? Evidence from Europe. *Economics Letters 80*(2), 247 – 253.

Charles, M. and D. B. Grusky (2004). *Occupational ghettos: The worldwide segregation of women and men (Studies in social inequality).* Stanford University Press.

Devroye, D. and R. B. Freeman (2001). Does inequality in skills explain inequality in earnings across advanced countries? Working Paper 8140, National Bureau of Economic Research.

Firpo, S., N. Fortin, and T. Lemieux (2007). Decomposing wage distributions using recentered influence function regressions unpublished manuscript.

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

Fortin, N., T. Lemieux, and S. Firpo (2011). *Decomposition Methods in Economics*, Volume 4 of *Handbook of Labor Economics*, Chapter 1, pp. 1–102. Elsevier.

Freeman, R. and R. Schettkat (2001). Skill compression, wage differentials, and employment: Germany vs the US. *Oxford Economic Papers 53*(3), 582–603.

Garibaldi, P., F. Giavazzi, A. Ichino, and E. Rettore (2012). College cost and time to complete a degree: Evidence from tuition discontinuities. *The Review of Economics and Statistics 94*(3), 699–711.

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

Goodman, J. (2008). Who Merits Financial Aid? Massachusetts' Adams Scholarship. *Journal of Public Economics 92*, 2121–2131.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review 73*(0), 103 – 130.

Heineck, M., M. Kifmann, and N. Lorenz (2006). A duration analysis of the effects of tuition fees for long-term students in Germany. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik) 226*(1), 82–109.

Leuven, E., H. Oosterbeek, and H. van Ophem (2004, 04). Explaining international differences in male skill wage differentials by differences in demand and supply of skill. *Economic Journal 114*(495), 466–486.

Mandel, H. and M. Semyonov (2006). A welfare state paradox: State interventions and women's employment opportunities in 22 countries. *American Journal of Sociology 111*(6), pp. 1910–1949.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. National Bureau of Economic Research, Inc.

Ministry of Education and Culture (2010). Ei paikoillanne, vaan valmiit, hep! Koulutukseen siirtymistä ja tutkinnon suorittamista pohtineen työryhmän muistio. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2010:11.

Ministry of Education and Culture (2012). Opintotuen rakenteen kehittäminen 2012. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2012:29.

Ministry of Education and Research (2002). Kvalitetsreformen. Veiledninger og brosjyrer 14.08.2002.

OECD (2000). Literacy Skills for the World of Tomorrow: Further Results from PISA 2000.

OECD (2013a). Education at a Glance 2013: OECD Indicators.

OECD (2013b). OECD Skills Outlook 2013: First Results from the Survey of Adult Skills.

Oreopoulos, P. and K. G. Salvanes (2011). Priceless: The nonpecuniary benefits of schooling. *The Journal of Economic Perspectives 25*(1), pp. 159–184.

Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources 46*(3), 614–646.

Statistics Norway (2013). Statistisk årbok 2013.

St.meld. nr.7 (2007-2008). Statusrapport for kvalitetsreformen i høgre utdanning.

UNDP (2010). *Human Development Report 2010 - The Real Wealth of Nations: Pathways to Human Development*. Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

UNDP (2011). *Human Development Report 2011 - Sustainability and Equity: A Better Future for All*. Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

UNDP (2013). *Human Development Report 2013 - The Rise of the South: Human Progress in a Diverse World*. Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

UNDP (2014). *Human Development Report 2014 - Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience*. Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Chapter 1

Early bird caught the worm?

The effect of a student aid reform on time-to-degree

Abstract

Delayed graduation from higher education is an issue that has received much attention in recent years. This paper studies students' response to a student aid reform in the early 1990s, aimed at increasing the share of students who graduated on time. In particular, the reform gave students enrolled in some study programs financial incentives to graduate on time by offering a reduction of their study loan. Using a difference in difference strategy and detailed Norwegian panel data on study progression, I find that one additional year of treatment, defined as the number of years enrolled in higher education during the reform period, increased the probability of graduating on time by 1.5 percentage points compared to a baseline probability of timely graduation of 15.5 percent. The reform also reduced duration of delays: one year of treatment reduced delay by 0.13 semesters. There was, however, considerable heterogeneity in the take-up of treatment. The treatment effect was largely driven by high ability students and students with highly educated parents. Further, the treatment effect varied by study program.

1.1 Introduction

Enormous amounts of money are spent on student aid worldwide,[1] and in many countries expenditures are increasing along with the student population. At the same time, there are concerns that current student aid systems are inefficient since a large fraction of graduates spend excess time in education[2] and dropout rates are high.[3] Consequently, there is a large public debate on how to redesign student aid systems that i) incentivize students to graduate faster, and ii) target financial aid at those who need it while iii) keeping total costs low. Many countries are implementing new student aid programs,[4] but there is surprisingly little empirical research supporting these reforms. The empirical literature finds mixed effects of student aid on academic achievement, and many of the existing studies cannot convincingly control for confounding factors.

The aim of this study is to analyze the effects of a student aid reform in Norway in the 1990s that aimed at increasing the number of students who graduated on stipulated time and thereby reducing delays in higher education in Norway in the 1990s. The reform, which is often referred to as the turbo reform [*turboreformen*], entitled students in certain graduate programs to an extra reduction of their state funded study loan if they graduated on stipulated time. The reduction was NOK 18,000, which corresponded to about 35 percent of the financial support for one year, or 9 percent of the total loan of a student who had taken up the full amount of loan for her entire course of study. The reform created a discontinuity in the incentives to graduate on time and offers an opportunity to estimate whether students respond to financial incentives by adjusting their study pace. This

[1]The OECD countries spend on average 1.6 percent of GDP on higher education institutions and 0.31 percent of GDP on student aid. But there is also considerable variation both in the spending as a share of GDP and in the share of public and private funding(OECD, 2014). The spending on student aid varies from 0.02 percent of GDP in the Czech Republic to 0.99 percent in the United Kingdom and in Norway (OECD, 2014).

[2]For instance, Brunello and Winter-Ebmer (2003) report that the share of students who expect to delay graduation ranges from close to zero in the UK and Ireland to about 31 % in Sweden and Italy. In Norway, only 29 % of the graduates from 5-year graduate programs and 44 % of the graduated from 3-year undergraduate programs completed on expected time in 2011-2012 (Statistics Norway, 2013).

[3]On average 30 percent of all entrants into higher education do not graduate with a degree in the OECD countries (OECD, 2013).

[4]For example, Norway implemented a progression dependent student aid system in 2002 (Ministry of Education and Research, 2002; St.meld. nr.7, 2008) and Finland is currently looking to reform their student aid system to more efficiently promote progression (Ministry of Education and Culture, 2010, 2012). In Germany and Italy higher tuition fees for those who are delayed have been introduced (Heineck et al., 2006; Garibaldi et al., 2012). In the US, several merit based student aid programs have been implemented in recent years (Goodman, 2008; Scott-Clayton, 2011; Angrist et al., 2009).

study is one of few to study a student aid reform that was directly targeted at reducing delays. Another contribution of this study is to highlight the presence of heterogenous treatment effects by studying how the take-up varies by parental background, student ability and field of study. Further, I discuss underlying mechanisms of the reform.

The study uses rich and accurate register data on education and student characteristics and links cognitive ability test scores from the military draft to the male students in the sample.

I find that on average one additional year of treatment, defined as the number of years enrolled in higher education during the reform period, resulted in a 1.5 percentage point increase in the probability of graduating on time. The effect is robust to a number of changes in the control group, but slightly smaller than what Gunnes et al. (2013) found in a recent paper analyzing the implications of the same reform. Given that the probability of timely graduation was 15.5 percent in the treatment group in the pre-reform period, this translates into a 58 percent increase in the probability of graduating on time for a student who was treated for six years. The reform also reduced duration of delays; one additional year of treatment resulted in a 0.13 semester reduction in delay. In addition, the treatment effect was strongest among students in the upper tail of the ability distribution and among students from stronger socioeconomic backgrounds.

The treatment effect also varied by study program. Available survey evidence by Berg (1994) is suggestive of why these differences may have occurred. For example, the treatment effect was largest among students in humanities and science who reported that they often delayed graduation to take extra courses. Law students, on the other hand, reported failing exams and struggling with extensive curricula as the most common causes for delaying graduation, and their treatment effect was significantly smaller than the average treatment effect. These findings indicate that students differed in their possibilities to respond to the reform, and also suggest that in order to further reduce delays structural reforms might be needed.

To further investigate the underlying mechanisms of the reform I present estimates of effects on student earnings while studying. While there is no evidence that fewer treated students worked in the reform period, their earnings decreased compared to the non-treated students, suggesting that the treated students cut back on hours worked in the reform period.

The paper is organized as follows. In Section 3.3, I discuss the related

literature, and in Section 1.3 I provide information about the institutional setting and the turbo reform. Data and empirical strategy are presented in Sections 3.4 and 1.5, respectively. The empirical results are presented in Section 1.6 and underlying mechanisms are investigated in Section 1.7. Section 3.9 concludes.

1.2 Related literature

There is a rapidly growing literature that studies the impact of financial incentives on student performance. The literature can be divided into studies that focus on the effect of student aid on the extensive margin, i.e., enrollment or access to higher education[5] and on studies that study the intensive margin, i.e., academic performance, effort and completion.

In more recent years, researchers have shown increasing interest in the effects of financial aid on the intensive margin. One strand of this literature is concerned with academic performance and effort,[6] while the research most in line with this study has focused on duration of study and completion. This literature is less conclusive than the evidence on enrollment and persistence. What is also problematic is that many of the studies cannot convincingly control for confounding factors as the effect is often identified from comparing students who graduated before and after a policy intervention.

One of the earliest contributions to this literature is a study by Häkkinen and Uusitalo (2003) who evaluate a Finnish student aid reform aimed at facilitating full-time studies and reducing study duration by increasing the total financial support. They find only limited effects of the reform and conclude that the absence of an effect is partly explained by increasing unemployment rates that reduced student employment possibilities. Heineck et al. (2006) find that the introduction of tuition fees for delayed students at a German university affected student behavior but that the effect varied by field of study. While tuition fees made students in some majors graduate faster, average duration increased in others. Tuition fees also increased the dropout rate in some majors, which was an unintended and undesirable effect of the reform. Glocker (2011) uses German panel data to study the relationship between student aid and duration and graduation probabilities. She finds that while higher levels of financial aid have no effect on study duration, higher levels of financial aid are positively correlated with

[5]Most studies find that student aid increases enrollment in higher education. For an overview of this literature see Dynarski (2002).

[6]See e.g. Angrist and Lavy (2009); Angrist et al. (2009); Leuven et al. (2010).

the probability of graduating with a degree. Students who receive financial aid are also found to graduate faster than students who fully rely on private funds. Scott-Clayton (2011) finds that graduation rates increased and time-to-degree decreased following the implementation of a merit based aid program in West Virginia that had GPA and course load targets as requirements for the yearly renewal of the aid.

Garibaldi et al. (2012) manage to circumvent the problem of confounding factors using discontinuities in the tuition fees at a private university in Italy as they estimate the effect of increased tuition on the probability of graduating on time. They find that students who face the threat of having to pay higher tuition fees after their expected graduation year are more likely to graduate on time than students who do not face the same threat of higher tuition fees.

The Norwegian turbo reform is also evaluated in a recent paper by Gunnes et al. (2013). The authors focus on the average effect of the reform and find that the reform significantly reduced delay and increased the share of students who graduated on stipulated time. The findings on non-completing are ambiguous, but if anything they indicate that the reform slightly increased completion rates. An interesting finding is that the timing of treatment is important as students who were treated only in the beginning of their studies show a positive and significant treatment effect.

Joensen and Mattana study the impact of student aid on academic achievement and labor market behavior in an dynamic discrete choice framework in a series of papers (Joensen, 2010; Joensen and Mattana, 2014). They find that uniformly increasing student aid reduces the risk of dropout and increases completion rates, but at the cost of longer enrollment. Simulations indicate that a more efficient way to alter the student aid system could be to introduce a merit based system or graduation bonuses (Joensen, 2010). In a recent working paper, Joensen and Mattana find that the relationship between grants and study loans is also important for academic performance. If most of the student aid is given as a grant, an increase in the grant share reduces graduation rates but if most of the student aid is given as a loan, the loan share can be further increased without affecting human capital accumulation (Joensen and Mattana, 2014).

The main message of this literature is that financial incentives have some impact on study duration and college completion, but that the effect on completion is not as strong as that on enrollment. Increasing the level of student aid seems to have a positive impact on graduation rates, but the effect on duration are

less clear (Glocker, 2011; Joensen, 2010). The mix of grants and loans also seems to be important (Joensen and Mattana, 2014). Merit based aid that is contingent on academic performance seems promising in reducing duration, increasing graduation rates and improving overall academic performance (Scott-Clayton, 2011; Angrist et al., 2009; Joensen, 2010). However, there seems to be heterogeneity in the responses. Both Häkkinen and Uusitalo (2003) and Heineck et al. (2006) find that students in different majors react differently to incentives and the latter study also finds increased dropout behavior following the introduction of tuition fees for delayed students.

What much of the existing literature fails to take into account is the heterogeneity in the student population by focusing on treatment effects that are averaged over all students (with the exception of the two studies mentioned above). This misses the point that students delay graduation for a wide variety of reasons, and that these reasons are likely to affect the response to various policy interventions. Two contributions of this paper are therefore to highlight the importance of acknowledging the heterogeneity in the student mass and to study mechanisms when evaluating student aid reform and making policy advice.

1.3 Institutional settings

The higher education system and the student aid system in Norway were both restructured in 2002/2003 as a part of the Bologna process, and what follows is therefore a description of the old systems as they were in the 1990s.

1.3.1 The Norwegian higher education system

The Norwegian higher education system consisted of universities, specialized universities and regional university colleges. All types of institutions offered both undergraduate and graduate courses, but the regional university colleges mostly provided shorter vocationally oriented programs (with a duration of two or three years). Most undergraduate programs at universities lasted for three or four years. The graduate programs were structured either as integrated study programs with a total duration of five to six years, such as medicine or law, or as a combination of an undergraduate and a related graduate program, also with a combined duration of five to six years. The education system was similar to that in the U.S. in that

most students would leave university after the undergraduate level.[7] Both the undergraduate and the graduate degrees in the 1990s were more comprehensive than the post-Bologna Bachelor's and Master's degrees.

The majority of the students were enrolled in public institutions.[8] Tuition fees, which were only paid in the private higher education sector, were low, making the direct private costs of higher education very low.[9]

1.3.2 The Norwegian State Loan Fund

The Norwegian State Loan Fund (NSLF) is the main provider of student financial aid in Norway. The purpose of the NSLF is to promote equality in society by enabling students to participate in education irrespective of age, gender, geographical, economic and social conditions and to ensure a satisfactory work environment for students (Lånekassen, 2012).

Thus, practically all citizens were entitled to financial support from the NSLF if enrolled in higher education. Since the cost of higher education was virtually zero, the NSLF provided loans and grants to cover living expenses during the academic year. The student support was not tested against parental income, but dependent on students' own income and wealth. Students were allowed to work during the academic year, but the allowed earnings were restricted to NOK 5,200 per month. Earnings in the summer months were not included in the calculations.

In the time period of the study 87 % of the financial support was distributed as a loan, and 13 % as a grant.[10] The total support during an academic year was decided upon every year by the Parliament and it ranged from NOK 52,000 to 60,000 in nominal value in 1991-1995.[11] If a student did not make any progress the support was cut, but there were special arrangements in the case of sickness, maternity leave etc., (Lånekassen, 2012).

The loans provided by NSLF were very favorable. Interest was not calculated while the student was enrolled, and repayments only started about ten months after graduation (or after dropping out). The interest rate was usually lower than

[7] Of the entry cohorts in years 1982-1995 roughly 2/3 students graduated with a degree within 10 years. Of these 25-30 percent completed a graduate degree, and less than one percent completed a doctorate degree (Statistics Norway, 2015).

[8] In 2011 the figure was 87 % (Kunnskapsdepartementet, 2012).

[9] Students in private institutions could apply for a loan from The Norwegian State Loan Fund to cover their tuition fees.

[10] Students would automatically receive both the loan and the grant, unless their earnings were too high.

[11] Corresponded to about USD 10,500 in July 2015.

the market interest rate. If a person could not repay her debt, for example because of illness or unemployment, the loan could be fully or partly cancelled.

The uptake of loan was high. Among the students who graduated with degree with a duration of more than five years in 1990, 97 % had some study loan, but only 28 % had taken up the full amount (Berg, 1997). In 1995 51 % of the 5th semester students had taken full loans, while 31 % had taken some loan and 18 % had not taken any loan or only the grant part for the support (Berg, 1997). Too high labor income, unwillingness to accumulate debt and living for free at home with parents were the most common explanations for not accessing full support (Berg, 1997).

Enrollment in graduate programs was low in the 1980s, and to stimulate enrollment in these programs, all students who graduated from a graduate program got a reduction of their loan after graduation, irrespective of time-to-degree. Until the academic year 1989/1990 this amount was fixed for all study programs (NOK 27,300 in 1989/1990), but from 1990/1991 it was differentiated by the duration of the program, ranging from NOK 28,400 for 10 semesters to NOK 43,400 for 13 semesters in 1990/1991. All study programs were affected similarly by this scheme, and it is not expected to influence the results.

1.3.3 The turbo reform

The *turbo reform* (turbostipendreformen) was made public on October 4th 1990 as a part of the National budget for 1991 and was motivated by a concern for students not exerting full effort in their studies. It entitled students in certain graduate study programs who completed their degree on stipulated time to a reduction of their student loan of about NOK 18,000 from the NSLF.[12] This corresponded to about 35 percent of the total student aid in one year, or 9 percent of the total study loan of a student who had followed normal study progression and taken up the full loan. The new rules applied to students who graduated after August 15th 1990, and thus a small number of students received the grant retrospectively.

Not all students were eligible for the restitution. Students in undergraduate programs were exempted, as well as students in certain fields of study. Delayed graduation was a widespread problem in higher education, and the reform was targeted specifically at programs where delays were common. These were mostly loosely structured study programs taught at universities, such as humanities, social sciences and natural sciences. The largest groups not covered by the

[12]This translated into USD 3,600 in July 2015.

reform were engineering and medical students. However, when the reform was first announced it was not clear which study programs were covered by the reform and the first official guidelines from the NSLF were not published until July 1991. Therefore, there was quite some uncertainty about the reform in the first year after implementation. There is no record of this reform being discussed in the media prior to the date it was announced, and therefore it is very unlikely that students could anticipate the reform.

The reform was debated from the start. The main arguments for the discontinuation of the turbo reform were that the rules were difficult to administer and that the restitution was likely to be given to students who would have graduated on time anyway, thus rewarding the good students and punishing the weaker students. Therefore, the grant was abolished, and students who graduated after August 14th 1995 were not eligible for the turbo reduction.

Even though the termination of the turbo scheme had been discussed by policy makers in the spring of 1994 (St. Meld. nr. 14, 3 94), students were unlikely to anticipate its ending. The changes were announced in the spring/summer 1995, but at that time the plan was to replace the turbo reform with a similar but more general scheme that would cover all students. Only later, in the fall of 1995, was it announced that the turbo grant would not be replaced after all.

1.3.4 Expectations from the reform

Before proceeding to the empirical analysis, it is useful to discuss what we can expect from the reform. The question is whether we would expect students to be willing (or able) to change their progression in response to the incentive offered through the turbo reform.

When answering this question, it is important to evaluate the effectiveness of the actual incentive. Should the turbo reduction of a students loan be considered a high powered incentive or not? The long time frame - students had to work hard for five to six years before receiving the reward - and the fact that the reward was a reduction of a loan rather than a cash reward might suggest that the incentive may not be very strong. There is, however, survey evidence suggesting that students at this time were reluctant to accumulating debt, which indicates that a loan reduction could be an attractive incentive for this group (Berg, 1994, 1997). In addition, the interest rate was high in this time period (10-12 percent) meaning that even a modest loan reduction would turn into a considerable amount of money saved given that the study loans were usually repayed over a long time

period (most commonly 20 years).

On the other hand, it is important to notice that students delay graduation for very different reasons, and that it might not be equally easy, or even desirable for the government, to counteract all these reasons. One motivation behind the reform was a concern for students not exerting full effort in their studies, and the hope was that the reform would induce students to spend more time studying. However, in a survey of graduates from graduate programs in 1990, Berg (1994) found that students delay graduation for a variety of reasons of quite different character. She also noted that a considerable share of students delayed graduation for reasons that were at least partly beyond their control. The single most important reason for delaying graduation was taking extra credits (42 percent reported this as a reason for delay), followed by work activities (26 percent), extensive curriculum (20 percent) and failing exams (19 percent).

The optimal response of a student depends on how costly it is to change her study pace. If the student is using a sub-optimal number of hours on studying to enjoy more leisure time, adjusting behavior is probably quite easy. Students who take extra courses can also easily graduate sooner by dropping courses. Graduating with fewer courses may not, however, be an attractive alternative if they perceive that the extra courses are needed to differentiate themselves in a competitive labor market. This concern was also expressed by Berg (1994) who noted that increased competition in the labor market due to rising unemployment rates and increasing supply of graduates led to hoarding of education in some programs in this period. If it was common to delay graduation in order to improve one's transcript of records, the loan reduction might not weigh up for the competitive advantage lost by not improving a grade or taking an extra course. Further, if sooner graduation meant entry into an unstable labor market, the extra reduction might not have been a very efficient incentive.

If the reason for delay was paid work, it is also not straightforward to anticipate what a student would do. A credit constrained student, for example, might find it impossible to cut down on working hours even when offered a loan reduction. The optimal response may also differ by the type of job a student holds. A student who holds a study related job, such as being a research assistant, might be less willing to reduce work hours than a student whose job is completely unrelated to her studies since there is less to gain in terms of complementarities and work experience.

An alternative response to taking fewer courses or working, could of course

be to spend less time on each course at the risk graduating with lower skills. While this is not an outcome intended by policy makers, it is not unrealistic if students perceive that taking extra courses and gaining work experience have high signalling value.

When students delay graduation because they struggle with extensive curricula, or fail exams, their possibilities of increasing their study pace are very limited even when they are encouraged to do so.

All in all, the turbo restitution was not necessarily a very high powered incentive, but it could still be valuable to students who were unwilling to accumulate debt. How well the restitution worked in practice, however, depended on how costly it was for students to change their behavior. The potential mechanisms of the reform are studied in Section 1.7.

1.4 Data

The study uses register data from Statistics Norway covering all students enrolled in higher education in 1974–2010. The data is reported directly from the educational institutions to Statistics Norway and is therefore considered to be very accurate. The data contains enrollment and graduation dates, completed degrees, institution from which the degree is obtained, duration of study program, as well as data on whether the student completed her degree on stipulated time, and if not, and the number of semesters delayed. The data also contains information on demographic characteristics, as well as parental education and income.

I focus on the students who completed a degree, although dropouts could be included to study dropout behavior.[13] I restrict the sample to students who were to expected graduate in 1986 or later. Students who enrolled in higher education in the fall 1991 or later are also excluded to avoid selection into treatment.[14]

To make the sample more homogenous, I restrict the sample to students who were aged 18-21 at high school graduation. Older students are less likely to rely on student aid and more likely to work and study part time, and therefore less likely to be affected by the incentives offered by the turbo reform. For this reason, I also restrict the sample to students who are aged 18-25 at first enrollment in

[13]Gunnes et al. (2013) find no effect of the reform on non-completion.

[14]The last cohort included in the study enrolled in the spring 1991, which means that they submitted their applications in the fall 1990. Thus, these students could have been aware of the turbo grant but since there was no information about the treatment status of the study programs at this point, it should not have affected the student's choice of study program.

higher education and no older than 40 years at graduation. Further, I truncate years of delay at the 1st and 99th percentile.[15]

Treatment status depends on two factors. First, not all study programs were covered by the reform. Eligibility was decided at the university-study program level, but in practice this often coincided with the study program level. The treated and non-treated study programs are listed in Table 1.1. Humanities, social sciences, science and law were the largest of the treated programs. Among those not treated, which I refer to as the control programs, were engineering, medicine and agriculture.[16] [17]

Second, eligibility depended crucially on the *expected* rather than the actual graduation date. Because delays were common, many students who graduated in the reform period had already passed their expected graduation date when the reform was implemented and were thereby not eligible for the restitution. Date of expected graduation is not recorded in the data, but I combine the date of first enrollment in higher education and the stipulated duration of the study program from which the degree was obtained (Column 3 in Table 1.1) to impute this date.

In in the treated group average delay was 3.85 semesters, or almost two years. This has important implications for the expected take-up of the reform. For many students who were approaching their stipulated graduation date when the reform was implemented, it was very hard to comply with the new rules even if they wanted to. Therefore, no big jump in the share of students graduating on time is expected at the time of implementation. Instead, I expect a gradual increase in this share for later cohorts who had more time to adapt their study habits and pace.

Put differently, students in different cohorts were treated at different intensities. A simple parametrization of treatment intensity is presented in Table 1.2 using information on the expected graduation date and duration of the study program. I define treatment as the number of years the student was studying in the reform period up to her expected graduation date. Given that the reform was implemented in 1990, I define students who were expected to graduate in

[15]Results are robust to truncation at the 5th and 95th, and 10th and 90th percentile.

[16]The degree obtained by the agriculture students is called *Cand.agric.*, which signals that is it related to agriculture and these students graduated from what is now called the Norwegian University of Life Sciences. In practice, the students had a number of different majors including engineering, business administration, resource management and biology. Thus, the agriculture students studied many of the same majors represented in the treatment group.

[17]The reform status of some study programs was unclear, most commonly because the status of the program changed during the reform period, or because NSLF could not determine the duration of the study program. Students in these programs are excluded from the sample.

1991 to be treated for one year, 1992 graduates to be treated for two years and so on. Students who were expected to graduate in the pre-reform period, and students who were expected to graduate in the reform period, but graduated before the reform was implemented are not treated. Students who were expected to graduate in the fall of 1990 were, strictly speaking, treated for two months, but their possibilities to comply with the reform were very limited and I treat these students as not treated. If there was a positive reform effect on these students, the estimates are downward biased. Students who were expected to graduate after the reform period ended were also partly treated, although not at the end of their studies. I include these students in the analysis as it is possible that these students changed their study habits early on in their studies and thus managed to graduate faster even in the absence of the reform.

The main goal of the turbo reform was to increase the share of students who graduated on stipulated time, and this is also the main outcome variable of the analysis. The outcome variable is a dummy variable indicating whether the student graduated on stipulated time or not and it is derived by Statistics Norway by combining data on the stipulated duration and the number of semesters a students was registered in higher education before graduation.

The turbo grant might also have had an impact on other dimensions of student behavior. In Appendix 1.C the analysis is repeated using delay measured in semesters as the outcome variable. Delay is also a policy relevant outcome because reductions in delay are associated with reductions in public spending on education both through student aid and through resources spent on teaching etc.

The share of students graduating on time by expected graduation year and treatment status is shown in Figure 1.1. The share who graduated on time was significantly lower in the treatment group than in the control group, which is expected since the turbo reform was targeted specifically at study programs where delays were common. As expected, there is no immediate jump in the outcome variables for the treated group straight after the implementation in 1990, but rather a gradual increase over the reform period. The treatment and control group follow the same pattern in the first two years of the reform period, but then the trends part as enough time has passed for students to adapt.

1.5 Empirical strategy

The fact that there are treated and non-treated study programs and a clear implementation date makes this a suitable application for a difference in difference estimator. There are ten treated study programs, and in the main analysis I estimate an average treatment effect controlling for study program fixed effects. Study program specific treatment effects are investigated in Section 1.6.2.

Just as there are several potential treatment groups, there are six non-treated study programs in the sample. Any combination of these can be used as a control group. The baseline results are estimated using an unweighted control group consisting of students in all of the six non-treated study programs. The choice of control group could, however, be motivated in many ways, and the robustness of the results is tested by using different control groups in Section 1.6.3.

The main specification is specified as follows:

$$y_i = \alpha + \beta treatment_i + s_i + c_i + \delta X_i + \epsilon_i \tag{1.1}$$

where $treatment_i$ measures treatment intensity (equal to zero in the control group and ranging from zero to five years in the treatment group). s_i and c_i are study program and expected graduation year dummies and X_i are control variables including demographic and family characteristics and ϵ_i is an error term. The coefficient β measures the effect of one additional year of treatment on the probability of timely graduation. Because of the short reform period most of the students were only partially treated. By multiplying β with the total study duration, it is possible to extrapolate the effect to a fully treated student.

When using difference in difference estimators unadjusted standard errors will often understate the true standard errors of the estimated coefficients due to the presence of unobserved group-level effects and/or serial correlation in the error term (Moulton, 1990; Wooldridge, 2003; Bertrand, Duflo, and Mullainathan, 2004; Donald and Lang, 2007). While there is consensus that the standard errors need to be adjusted when applying difference in difference estimators, there is less agreement on the best way to adjust them. In cases where there are many groups or clusters, the most straightforward approach is to cluster the standard errors at the group level.[18] When the number of clusters is small, however, clustering is not reliable (Bertrand et al., 2004).

[18]In practice this can be done by using the cluster option in STATA. This procedure allows for general within-group covariance and heteroscedasticity (Wooldridge, 2003; Donald and Lang, 2007).

Even though the group variable used in the analysis is study program, I cluster the standard errors at the university-study program level for three reasons.[19] First, this is the level at which treatment status is determined even though, in most cases, it coincides with the study program level. Second, one can easily argue that if there are common group effects or shocks, these are most likely to appear at the university-study program level. Consider social science students at two different universities; they study the same major, but the course structure of the program, the labor market and other factors that might affect study progression might differ between the universities (and cities). Third, by clustering at the university-study program level I increase the number of clusters from 16 study programs, to 56 university-study program clusters. This improves the reliability of clustering, as the method is only consistent if the number of clusters is large. The minimum number of clusters required to obtain reliable standard errors is often said to be 50 (Bertrand et al., 2004; Donald and Lang, 2007).

1.5.1 Validity of the difference in difference estimator

The validity of the difference in difference estimator relies on a number of assumptions. First, identification is threatened if students can manipulate their treatment status either through changing their expected graduation date or by switching between the treatment and the control group. The fact that the reform was retrospectively implemented is comforting because there was no way a student could manipulate her graduation date relative to the implementation date. It is also very unlikely that a student would move from the control group to the treatment group. When the reform was announced, all the students in the sample were already enrolled or in the process of enrolling in the study program they later graduated from. Changing their treatment status would mean enrolling in a different study program and starting over again, which would make them non-

[19] In Table 1.A.2 in the Appendix I compare different standard error corrections. Going from left to right the columns show unadjusted standard errors, standard errors clustered at the study program level, standard errors clustered independently at the study program year level following Cameron and Miller (2015) and standard errors clustered at the university study program level. From this table it is clear that the unadjusted standard errors are probably too small, but the table is not informative on which of the alternatives for clustering are preferable. In Columns 2 and 3, the estimated effects of the reform on delay are not statistically significant, while both estimates are significant at the 5 percent level in Column 4. In order not to overestimate the significance of the estimates, it is preferable to use a more restrictive standard error correction. But since simulation studies have shown that too few clusters can lead to overrejection I choose to cluster on the university-study program level, both because it is the logical choice and because the number of clusters is sufficiently large.

eligible for the restitution since it would cause them to spend too long obtaining their degree. Thus, I am confident that the selection into treatment is not an issue.

A potential concern is that even though students could not manipulate their treatment status directly, choice of study program (which in turn determines treatment status) is not random. But even in that case, the choice of study program was made before treatment status of the study programs was assigned. Therefore, the choice of study program should be uncorrelated with take-up of the reform.

The identifying assumption of the difference in difference estimator says that the reform effect can be estimated if the time trend of the outcome variable in the treatment and control group would have been the same, had it not been for the reform. The difference in difference estimator automatically deals with any differences in levels of the outcome variables. Therefore it is not a problem that the treatment and control group differ in their likelihood of graduating on time as long as their time trends are parallel.

The assumption of parallel trends is ultimately not testable, but there are some ways of assessing its plausibility. A first step is to graphically compare the pre-reform trends of the treatment and control group, which is done in Figure 1.1. The figure shows that the pre-reform trends are fairly parallel. The parallel trend assumption is studied further in Table 1.A.1 in Appendix 1.A, where the difference in pre-reform trends is estimated using both a linear time trend and year dummies. The assumption of parallel time trends in the pre-reform period cannot be rejected in either case.

The similarity of the treatment and control group is further investigated in Table 1.3 by performing a balancing test of pre-determined characteristics. As one could expect based on the fact that the treatment and control group students study different majors, there are some differences in background characteristics. The control group students are less likely to be female and have higher IQ score (available only for male students). The control group students also come from families with slightly higher income and parental educational attainment.

The last column of Table 1.3 reveals whether there are differential trends in the pre-determined characteristics that could explain the reform effect. The sample is balanced on parental education and ability score, but the share of female students increased at a significantly higher rate in the treatment group in the reform period. Increasing female educational attainment is a well-documented phenomenon over

this period, and it is therefore not surprising that the share of female students increased rapidly. However, some of the study programs in the control group (such as engineering) are and have always been very male dominated, and it is also well-known that the female share has not increased as much in these programs over time. The sample is also unbalanced on age at high school graduation. However, the difference is less than 0.04 years (14 days), which is unlikely to make a big difference in practice. Family income (measured at age 16) also increased slightly more in the treated group than in the control group, and the difference is significant at the 10 percent level. In the analysis, I control for pre-determined variables to ensure they are not driving the results.

1.6 Results

The main results are presented in this section. The baseline results using the unweighted control group are presented in Section 1.6.1. In Section 1.6.2, I study whether the reform effect differs by student characteristics such as gender, parental background and ability, as well as by study program. The robustness of the results are investigated in Section 1.6.3 where I use alternative control groups and test for other possible confounding factors such as the unemployment rate and increasing enrollment in higher education.

1.6.1 Probability of graduating on time

In Table 1.4, I first ignore the information that some students were treated for longer than others and estimate the average treatment effect for those who were expected to graduate in the reform period and for those who were expected to graduate in the post-reform period, respectively. The estimates in Column 1 suggest that students who were expected to graduate in the reform period were on average 3.7 percentage points more likely to graduate on time, while students who were expected to graduate in the year after the reform period ended were 4.7 percentage points more likely to graduate on time. However, the reform period estimate is only statistically significant when controls for student background characteristics are included in Column 2. The inclusion of control variables also makes the estimated coefficients slightly larger, and students who were expected to graduate in the reform and post-reform periods are now 4 and 5 percentage points more likely to graduate on time, respectively. Table 1.4 also suggests that female students and students with wealthier and more highly educated parents

(especially fathers) were less likely to graduate on time, while there were no significant differences between students of different age at high school graduation or between natives and immigrants.

The intensity of treatment is taken into account in Table 1.5, where I estimate effect of one additional year of treatment on the probability of graduating on time. In Column 1, control variables not included, and I find that one additional year of treatment increased the probability of graduating on time by about 1.5 percentage points. Given that only 16 percent of the students in the treated programs graduated on time in the pre-reform period, it must be seen as a relatively large effect. If the effect extrapolated to a student who was treated for six years (which was the most common duration in the treated group) the accumulated effect corresponds to a 9 percentage point increase in the probability of graduating on time. This again corresponds to a 58 percent increase in the probability of graduating on time. The inclusion of control variables in Column 2 does not significantly change the estimates, which suggests that selection into treatment and student characteristics are not driving the results. In the remainder of the paper, control variables will be included unless otherwise stated to improve precision.

In Figure 1.3, I re-estimate Equation 3.1 replacing the continuous treatment variable with dummies for each number of years treated to test for a non-linear treatment effect. However, I estimate a separate effect for students who were expected to graduate in 1996, who were treated for four or five years, but not in their last year of studies. The estimated effect is very close to linear (except for the group of students who were not treated in their last year). Confidence intervals at the 5 and 10 percent level are drawn in the figure and indicate that only the effect of being exposed for three or more years are significant, and only at the 10 percent level.

1.6.2 Heterogenous treatment effects

In this section, I exploit the data further to see whether the treatment effect varies with student characteristics, such as gender, family background and ability. I do this by interacting the treatment variable with the student characteristic of interest. Each of Columns 3 to 6 in Table 1.5 investigates differential treatment effects along one dimension of student characteristics.

From Table 1.4 we learned that female students were less likely to graduate on time than male students. However, the estimates in Column 3 of Table 1.5

suggest that the treatment effect was not significantly different for male and female students.

In Column 4, I test whether the treatment effect depends on family income. The students are divided into quartiles based on family income at age 16 and the quartiles are interacted with the treatment variable.[20] The treatment effect is not systematically correlated with family income. One interpretation of this is that wealthier parents do not to a larger extent than poorer parents provide financial support to their children so that they can focus more on their studies in response to the reform.

In Column 5, I find that the treatment effect increases with parental education and ranges from 0.4 and insignificant for students with parents who have less than high school education to 1.0 and 1.7 percentage points and significant at the 5 % level for students of parents with intermediate and higher education, respectively. From Column 4, we know that the positive effect of parental education on take-up works through other channels than income. Examples of such mechanisms could be attitudes or skills that are more often taught in homes of highly educated parents, such as motivation, determination or effort, or innate ability.

As a part of the military draft, all Norwegian males do a cognitive ability (IQ) test at age 18. I link the military test score data to the sample of male students to study the relationship between treatment and ability.[21] In Column 6, I divide the sample of male students into quartiles based on their cognitive test scores and interact them with the treatment variable. The estimates suggest that students in the highest quartile of the ability distribution are driving the treatment effect. One additional year of treatment increases the probability that a student in the top quartile of the ability distribution graduates on time by 3.2 percentage points relative to non-treated students and 2.1 percentage points relative to the students in the lowest quartile. It is not very surprising that the most able students are the ones that respond most strongly to the reform. The bright students who for some reason did not graduate on time in the absence of the reform are the most likely to have the capacity to graduate faster once faced with an incentive to do so.

[20] Family income measured at the year of enrollment in higher education gives the same results.
[21] See Black, Bütikofer, Devereux, and Salvanes (2013) for a description of these data.

Study program

In this section, I estimate study program specific treatment effects. A first motivation for doing so is simply to test whether there is evidence that students in different programs respond differently to changes in student aid policies. Second, I compare the program specific estimates to survey evidence on reasons for delaying graduation to try to learn about the mechanisms of the reform. This approach is very indirect and highly suggestive, but it may offer some indications for further investigation.

When estimating the study program specific treatment effects a few points are worth noting. First, when the data is split by study program, the precision of the estimates decreases, and since the number of clusters also decreases, the standard errors reported in Table 1.6 are not adjusted by clustering. Also, the unweighted control group does not work equally well for all treated study programs, partly because the small number of treated individuals observed in a given year causes the trend in the outcome variable to vary considerably from year to year (Figure 1.4). For these reasons the study program specific estimates should be interpreted with care, and I focus on the largest study programs for which the estimates are most precisely estimated. These programs are the humanities, social sciences, science, law and psychology.

Before turning to the results in Table 1.6, I summarize the findings of Berg (1994) from a survey on study progression. She found that students in the humanities were among those who took the most extra credits (59 percent reported delay for this reason) and worked the most (47 percent reported they worked so much that progression suffered and 33 percent worked without it affecting delay) and 25 percent reported that care responsibilities prevented them from following normal study progression. Science students were found to delay for very similar reasons as students in the humanities, although science students were even more likely to take extra courses (66 percent delayed graduation for this reason), in addition to failing exams and starting courses without finishing them. Social science students were found to delay because of work activities, care responsibilities and extensive curricula. Law students were different from the other students in that they failed exams to a larger extent (35 percent said it affected progression and 28 percent had failed without it affecting progression). In addition, law students often improved grades and struggled with extensive curricula. Psychology students reported that limited supply of courses and work activities affected their study progression, but not to a large extent.

Based on the findings of Berg, I can hypothesize about which study programs are most likely to respond to the reform. It seems as if students in humanities and science, of whom about 60 percent take too many course credits, could adapt quite easily to the new rules. On the other hand they also worked a lot, which might be harder to change in some cases. Social science students may have found it harder to comply than the former groups since they reported work and care responsibilities as main causes for delay. Of all the programs discussed above, however, law students seem the least likely to comply since they struggle with passing exams and extensive curricula.

The estimates in Table 1.6 are in line with the hypotheses stated above. Even though only the estimate for law students is statistically different from that for the other large programs, the ordering of the estimates is as expected. Students in science, humanities and social sciences responded most strongly to the reform with treatment effects ranging between 1.79-1.97 percentage points per year treated. Law students were significantly less likely to comply and the estimated effect of treatment is only 0.67 percentage points. Psychology students are the only ones that did not quite fit into the hypothesis as they responded quite well to the reform despite reporting work and limited supply of courses as main reasons for delay - reasons that may be hard to do something about.

While this exercise is interesting, it is not enough to draw any conclusions on the mechanisms of the reform. In order to be more conclusive, more data on variables related to delay is needed.[22] Since one in four students reported that working activities affected study progression, it is important to investigate this channel in more detail. This is done in Section 1.7.1.

1.6.3 Sensitivity analysis

In this section, I investigate the robustness of the results. First, I test whether the results are robust to changes in the control group. Later, I discuss how the economic conditions in Norway might have had an impact on the outcomes.

Alternative control groups

As described in Section 1.3.3, the turbo reform was targeted at study programs where delays were common. This, together with the fact that the treated and the non-treated study programs are different in terms of the level of the outcome

[22] In future work, I will link the sample to course data to investigate whether the reform worked through taking fewer courses.

variables (Table 1.1), characteristics of the students (Table 1.3) and in terms of their structure raises potential concerns about the validity of the choice of control group. In this section, I test the sensitivity of the estimates to changes in the control group.

Ideally, I would like the treated and control group students to study the same majors. Although that is not feasible, it is possible to make the treatment and control group more similar by restricting the analysis to only the programs that are the most similar. Most of the treated study programs were loosely structured programs with little organized teaching and a large degree of freedom of choice (and responsibility for one's progression) for the students, while the (health related) non-treated programs were tightly structured programs with classroom-like teaching. The treated programs had no or very low admission criteria, while competition for study places was fierce in most of the non-treated programs. The labor market prospects of the two groups also differed. While the students in the treated programs were likely to be employed in both the private and the public sectors, the majority of the students in the health related non-treated programs were likely to be employed in the public sector. In general, students in treated programs faced more volatile labor markets and were more sensitive to fluctuations in the unemployment rate. All of these factors might in turn affect study progression of the two groups differently.

Following these lines of argument, I exclude medicine, veterinary medicine and pharmaceutical science from the control group. Of the remaining three programs, I also exclude architecture because it had some features which makes it hard to compare to the other programs. That leaves engineering and agriculture in the control group. Studies in engineering are more tightly structured than some of the treated programs, but all in all, the programs are relatively similar. The engineering students also face very similar labor market conditions as the treated students in terms of sensitivity for recessions and unemployment rates. The same holds for the agriculture students (see also footnote 16).

In Column 2 of Table 1.7, I estimate the reform effect using the control group consisting only of engineering and agriculture students. Compared to the baseline estimates in Column 1, the estimated reform effect is now marginally smaller and only significant at the 15 percent level.[23] The R-squared is slightly higher, suggesting that the control group with only engineering and agriculture students might be a better fit than the baseline control group although precision is lower.

[23]Note, however, that the number of clusters decrease when I drop programs from the control group, and this may lead to overrejection.

Another way to make the treatment and control group more similar is by using propensity score matching. By estimating a propensity score and restricting the sample to the region of common support, I can make the sample more homogenous in terms of student characteristics. To be more specific, I estimate the probability of being in the treatment group based on a linear function of all the predetermined variables (except ability) in Table 1.3, and restrict the sample to the region of the probability distribution that is covered by both the treated and the non-treated group. In this way, I lose 1899 individuals (5.5 percent of the sample), who are proportionally distributed across study programs (except a slightly smaller share of engineering students compared to the full sample).

In Column 3, I restrict the sample to the region of common support and estimate the baseline specification. This gives me an estimate that is very close to that in the baseline specification, and of similar precision. In Column 4, I weight the sample by the propensity score. In this case the estimated effect is 0.0147 and significant at the 5 percent level.

Control groups are often chosen based on economic reasoning, but Abadie et al. (2010, 2012) have developed a data driven method for constructing a control group. The synthetic control method was originally developed for case studies where there is one treated unit[24] and a number of possible control units, but where none of the control units serve as good control group either on their own or in an unweighted combination. The idea of Abadie and co-authors is to create a synthetic control group based on a weighted combination of the possible control units based on matching of pre-intervention values of the outcome variable and/or pre-intervention values of predictors of the outcome variable.

The matching procedure is explained in more detail in Appendix 1.B. In short, I use the study program specific share of students graduating on time in each of the years 1986-1989 as predictor variables. Because the matching algorithm matches on levels rather than trends, and because of the large differences in levels between the treated and the control units, I normalize the outcome variable to have mean zero in 1990. The result of the matching is presented in Figure 1.2 where the trends of the unweighted and the synthetic control group are compared.

The difference in difference estimates using the synthetic control group are presented in Column 5 of Table 1.7. The estimated reform effects are smaller than the baseline specification in Column 1 predicts: one additional year of treatment is predicted to increase the probability of graduating on time by 1.33 percentage

[24]In this case there is of course more than one treated unit but I pool all the treated programs into one unit.

points and the estimate is only statistically significant on the 11 % level.

As mentioned earlier, Gunnes et al. (2013) found large and significant effect of the same reform on average delay and the probability of graduating on time. More specifically they estimated that one additional year of treatment increased the probability of graduating on time by 2 percentage points using a slightly different estimating equation and a different sample. One of the main differences between this study and that of Gunnes et al. is that they do not include engineering in the control group. In Column 6, I exclude engineering from the control group and find a highly significant reform effect of 2.2 percentage points. This is quite close to what Gunnes et al. found, and suggests that much of the difference between the estimates is due to our treatment of engineering students.

Taken together, the results are robust to changes in the control group, although some of the estimates are imprecise. All estimates are in the same ball park and suggest that the one additional year of treatment increased the probability of graduating on time by 1.3-1.5 percentage points.[25]

Unemployment and increasing enrollment

Despite seemingly parallel trends in the outcome variables in the pre-reform period, the validity of the difference in difference estimator can be threatened if there is some confounding factor that coincided with the implementation of the reform and that had a different impact on the treatment and control group over time. In this section, I study the presence of such factors.

The most serious potential confounder is the recession that hit Norway in the late 1980s and that led to increasing unemployment rates in 1987-1993. This was likely to affect the employment possibilities of recent graduates and students looking for part time jobs in the reform period. As pointed out above, however, increasing unemployment rates is only a threat to identification if the treated and non-treated students are affected differently. This can not be assumed without further ado for a number of reasons discussed below.

In Section 1.6.3, I argued that the treated students probably faced more volatile labor markets than the control students. Linking this to the literature that suggests that there are severe and persistent negative effects on labor market outcomes from graduating in a recession (Kahn, 2010; Oreopoulos, von Wachter, and Heisz, 2012; Liu, Salvanes, and Sørensen, 2012) treated students

[25]Based on the tests discussed in Section 1.5.1, I cannot reject the parallel trend assumption using these alternative control groups. Results are not presented here for brevity, but are available on request.

could have stronger incentives to strategically delay graduation in bad times. If treated students strategically delayed graduation in response to the increasing unemployment rates to a larger extent than the students in the control group, the reform effect would be underestimated.

Häkkinen and Uusitalo (2003) offer an alternative hypothesis relating unemployment and graduation behavior. They found that higher unemployment rates in the 1990s in Finland lead to fewer student jobs, forcing students to study full-time and to graduate faster despite the bad economy.

Working while studying is and has historically been very common in Norway. In a student survey conducted in the mid 1990s, about 60 percent of the respondents agreed at least to some extent with the claim that they had to work to manage financially (Berg, 1997), and, in another study, 80 percent of the respondents answered they had worked during their studies and 26 percent reported that work activities had affected their study progression (Berg, 1994). About 59 percent of the students in the sample of this study worked at some point while enrolled in higher education, and working was more common among treated students (66 vs 49 percent). If there were fewer students jobs available during the recession, this probably hit the treated students harder, which would lead to an overestimation of the reform effect.

In Column 2 in Table 1.8, I control for unemployment by interacting the study program dummies with the national unemployment rate in the year of expected graduation for individuals aged 25-54.[26] The interaction terms (not reported for brevity) are mostly small and insignificant, but there seem to be some differences between study programs. The estimated reform effect is 1.4 percentage points and significant at the 10.2 % level, suggesting that if anything, not controlling for unemployment rates actually lead to overestimating the reform effect slightly, which is in line with what Häkkinen and Uusitalo (2003) found. This issue is further discussed in Section 1.7.1 where I discuss whether the work behavior of treated students changed.

Youth unemployment increased rapidly during the recession and in order to keep youth unemployment rates at acceptable levels the government increased the number of study places at the universities. This led to an enormous increase in the number of students enrolled in higher education between 1987-1994.[27] This

[26]Adding the unemployment rate one and two years before expected graduation had no big impact on the estimates.

[27]Between 1986 and 1994 the number of study places increased from 101,000 to 169,000 mostly as ad hoc solutions to the unemployment problem. Despite wishes that the new places would mainly be created in the the college sector where most study programs were short and

expansion was mainly concentrated to the treated programs simply because it was easiest to expand capacity in the loosely structured study programs (Try, 2000). A concern might be that this expansion affected study progression, but it is not clear in what direction. On the one hand, the increased number of students could lead to increases in delay if there was congestion in the education system due to failure to provide resources necessary to meet the needs of the bigger cohorts. Furthermore, as more students enrolled in the universities, the composition of the student mass probably also changed. In particular, average quality of students in terms of ability, study preparedness and motivation might have decreased, which could impact academic performance and progression negatively. On the other hand, increased competition for jobs between the students might have put more pressure on the students to graduate faster to signal ability and motivation to future employers.

In Columns 3 to 5 in Table 1.8, I control for cohort size. In Column 3, I include a control for the logarithm of the total graduation cohort size in the year of expected graduation, which I interact with study program to allow students in different programs to differ in their sensitivity to competition. In this specification, students in different programs are assumed to be substitutes. In Column 4, I instead include a control for the logarithm of study program specific cohort size. In this specification, students only face competition from their peers in the same study program. Both variables are included in Column 5.

Once cohort size is taken into account, one additional year of treatment is estimated to increase the probability of graduating on time by 1.1-1.4 percentage points, which is lower than the baseline estimate but must still be considered a non-negligible effect. This is in line with the idea that students increase their study progression in response not only to the reform, but also in response to the increased competition.

1.7 Mechanisms

I Section 1.3.4, I discussed potential responses to the reform, and this discussion also gave some indications of underlying mechanisms through which the reform could have worked. In this section, I discuss and investigate some potential mechanisms.

vocationally oriented, the university sector grew faster than the college sector in this period (Try, 2000).

1.7.1 Working while studying

Since the majority of students work at some point during their studies and since about one in four reported that this affected their study progression, it is interesting to study whether the reform had an impact on the working activities of the students. In Table 1.9, I study whether the reform affected both the probability of having a job and the annual earnings of students.

I use two definitions of having a job. The first is based on job spell data, but many missing stop dates adds some uncertainty to the variable. The second definition is based on earnings, and it is a dummy that takes on the value one if the student had positive earnings in a given year. However, the data on pension qualifying earnings also include other types of income than labor income, and therefore the share of students with some earnings is higher than the share of students who actually work.

I estimate the effect of the reform both on total pension qualifying earnings and on the probability of having pension qualifying earnings in certain intervals.[28]

The variables are measured in every calender year from enrollment up to the year prior to expected graduation. In the expected graduation year, many of the students graduate and enter the labor market which of course affects earnings in that year.

I use a difference in difference strategy, but the year fixed effects now refer to the year when the work activity or earnings were recorded. The difference in difference variable of interest measures the average effect over the reform period. Study program dummies and background characteristics are included as before. I also control for the national unemployment rate (interacted with study program) since changes in the unemployment rate might affect the employment possibilities of students.

The effect on the probability of having a job is investigated in Columns 1 and 2 of Table 1.9. Both estimates are small and insignificant, and there is no evidence that the reform affected the probability of working. In Column 3, I estimate the effect on total earnings and find a small negative but statistically insignificant effect on total earnings.

In Table 1.10, I estimate the reform effect on having earnings in certain earnings intervals. The average annual earnings in the sample was NOK 36,000 in 1990 value and the median was NOK 26,400. The results indicate that there was

[28]The earnings intervals are 1-5,000, 5,001-10,000, 10,001-20,000, 20,001-30,000, 30,001-40,000 and more than NOK 40,000 in 1990 value.

no effect of the reform on the probability of earning more than NOK 40,000.[29] This is natural since students with high earnings were unlikely to rely on student aid in the first place and thereby had no incentive to work less and graduate sooner. Among the lower earning students, on the other hand, there are significant changes. The probability of earning NOK 20-30,000 decreased by 2.15 percentage points while the probability of earnings less, but still positive sums increased significantly. In other words, the results suggest that students reduced working hours in the reform period.

One potential mechanism that would be important to study is the one that goes through quality. A potential response of students who cannot reduce working hours or other time consuming activities is to reduce the time spent studying. This could lead to the student graduating with lower skills, which is not a desirable outcome. A direct way of studying this mechanism would be to estimate the impact on grades, but these data are not available. Another, less direct, way which was also suggested by Gunnes et al. (2013) is to study the impact on earnings after graduation. If students graduated with lower quality, one could expect a negative impact on earnings for the treated students. Such an analysis is challenging in practice, since the treated and control students followed different earnings trajectories even before the reform was implemented. Thus, the difference in difference estimator is not valid in this case.[30]

1.8 Conclusion

In many industrialized countries there is a strong political will to incentivize students to graduate faster because of high costs and other inefficiencies related to students spending excess time obtaining their degrees. Still, the empirical evidence of how students respond to financial incentives is scattered. A particular issue that the existing literature has failed to acknowledge is that students delay graduation for a variety of reasons. This has three implications. First, not all students want to increase their study pace. Two examples of this could be a students who works part time in jobs that complement their studies, or a student who feels pressure to graduate with good grades or extra courses to be able to succeed in a competitive labor market. Second, not all students can impact their progression. Sometimes the students' possibilities of action are limited by structural issues, such as limited

[29] As noted earlier, students earning more than NOK 5,200 a month were not eligible for the student loan, and were thus not expected to respond to the reform.

[30] Results are available on request.

supply of courses, poor supervision or extensive curricula. Third, the are instances when compliance is not desirable either from the student's or from the societal perspective. Again, an example of this could be if a student works in a study related job, such as being a research assistant, or when a student spends less time on her studies to increase progression, leading to lower skills at graduation.

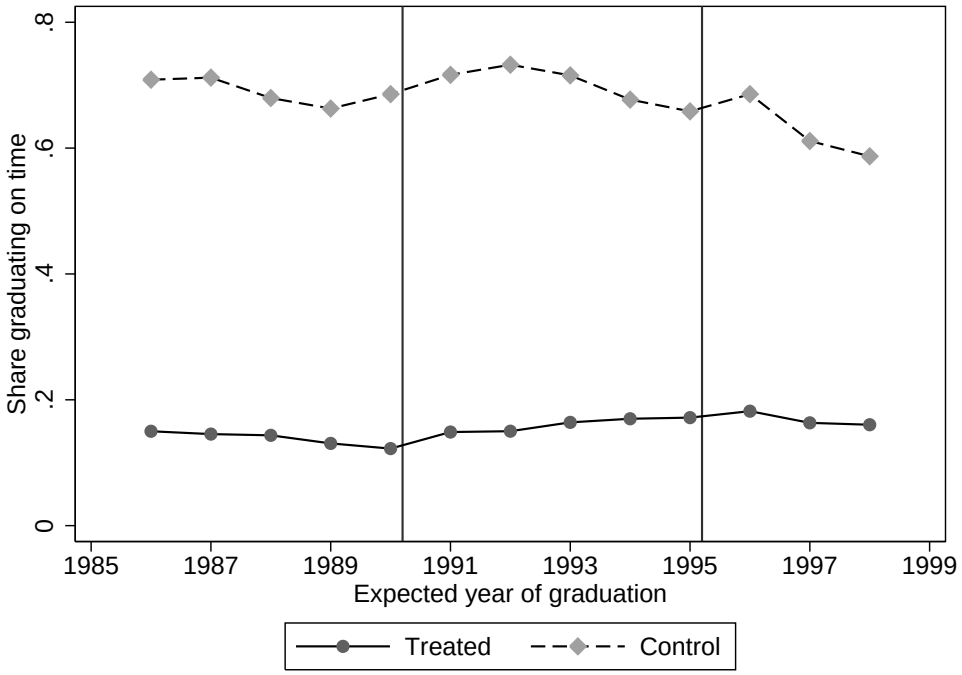
This paper evaluates one of the earliest policies to target delay; the turbo grant reform that was implemented in the 1990s in Norway. Students who graduated from certain study programs on stipulated time were entitled to a reduction of their student loan of NOK 18,000, which corresponded to about 9 percent of the total loan of a student who had taken up the full amount of loan.

I find significant effects of the reform both on the share of students who graduate on stipulated time and on average delay and the results are significant both in statistical and economic terms. But even so, the share of treated students who graduated on time remained well below 30 percent.

The data offers limited possibilities to study mechanisms of the reform, but it is possible to speculate about possible mechanisms of the reform based on survey evidence on why these students delay graduation in the first place (Berg, 1994). She reported that students vary in their motives for delaying graduation, and that the reason for delay is sometimes outside of the control of the students. Further, she found that the reasons for delaying graduation were often correlated with study program, which in turn indicated that structural differences between study programs might explain differences in take-up. Initial analysis suggests that the reform, at least in part, worked through reducing working hours.

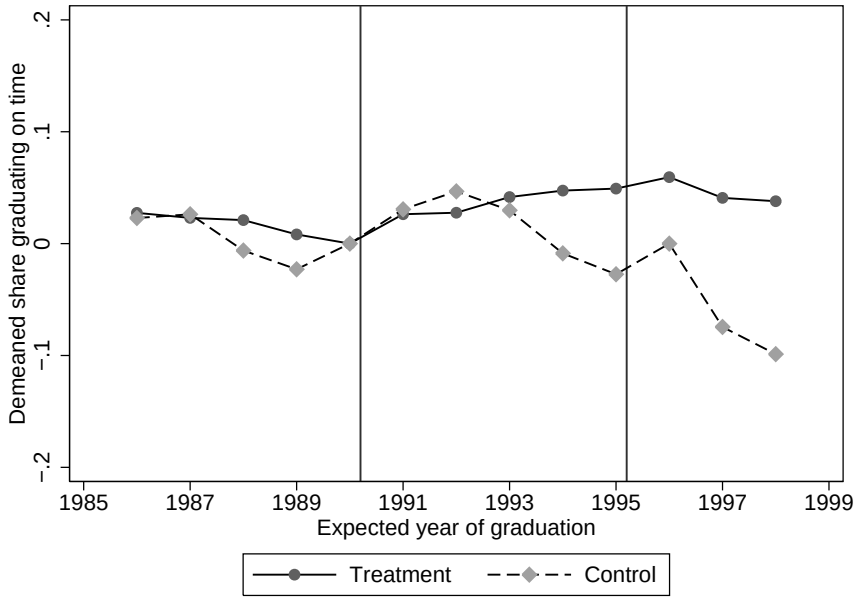
These results suggest that when designing financial incentives for students, it is useful to acknowledge that students delay graduation for a variety of reasons and that this affects the degree to which they are able and willing to respond to incentives. If the aim is to reduce delay, other interventions, such as restructuring of study programs or improved supervision, might be needed as well.

Figure 1.1: Share of students graduating on time by treatment status

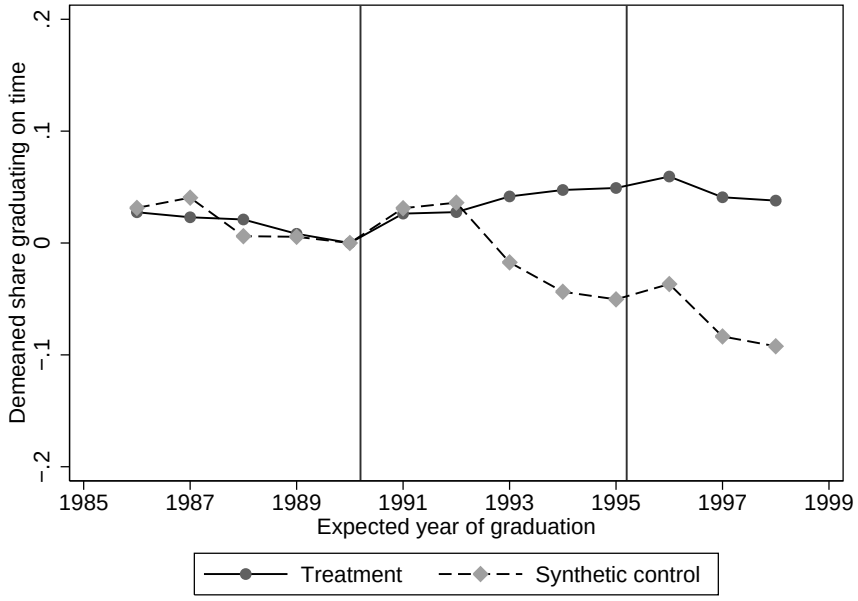


Note: The treated group contains individuals enrolled in humanities, social sciences, science, law, arts, theology, business administration, psychology, dentistry and fishery. The control group consists of individuals enrolled in medicine, agriculture, engineering, pharmaceutical science, veterinary medicine and architecture. The vertical lines refer to the implementation and termination date of the reform, respectively.

Figure 1.2: Demeaned share of students graduating on time by treatment status



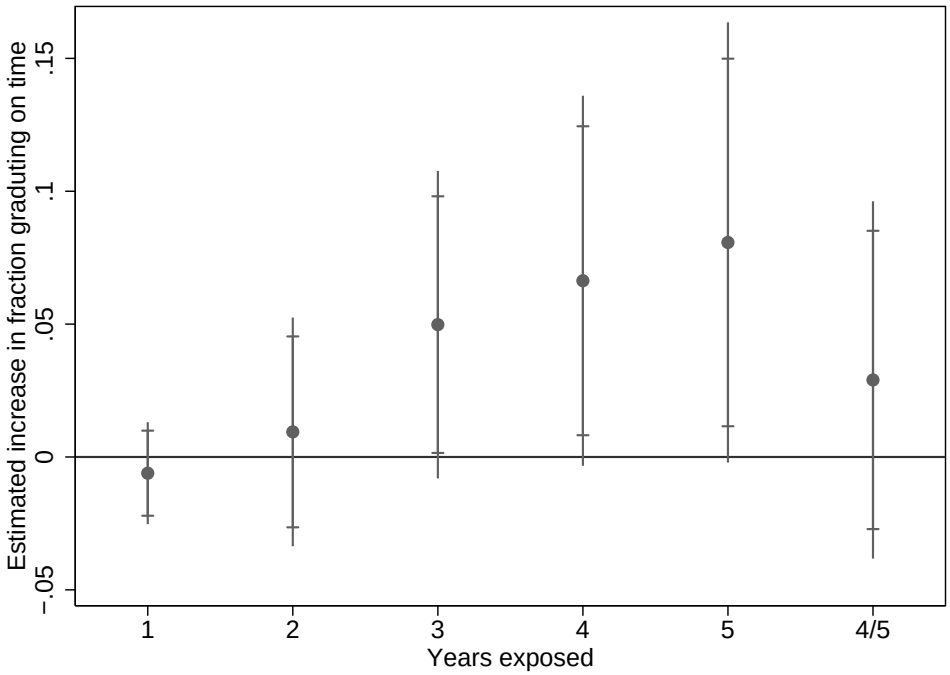
(a) Unweighted control group



(b) Synthetic control group

Note: Trends are demeaned by subtracting the group specific average of the outcome variable in 1990 from the group average in each year. The treated group contains individuals enrolled in humanities, social sciences, science, law, arts, theology, business administration, psychology, dentistry and fishery. The control group in (a) as defined in Figure 1.1 and the control group in (b) as described in Appendix 1.B. The vertical lines refer to the implementation and termination date of the reform, respectively.

Figure 1.3: Testing for non-linear treatment effects



Note: Estimates are obtained by regressing the outcome variable on dummies for years treated (see Section 1.5) to test for non-linear treatment effects. The last estimate corresponds to individuals that were expected to graduate in 1996, who were either treated for four of five years. 90 and 95 percent level confidence intervals included.

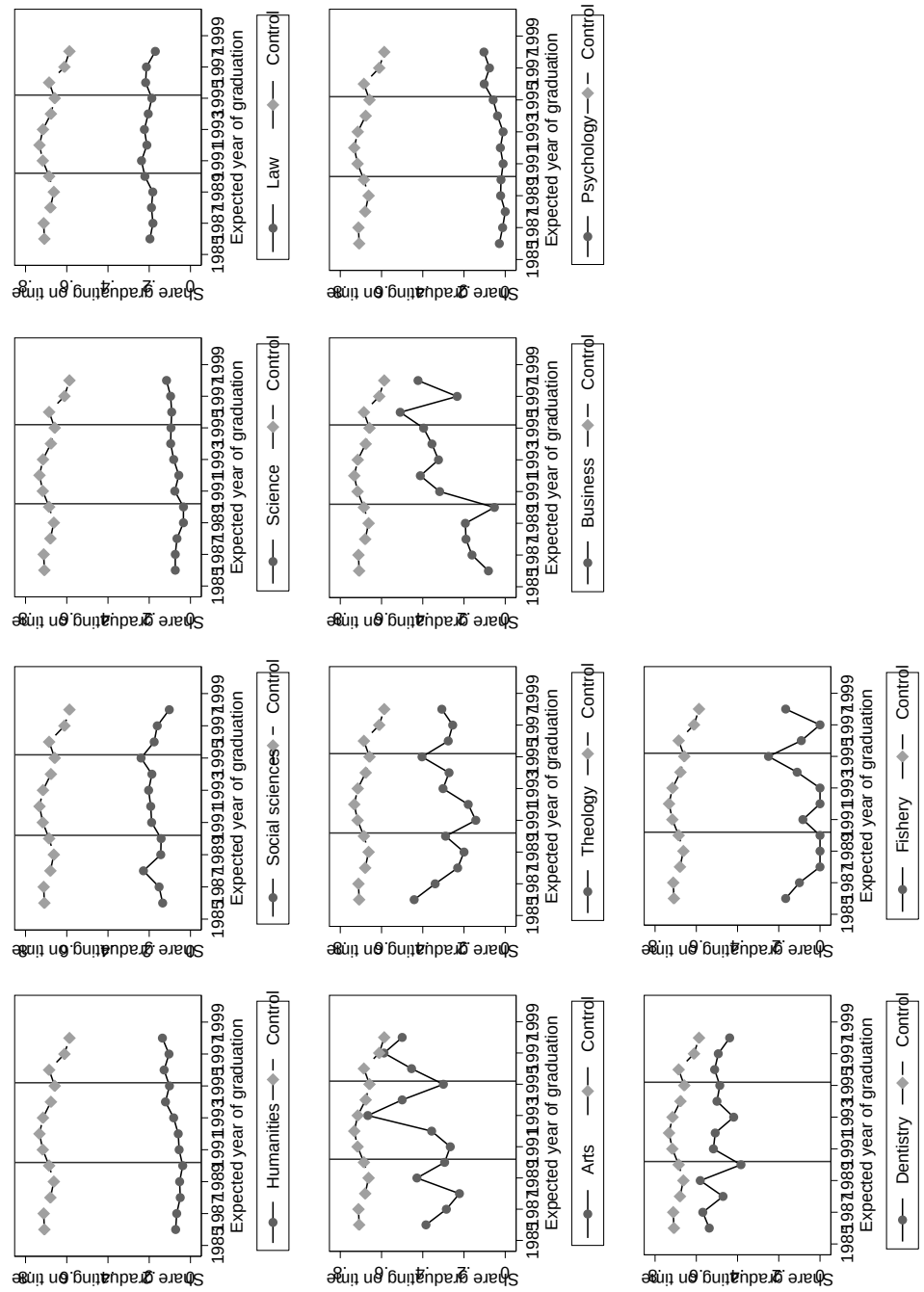


Figure 1.4: Study program specific share of students graduating on time

Note: The control group consists of individuals enrolled in medicine, agriculture, engineering, pharmaceutical science, veterinary medicine and architecture. The vertical lines refer to the implementation and termination date of the reform, respectively.

Table 1.1: Sample distribution by study program

Treated	No of students	Percentage	Expected duration	Delay in semesters	Percentage on time
Humanities	2,829	8.27	6	4.93 (3.57)	8.52 (27.92)
Social Sciences	3,212	9.39	6	3.95 (3.82)	18.57 (38.91)
Science	4,580	13.38	5	4.49 (3.08)	7.03 (25.57)
Law	5,188	15.16	6	2.98 (3.17)	20.51 (40.38)
Arts	158	0.46	6	1.96 (2.50)	37.34 (48.52)
Theology	481	1.41	6	2.17 (2.96)	28.90 (45.38)
Business administration	520	1.52	5.5/6	2.31 (3.03)	29.04 (45.44)
Psychology	1,129	3.30	6.5	5.52 (3.03)	3.45 (18.27)
Dentistry	623	1.82	5	1.16 (1.66)	49.60 (50.04)
Fishery	87	0.25	5	3.82 (2.55)	5.75 (23.41)
	18,807	55.0		3.85 (3.43)	15.55 (36.25)
Control					
Medicine	2,874	8.40	6	1.63 (1.97)	26.06 (43.90)
Agriculture	1,974	5.77	5	-0.29 (2.02)	80.50 (39.63)
Engineering	9,419	27.52	5	-0.40 (2.05)	81.75 (38.63)
Pharmaceutical Science	303	0.89	5	-0.52 (2.05)	76.57 (42.43)
Veterinary Medicine	368	1.08	5.5	0.11 (1.22)	80.98 (39.30)
Architecture	475	1.39	5.5	3.37 (2.71)	5.05 (21.93)
	15,413	45.0		0.11 (2.26)	68.72 (46.36)
N	34,220	100		2.17 (3.50)	39.50 (48.89)

Notes: Standard deviations of delay and share on time in parenthesis.

Table 1.2: Parametrization of treatment intensity

	Stipulated duration of study in years			
Expected graduation year	5	5.5	6	6.5
1986-1990	0	0	0	0
1991	0/1	1	1	1
1992	1/2	2	2	2
1993	2/3	3	3	3
1994	3/4	4	4	4
1995	4/5	5	5	5
1996	4	4/5	5	5

Notes: The intensity of treatment is determined by expected graduation year and by the duration of the study program. Science students (5 year program) became eligible in 1991, and are thereby treated one year less than other students.

Table 1.3: Means of predetermined variables by treatment status

	Treated			Control			Difference-in-Difference
	Pre-reform (1)	Post-reform (2)	Difference (2)-(1)	Pre-reform (3)	Post-reform (4)	Difference (4)-(3)	[(2)-(1)] -[(4)-(3)]
Female	.4638	.5343	0.0705***	.3367	.3391	0.0024	.0681***
	[.006]	[.005]	[.008]	[.006]	[.005]	[.008]	[.011]
High school	19.093	19.074	-0.0180***	19.043	19.062	0.0196***	-.0376***
graduation age	[.006]	[.004]	[.007]	[.005]	[.004]	[.007]	[.010]
Mother's years	11.36	11.67	0.318***	11.43	11.77	0.332***	-.0143
of education	[.038]	[.028]	[.047]	[.037]	[.033]	[.049]	[.068]
Father's years	13.06	13.15	0.098*	13.12	13.29	0.168***	-.0700
of education	[.047]	[.034]	[.058]	[.047]	[.040]	[.062]	[.085]
Family income	523543	567601	44058***	531264	565996	34622***	9436*
at age 16	[2894]	[2366]	[3844]	[2804]	[2629]	[3874]	[5482]
Ability (males)	7.041	6.901	-0.139***	7.616	7.502	-0.114***	-0.0249
	[.025]	[.020]	[.032]	[.020]	[.018]	[.027]	[.042]

Notes: Means and differences in means of predetermined variables. Ability score available only for male students. Standard errors in square brackets. $*p < 0.10, **p < 0.05, ***p < 0.01$.

Table 1.4: Average effect over reform and post-reform period

	(1)	(2)	(3)	(4)
Outcome variable:	On time	On time	Delay	Delay
Reform period*Treated	0.0370	0.0397*	-0.370*	-0.391*
	(0.0231)	(0.0224)	(0.210)	(0.203)
Post-reform period*Treated	0.0468*	0.0501*	-0.635***	-0.654***
	(0.0275)	(0.0268)	(0.210)	(0.208)
Female		-0.0379***		0.351***
		(0.00552)		(0.0549)
High school graduation age		-0.00278		0.108
		(0.00535)		(0.0960)
Mother's Education Middle		0.00195		-0.0346
		(0.00938)		(0.0872)
Mother's Education High		-0.0165		0.0887
		(0.0136)		(0.132)
Father's Education Middle		-0.00964		-0.0443
		(0.00642)		(0.0593)
Father's Education High		-0.0254**		0.0872
		(0.0112)		(0.0999)
log Family income at age 16		-0.0123***		0.0733*
		(0.00332)		(0.0435)
Immigrant status		0.00964		-0.0412
		(0.0114)		(0.113)
Constant	0.282***	0.523***	1.690***	-1.416
	(0.0238)	(0.106)	(0.104)	(1.588)
R^2	0.427	0.431	0.371	0.377
Observations	34220	34220	34220	34220

Notes: Difference in difference estimates of the effect of expected graduation in the reform and post-reform period. All specifications include study program and cohort fixed effects. Columns 2 and 4 also contain dummies for region of residence at age 16 and unknown parental education. Parental education relative to low education, where parental education is defined as low=less than high school, middle=high school, high=tertiary education. Standard errors in parentheses clustered at study program university level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: The effect of the turbo reform on the probability of graduating on time

	(1)	(2)	(3)	(4)	(5)	(6)
	Raw	Baseline	Female	Family income	Parental education	IQ
Years treated	0.0148*	0.0153**	0.0140*	0.0165*	0.00374	0.0108
	(0.00777)	(0.00760)	(0.00791)	(0.00830)	(0.00970)	(0.0138)
Female		-0.0379***	-0.0411***	-0.0377***	-0.0376***	
		(0.00551)	(0.00698)	(0.00551)	(0.00550)	
Female*Years treated			0.00248			
			(0.00268)			
Family income 2nd quartile				0.000152		
				(0.00559)		
Family income 3rd quartile				-0.00388		
				(0.00667)		
Family income 4th quartile				-0.0359***		
				(0.00712)		
Family inc 2nd q*Years treated				-0.00567*		
				(0.00324)		
Family inc 3rd q*Years treated				-0.00287		
				(0.00393)		
Family inc 4th q*Years treated				0.00236		
				(0.00489)		
Medium parental education					-0.0101	
					(0.00834)	
High parental education					-0.0384**	
					(0.0175)	
Medium par educ*Years treated					0.00973**	
					(0.00406)	
High par educ*Years treated					0.0133**	
					(0.00602)	
Ability 2nd quartile						-0.000954
						(0.0187)
Ability 3rd quartile						0.00125
						(0.0260)
Ability 4th quartile						-0.0205
						(0.0222)
Ability 2nd q*Years treated						0.00319
						(0.00599)
Ability 3rd q*Years treated						0.00965
						(0.00710)
Ability 4th q*Years treated						0.0209***
						(0.00739)
Constant	0.285***	0.519***	0.522***	0.374***	0.531***	0.611***
	(0.0244)	(0.106)	(0.106)	(0.0915)	(0.102)	(0.136)
R^2	0.427	0.431	0.432	0.432	0.431	0.442
Observations	34220	34220	34220	34220	34220	18723

Notes: Difference in difference estimates of the effect of one additional year of treatment on the probability of graduating on time. In Column 1 Eq. 3.1 is estimated without control variables, and Columns 2 to 6 are estimated including controls for gender, age at high school graduation, region of residence, immigrant status, parental education and log family income at age 16. All specifications include study program and cohort fixed effects. In Column 3 the treatment variable is interacted with gender and in Column 4 dummies for family income at age 16 quartiles are interacted with treatment. In column 5 treatment is interacted with parental education dummies, where high education means that at least one parent has higher education, intermediate education means that at least one parent has a high school degree, but no more and low education means that parents have not finished high school. In Column 6 the treatment variable is interacted with dummies for ability quartile restricting the sample to male students. Standard errors that are clustered at study program university level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.6: The estimated effect on the probability of graduating on time by study program

	(1) Humanities	(2) Social Sciences	(3) Science	(4) Law	(5) Arts
Years treated	0.0185***	0.0179***	0.0197***	0.00673**	0.0290*
	(0.00402)	(0.00388)	(0.00401)	(0.00328)	(0.0150)
Constant	0.645***	0.480***	0.555***	0.680***	0.572***
	(0.144)	(0.149)	(0.136)	(0.142)	(0.166)
R^2	0.413	0.362	0.461	0.367	0.287
Observations	18242	18625	19993	20601	15571

	(6) Theology	(7) Business	(8) Psychology	(9) Dentistry	(10) Fishery
Years treated	0.0148	0.0664***	0.0150***	-0.00704	0.0361
	(0.00908)	(0.00923)	(0.00578)	(0.0110)	(0.0319)
Constant	0.584***	0.639***	0.596***	0.580***	0.621***
	(0.164)	(0.164)	(0.155)	(0.164)	(0.165)
R^2	0.294	0.296	0.368	0.278	0.293
Observations	15894	15933	16542	16036	15500

Notes: Study program specific difference in difference estimates of the effect of one additional year of treatment on the probability of graduating on time using the unweighted control group. Control variables included, see Notes in Table 1.5. Standard errors (not clustered because of small number of clusters) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: The effect on the probability of graduating on time using alternative control groups

	(1) Baseline	(2) Engineering+ Agriculture	(3) P-score com. sup.	(4) P-score weighted	(5) Synthetic control	(6) Excluding engineering
Years treated	0.0153**	0.0147	0.0155**	0.0147**	0.0133	0.0220***
	(0.00760)	(0.0102)	(0.00751)	(0.00704)	(0.00820)	(0.00619)
Constant	0.519***	0.949***	0.581***	0.582***	0.444***	0.583***
	(0.106)	(0.0962)	(0.107)	(0.109)	(0.112)	(0.109)
R^2	0.431	0.454	0.434	0.430	0.372	0.269
Observations	34220	30200	32321	32324	33074	24801

Notes: Difference in difference estimates of the effect of one additional year of treatment on the probability of graduating on time. All specifications include study program and cohort fixed effects and control variables as described in Table 1.5. Standard errors that are clustered at study program university level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.8: Sensitivity checks: the effect of unemployment and cohort size

	(1) Baseline	(2) Unemployment	(3) Supply 1	(4) Supply 2	(5) Supply 1+2
Years treated	0.0153**	0.0140	0.0140	0.0109*	0.0124*
	(0.00760)	(0.00844)	(0.00863)	(0.00582)	(0.00722)
Constant	0.519***	0.546***	0.447*	0.536***	0.607***
	(0.106)	(0.0983)	(0.226)	(0.0978)	(0.226)
R^2	0.431	0.432	0.431	0.432	0.432
Observations	34220	34220	34220	34220	34220

Notes: Difference in difference estimates of the effect of one additional year of treatment on the probability of graduating on time. All specifications include study program and cohort fixed effects and control variables as described in Table 1.5. Unemployment rate is national average unemployment rate for ages 25-54 measured at expected year of graduation. Cohort size refers to the number of students in the sample expected to graduate in a given year. In Column 3, study program specific cohort size is included and in Column 4 the total cohort size is included. In Column 5 both cohort measures are simultaneously included. Standard errors that are clustered at study program university level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.9: Reform effect on the probability of working and earnings while studying

	(1) P(Work)	(2) P(Earn>0)	(3) log(Earnings)
Reform period*Treated	0.0241	-0.00435	-0.0315
	(0.0187)	(0.00413)	(0.0312)
Constant	-0.352***	0.416***	9.688***
	(0.0923)	(0.0441)	(0.170)
R^2	0.119	0.031	0.033
Observations	185641	185641	172903

Notes: Estimated average effect of the reform on the probability of working (1) or having pension qualifying earnings over certain thresholds (2)-(7) and total log earnings while studying (8). All specifications include study program and year fixed effects as well as control variables as described in Table 1.5. Standard errors that are clustered at study program university level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.10: Reform effect on the probability of being in different earnings intervals

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: Earnings interval	1-5000	5001-10000	10001-20000	20001-30000	30001-40000	40000+
Reform period*Treated	0.00587*	0.00691*	0.0175*	-0.0215***	-0.0130	-0.000137
	(0.00342)	(0.00395)	(0.0103)	(0.00787)	(0.0109)	(0.0152)
Constant	0.0869***	0.0896***	0.162***	0.150***	-0.0653	-0.00647
	(0.0318)	(0.0221)	(0.0547)	(0.0447)	(0.0394)	(0.0774)
R^2	0.007	0.004	0.014	0.009	0.003	0.039
Observations	185641	185641	185641	185641	185641	185641

Notes: Estimated average effect of the reform on the probability of having earnings that correspond to different earnings intervals. All specifications include study program and year fixed effects as well as control variables as described in Table 1.5. Standard errors that are clustered at study program university level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix

1.A Additional tables

Table 1.A.1: Pre-reform trends in the probability of timely graduation and delay

	(1) Ontime	(2) Ontime	(3) Delay	(4) Delay
Treatment group*Year	0.00647		-0.0236	
	(0.00500)		(0.0326)	
Year	-0.0119***		0.0257	
	(0.00295)		(0.0219)	
Treatment group*1987		-0.0149		0.113
		(0.0129)		(0.168)
Treatment group*1988		0.0195		-0.0170
		(0.0197)		(0.155)
Treatment group*1989		0.0280		-0.0484
		(0.0177)		(0.158)
Treatment group*1990		0.0121		-0.0306
		(0.0219)		(0.148)
1987		0.00944		-0.0745
		(0.00972)		(0.0779)
1988		-0.0233*		0.0698
		(0.0130)		(0.0957)
1989		-0.0445***		0.148*
		(0.0113)		(0.0794)
1990		-0.0334**		0.0221
		(0.0155)		(0.0865)
R^2	0.454	0.455	0.404	0.404
Observations	12847	12847	12847	12847

Notes: All specifications include study program fixed effects and control variables. Standard errors are clustered at study program-univeristy level and are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.2: Comparison of standard errors using different levels of clustering

	(1) Unadjusted	(2) Study program	(3) Study program and Year	(4) Study program * University
Outcome variable: On time				
Years in treatment	0.0153***	0.0153**	0.0153***	0.0153**
	(0.00221)	(0.00521)	(0.00557)	(0.00760)
No. of clusters	0	16	16 + 11	56
R^2	0.431	0.431	0.431	0.431
Observations	34220	34220	34220	34220
Outcome variable: Delay in semesters				
Years in treatment	-0.133***	-0.133	-0.133	-0.133**
	(0.0165)	(0.0786)	(0.0813)	(0.0596)
No. of clusters	0	16	11 + 16	56
R^2	0.378	0.378	0.378	0.378
Observations	34220	34220	34220	34220

Notes: Cohort and study program fixed effects included, as well as control variables as described in Table 1.5. Column 1 uses unadjusted OLS standard errors. In Column 2 standard errors are clustered at the study program level. Column 3 uses standard errors clustered at the study program times expected graduation year level, and Column 4 uses standard errors clustered at the study program by university level. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.B Synthetic control method

The synthetic control method was developed for analyses of aggregate data where there is one treated unit and several possible control units, typically states in the US (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2012). It is, however, possible to apply the method to individual level data with some simple modifications. In order to perform the synthetic control matching I aggregate the individual level data to study program by expected graduation year level and since the synthetic control method only allows for one treated unit I further aggregate all the treated programs into one unit.

The matching algorithm aims at creating a control group that is as close as possible to the treated group in terms of the level of the outcome variable. Because the levels of the outcome variable in the treatment and control group are very different, I use the demeaned values of the outcome variables when performing the matching.[31]

The matching is then performed separately for both outcome variables, the probability of graduating on time and delay. The predictor variables that are used to construct the synthetic control group are simply the group average of the outcome variable in each of the pre-intervention years 1986-1989.

The results from the matching are presented in Table 1.B.1. For both outcome variables, the root mean squared prediction error (RMSPE) is very low, which suggests that the fit of the synthetic control group is good. The weights of the control units are presented in the Panel B. In both cases, the most weight is given to engineering, followed by agriculture and medicine, which are the programs that one would also chose based on logical reasoning (see also Section 1.6.3). The predictor variable means of the unweighted and the synthetic control groups are displayed in Panel C, and it is clear that the sample means of the synthetic control groups are closer to the mean of the treated group than the mean of the unweighted control group. These results, in combination with Figure 1.2, suggest that the synthetic control method was successful in generating a synthetic control group from the treated study programs.

[31] This is done by normalizing the level of the outcome variable to zero in 1990.

Table 1.B.1: Comparison of baseline and synthetic control groups

	Treated	Unweighted control	Synthetic control	
			On time	Delay
Panel A: Root Mean Squared Prediction Error				
RMSPE			.0117489	.0577496
Panel B: Weights				
Medicine			.226	.267
Agriculture			.355	.102
Engineering			.419	.631
Pharmaceutical science			0	0
Veterinary science			0	0
Architecture			0	0
Panel C: Predictor balance based on aggregate data				
Demeaned ontime 1986	.0275	.04795607	.0313999	
Demeaned ontime 1987	.0230161	.04700553	.0405342	
Demeaned ontime 1988	.0210407	-.02303465	.0061095	
Demeaned ontime 1989	.0082371	-.03673062	.0055667	
Demeaned delay 1986	-.0349998	-.06151698		-.0875597
Demeaned delay 1987	-.0449406	-.10231249		-.0530882
Demeaned delay 1988	.0042745	.25325278		.1060128
Demeaned delay 1989	.0333242	.32804474		.0459992

Notes: The synthetic control method selects a control group according to the weights in Panel B, by matching the pre-reform values on the outcome variable in Panel C.

1.C Duration of delay

The turbo grant reform aimed at increasing the share of students who graduated on time, but the data also offers an opportunity to study the effect on the duration of delay. This variable is of at least as big interest as the share of students graduating on time since it affects the resources that are spent on each delayed student. If the goal of the government is to decrease education spending, it should be concerned with reducing delay as this affects both spending at the educational institutions and student aid.

In the data, reported delay varies between -5 and 16 years, and I suspect that some of the extreme values are due to reporting error. I drop observations below the 1st and above the 99th percentile in the delay distribution, but this does not largely affect the results. The results are also robust to excluding 5 and 10 percent in the tails.

Figures 1.C.1 and 1.C.4 and Tables 1.C.1 to 1.C.2 are identical to those in Section 1.6 only the outcome variable is different. The main results are presented in Table 1.C.1. Columns 1 and 2 suggest that one additional year of treatment reduced delay by about .13 semesters, and that the inclusion of predetermined variables does not affect the estimate. If I extrapolate the result to a treatment of six years the accumulated effect is a reduction of 0.8 semesters. This corresponds to a 20 percent reduction compared to the average pre-reform delay in the treatment group.

In Columns 3 to 6, I study the effect on delay by student characteristics and I find the similar but not identical patterns as in the main analysis. The reduction in delay is to some extent driven by high ability students, but the role of parental education is less pronounced and not statistically significant. While there was no significant gender difference in the effect on timely graduation, female students reduced their delay significantly more than their male peers.

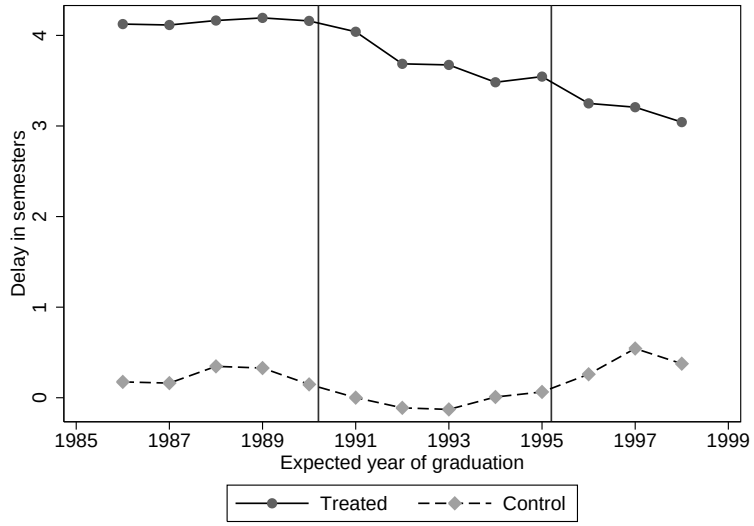
Tables 1.C.3 and 1.C.4 suggest that the results are robust to changes both in the control group and to the inclusion of controls for unemployment rates and cohort size, although some of the estimates are a bit smaller.

When estimating study program specific treatment effects, I find that students in the humanities experienced the largest reduction in delay following the reform. One year of treatment resulted in a 0.38 semester reduction. Science and social science students also experienced larger than average reductions in delay. Among law and psychology students there was no sign of change in average delay following the reform, even though the share of students who graduated on time increased.

This indicates that some delay increased in some parts of the delay distribution.

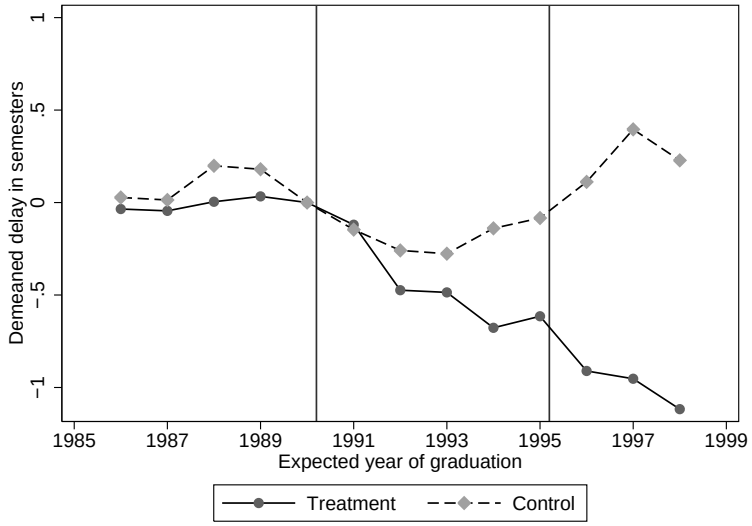
The reform effect on delay is perhaps not as striking as the effect on the probability to graduate on time, but this is expected. Average delay was almost two years in the treatment group, but a considerable share of the students were delayed by a lot more than that. These students were, however, not likely to respond to the incentives and they would therefore keep the average delay high. The compliers are likely to be students who would otherwise have been only a little delayed, and these students therefore only contribute to a small reduction in average delay. That said, and taken into account that reducing delay was not an explicit goal of the reform, the impact of delay is considerable.

Figure 1.C.1: Delay by treatment status

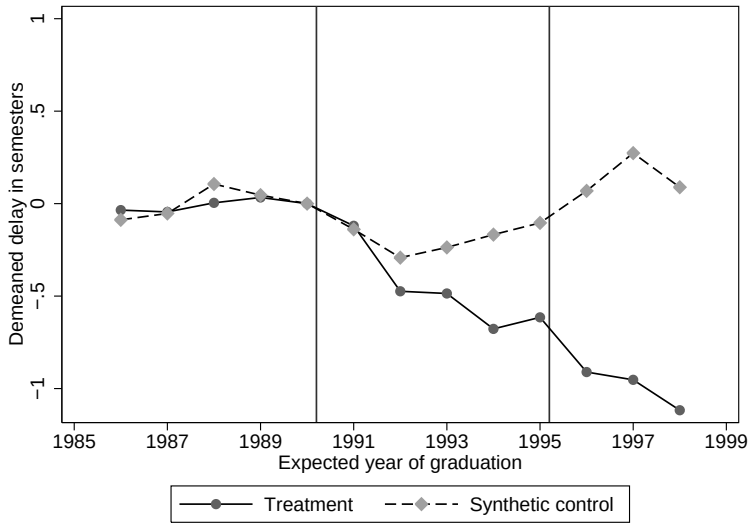


Note: The treated group contains individuals enrolled in humanities, social sciences, science, law, arts, theology, business administration, psychology, dentistry and fishery. The control group consists of individuals enrolled in medicine, agriculture, engineering, pharmaceutical science, veterinary medicine and architecture. The vertical lines refer to the implementation and termination date of the reform, respectively.

Figure 1.C.2: Demeaned delay by treatment status



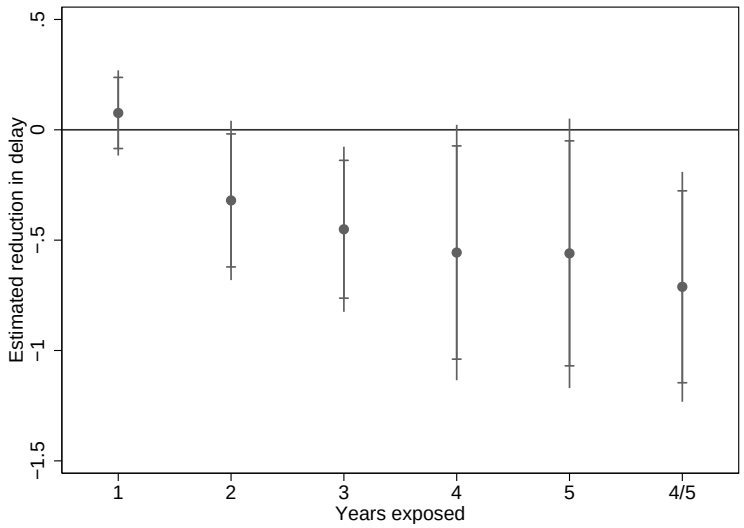
(a) Unweighted control group



(b) Synthetic control group

Note: Trends are demeaned by subtracting the group specific average of the outcome variable in 1990 from the group average in each year. The treated group contains individuals enrolled in humanities, social sciences, science, law, arts, theology, business administration, psychology, dentistry and fishery. The control group in (a) as defined in Figure 1.1 and the control group in (b) as described in Appendix 1.B. The vertical lines refer to the implementation and termination date of the reform, respectively.

Figure 1.C.3: Testing for non-linear treatment effects



Note: Estimates are obtained by regressing the outcome variable on dummies for years treated (see Section 1.5) to test for non-linear treatment effects. The last estimate corresponds to individuals that were expected to graduate in 1996, who were either treated for four of five years. 90 and 95 percent level confidence intervals included.

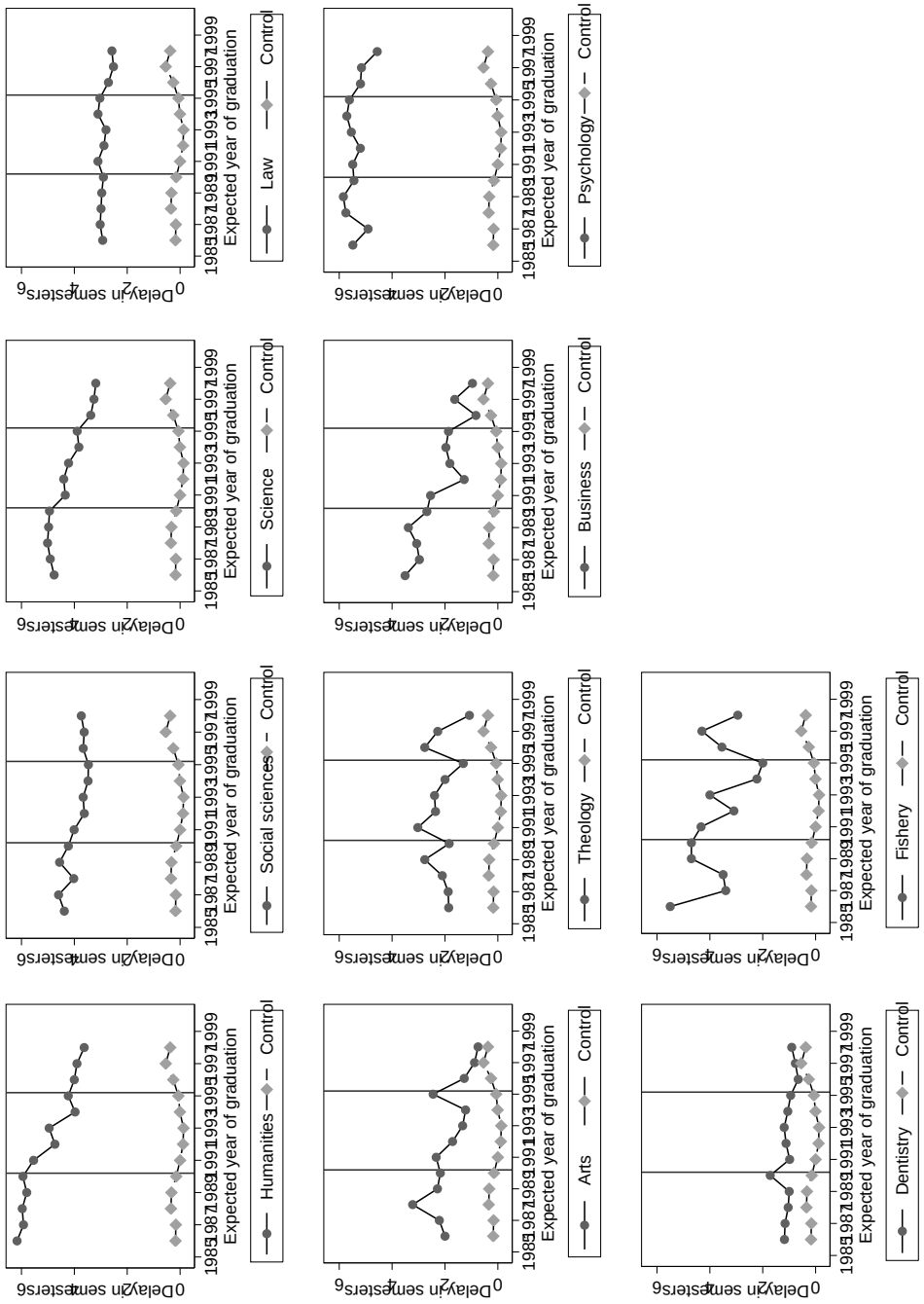


Figure 1.C.4: Study program specific trends in delay

Note: The control group consists of individuals enrolled in medicine, agriculture, engineering, pharmaceutical science, veterinary medicine and architecture. The vertical lines refer to the implementation and termination date of the reform, respectively.

Table 1.C.1: The estimated effect on delay by student characteristics

	(1) Raw	(2) Baseline	(3) Female	(4) Family income	(5) Parental education	(6) IQ
Years treated	-0.130**	-0.133**	-0.111*	-0.124*	-0.0488	-0.0435
	(0.0612)	(0.0596)	(0.0588)	(0.0701)	(0.0953)	(0.0874)
Female		0.351***	0.406***	0.350***	0.349***	
		(0.0557)	(0.0697)	(0.0558)	(0.0552)	
Female*Years treated			-0.0433**			
			(0.0214)			
Family income 2nd quartile				0.0842		
				(0.0534)		
Family income 3rd quartile				0.0778		
				(0.0780)		
Family income 4th quartile				0.192		
				(0.117)		
Family inc 2nd q*Years treated				-0.0176		
				(0.0233)		
Family inc 3rd q*Years treated				-0.0119		
				(0.0316)		
Family inc 4th q*Years treated				-0.00683		
				(0.0438)		
Intermediate parental education					-0.0867	
					(0.101)	
High parental education					0.146	
					(0.169)	
Intermediate par educ*Years treated					-0.0729	
					(0.0492)	
High par educ*Years treated					-0.0968	
					(0.0625)	
Ability 2nd quartile						0.0114
						(0.174)
Ability 3rd quartile						0.0658
						(0.319)
Ability 4th quartile						0.311
						(0.362)
Ability 2nd q*Years treated						-0.0223
						(0.0560)
Ability 3rd q*Years treated						-0.136
						(0.0871)
Ability 4th q*Years treated						-0.237**
						(0.0971)
Constant	1.690***	-1.357	-1.409	-0.522	-1.179	-1.321
	(0.0982)	(1.590)	(1.607)	(1.677)	(1.505)	(1.588)
R^2	0.372	0.378	0.378	0.378	0.377	0.395
Observations	34220	34220	34220	34220	34220	18723

Notes: Difference in difference estimates of the effect of one additional year of treatment on the duration of delay. In Column 1 Eq. 3.1 is estimated without control variables, and Columns 2 to 6 are estimated including controls for gender, age at high school graduation, region of residence, immigrant status, parental education and log family income at age 16. All specifications include study program and cohort fixed effects. In Column 3 the treatment variable is interacted with gender and in Column 4 dummies for family income at age 16 quartiles are interacted with treatment. In column 5 treatment is interacted with parental education dummies, where high education means that at least one parent has higher education, intermediate education means that at least one parent has a high school degree, but no more and low education means that parents have not finished high school. In Column 6 the treatment variable is interacted with dummies for ability quartile restricting the sample to male students. Standard errors that are clustered at study program university level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.2: The estimated effect on delay by study program

	(1) Humanities	(2) Social Sciences	(3) Science	(4) Law	(5) Arts
Years treated	-0.378***	-0.189***	-0.235***	0.0203	-0.126
	(0.0245)	(0.0239)	(0.0252)	(0.0196)	(0.0768)
Constant	-0.128	0.627	-0.861	-1.716**	0.607
	(0.879)	(0.914)	(0.855)	(0.844)	(0.848)
R^2	0.436	0.343	0.446	0.299	0.213
Observations	18242	18625	19993	20601	15571

	(6) Theology	(7) Business	(8) Psychology	(9) Dentistry	(10) Fishery
Years treated	-0.0208	-0.282***	0.0251	0.0123	-0.586***
	(0.0471)	(0.0481)	(0.0317)	(0.0552)	(0.164)
Constant	0.494	0.483	0.740	0.501	0.321
	(0.849)	(0.855)	(0.849)	(0.825)	(0.849)
R^2	0.218	0.222	0.395	0.211	0.221
Observations	15894	15933	16542	16036	15500

Notes: Study program specific difference in difference estimates of the effect of one additional year of treatment on the probability of graduating on time using the unweighted control group. Control variables included, see Notes in Table 1.5. Standard errors (not clustered because of small number of clusters) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.C.3: The estimated effect on delay using alternative control groups

	(1) Baseline	(2) Engineering+ Agriculture	(3) P-score com. sup.	(4) P-score weighted	(5) Synthetic control	(6) Excluding engineering
Years treated	-0.133**	-0.119*	-0.132**	-0.134**	-0.0941	-0.183**
	(0.0596)	(0.0648)	(0.0577)	(0.0570)	(0.0592)	(0.0694)
Constant	-1.357	-2.997	-1.846	-1.911	-1.406	-2.733**
	(1.590)	(1.811)	(1.571)	(1.551)	(1.670)	(1.292)
R^2	0.378	0.384	0.378	0.373	0.326	0.245
Observations	34220	30200	32321	32324	33074	24801

Notes: Difference in difference estimates of the effect of one additional year of treatment on delay measured in semesters. All specifications include study program and cohort fixed effects and control variables as described in Table 1.5. Standard errors in parentheses clustered at study program university level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.C.4: Sensitivity checks: the effect of unemployment and cohort size

	(1) Baseline	(2) Unemployment	(3) Supply 1	(4) Supply 2	(5) Supply 1+2
Years treated	-0.133**	-0.133**	-0.0904*	-0.140**	-0.0869*
	(0.0596)	(0.0555)	(0.0476)	(0.0549)	(0.0469)
Constant	-1.357	-1.700	1.088	-1.528	1.051
	(1.590)	(1.570)	(2.630)	(1.565)	(2.349)
R^2	0.378	0.380	0.378	0.380	0.380
Observations	34220	34220	34220	34220	34220

Notes: Difference in difference estimates of the effect of one additional year of treatment on delay. All specifications include study program and cohort fixed effects and control variables as described in Table 1.5. Unemployment rate is national average unemployment rate for ages 25-54 measured at expected year of graduation. Cohort size refers to the number of students in the sample expected to graduate in a given year. In Column 3, study program specific cohort size is included and in Column 4 the total cohort size is included. In Column 5 both cohort measures are simultaneously included. Standard errors in parentheses clustered at study program university level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association 105*(490), 493–505.

Abadie, A., A. Diamond, and J. Hainmueller (2012). Comparative politics and the synthetic control method. MIT Political Science Department Research Paper No. 2011-25, MIT Political Science Department.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review 93*(1), 113–132.

Angrist, J., D. Lang, and P. Oreopoulos (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics 1*(1), 136–163.

Angrist, J. and V. Lavy (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *American Economic Review 99*(4), 1384–1414.

Berg, L. (1994). Hvilke forhold forstyrrer studieprogresjonen? In N. Vibe and T. Nygaard (Eds.), *Utdanning og arbeidsmarked 1994*, pp. 58–69. Utredningsinstituttet for forskning og høyere utdanning.

Berg, L. (1997). Leve på lån. Studie- og låneatferd fem semestre etter studiestart (Delrapport 1 fra Studiefinansieringsprosjektet). Rapport 11/97, Norsk institutt for studier av forskning of utdanning.

Bertrand, M., E. Duflo, and S. Mullainathan (2004, February). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Black, S. E., A. Bütikofer, P. J. Devereux, and K. G. Salvanes (2013, April). This Is Only a Test? Long-Run Impacts of Prenatal Exposure to Radioactive Fallout. Working Paper 18987, National Bureau of Economic Research.

Brunello, G. and R. Winter-Ebmer (2003). Why do students expect to stay longer in college? Evidence from Europe. *Economics Letters 80*(2), 247 – 253.

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

Donald, S. G. and K. Lang (2007). Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics 89*(2), 221–233.

Dynarski, S. (2002). The behavioral and distributional implications of aid for college. *American Economic Review 92*(2), 279–285.

Garibaldi, P., F. Giavazzi, A. Ichino, and E. Rettore (2012). College cost and time to complete a degree: Evidence from tuition discontinuities. *The Review of Economics and Statistics 94*(3), 699–711.

Glocker, D. (2011). The effect of student aid on the duration of study. *Economics of Education Review 30*(1), 177 – 190.

Goodman, J. (2008). Who Merits Financial Aid? Massachusetts' Adams Scholarship. *Journal of Public Economics 92*, 2121–2131.

Gunnes, T., L. J. Kirkebøen, and M. Rønning (2013). Financial incentives and study duration in higher education. *Labour Economics 25*, 1 – 11.

Häkkinen, I. and R. Uusitalo (2003). The effect of a student aid reform on graduation: A duration analysis. Working Paper Series 2003:8, Uppsala University, Department of Economics.

Heineck, M., M. Kifmann, and N. Lorenz (2006). A duration analysis of the effects of tuition fees for long-term students in Germany. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik) 226*(1), 82–109.

Joensen, J. S. (2010). Timing and incentives: Impacts of student aid on academic achievement. mimeo, Stockholm School of Economics.

Joensen, J. S. and E. Mattana (2014). Student aid, academic achievement, and labor market behavior: Grants or loans? mimeo, Stockholm School of Economics.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics 17*(2), 303 – 316.

Kunnskapsdepartementet (2012). Høyere utdanning 2012. Tilstandsrapport, Kunnskapsdepartementet.

Lånekassen (2012, May). Lånekassen www.lanekassen.no.

Leuven, E., H. Oosterbeek, and B. van der Klaauw (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association 8*(6), 1243–1265.

Liu, K., K. G. Salvanes, and E. Ø. Sørensen (2012). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. Technical report, NHH Norwegian School of Economics, Department of Economics.

Ministry of Education and Culture (2010). Ei paikoillanne, vaan valmiit, hep! Koulutukseen siirtymistä ja tutkinnon suorittamista pohtineen työryhmän muistio. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2010:11.

Ministry of Education and Culture (2012). Opintotuen rakenteen kehittäminen 2012. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2012:29.

Ministry of Education and Research (2002). Kvalitetsreformen. Veiledninger og brosjyrer 14.08.2002.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro unit. *The Review of Economics and Statistics 72*(2), 334–38.

OECD (2013). Education at a Glance 2013: OECD Indicators.

OECD (2014). Education at a Glance 2014: OECD Indicators.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012, January). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources 46*(3), 614–646.

St. Meld. nr. 14 (1993-94). Studiefinansiering og studentvelferd.

Statistics Norway (2013). Statistisk årbok 2013.

Statistics Norway (2015). Throughput of students in tertiary education.

St.meld. nr.7 (2007-2008). Statusrapport for kvalitetsreformen i høgre utdanning.

Try, S. (2000). Veksten i høyere utdanning: Et vellykket arbeidsmarkedstiltak? Rapport 2/2000, Norsk institutt for studier av forskning og utdanning.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review 93*(2), 133–138.

Chapter 2

Skills, education and wage inequality[*†]

Abstract

We study whether the high level of wage inequality in the US is driven by the composition of skills or the returns to skills as compared to a Nordic country with low wage inequality and a low dispersion of skills. We use data from the Survey of Adult Skills (PIAAC) collected in 2011-20012 and a decomposition method based on recentered influence function (RIF) regressions (Firpo, Fortin and Lemieux, 2007, 2009) that allows us to investigate the role of skills across the entire wage distribution. We find that changing the US distribution of numeracy skills has a small but positive impact on wages across the wage distribution, but wage dispersion increases, which is explained by the high returns to skill in the US. Introducing the skill prices of Finland on average reduces US wages, and again, wage inequality increases slightly. Motivated by the observation that the low educated in the US are a particularly low performing group, we study the impact of skills within education. We find that on average, low educated workers in the US would have most to gain from having the skill distribution of Finnish workers, but that the wage inequality in this group would increase considerably. Our results imply that differences in skills and skill prices cannot explain the cross-country differences in wage inequality, but that substantial wage gains stand to be made in the US by raising the proficiency of US workers.

[*] Joint work with Stephen Machin and Kjell G. Salvanes.

[†] We would like to thank Mariagrazia Squicciarini at the OECD for letting us access the PIAAC data. Susanna Sten Gahmberg also acknowledges the Directorate for Science, Technology and Industry at the OECD for their hospitality during her stay in Paris, and the Ministry of Trade, Industries and Fisheries for financial support.

2.1 Introduction

Differences in the levels and dispersion of adult basic skills have commonly been discussed as a reason why some countries have higher levels of wage inequality. A particular concern has been a higher frequency of adults with low levels of literacy and numeracy skills in some countries, notably the US and the UK, whereas this lower tail is largely absent in other countries. One key issue concerns the source of these basic skills deficiencies, specifically how it can be traced to how well or poorly the schooling systems of different countries deliver literacy and numeracy education. A second issue concerns the wage penalty that poor basic skills impart on workers and how these can explain wage inequality differences across countries.[1]

Much of what we know on these issues can be traced to a series of papers based on the International Adult Literacy Survey (IALS) which took place in the mid-1990s (Blau & Kahn, 2005; Devroye & Freeman, 2002; Freeman & Schettkat, 2001; Leuven, Oosterbeek, & Ophem, 2004). These papers rely on the observation that there are considerable differences between countries in the distribution of cognitive skills in IALS, and that there is a positive correlation between the inequality of skills and the inequality of wages across countries. The conclusion of these studies is, however, that differences in skills inequality only explain a modest part of the differences in wage inequality. These studies only look at various specific summary measures of wage dispersion, such as percentile ratios, and not at the full distribution.

We reconsider these findings, first by using a method introduced by Firpo, Fortin and Lemieux (Firpo, Fortin, & Lemieux, 2007, 2009; Fortin, Lemieux, & Firpo, 2011) that allows us to study the impact of skills on the entire distribution of wages, and second by looking at more recent data from the Survey of Adult Skills (PIAAC) (OECD, 2013a). Further, we study the effects of basic skills within educational groups to highlight the fact that the performance gap between countries varies across education groups. We place a focus on two countries, one that has high inequalities in basic skills (the US), and one that has low inequalities in basic skills (Finland). This focus permits

[1] This positive correlation is also found in a number of studies that uses aggregate data on test scores at young age matched to aggregate data on earnings for the same cohorts later in life. See, for example Gregorio and Lee (2002), Bedard and Ferrall (2003) and Checchi and van de Werfhorst (2014).

us to consider how inequality in basic skills and in earnings can vary right across the entire distribution.

Similarly to what previous studies have found using the IALS data, we find that the level of wage and skill inequality is higher in the US than in most other countries. In our decomposition exercise, we find that changing the distribution of skills has a positive impact on the distribution of wages in the US. Contrary to what one could have expected, imposing Finland's skill distribution on the US actually increases wage dispersion in the US. This is explained by the higher returns to skills among workers in the higher end of the wage distribution. Changing the skill prices has a larger effect on the wage distribution in the US. When Finland's skills prices are imposed on the US, US wages are on average reduced, and again, the level of inequality increases slightly.

Our descriptive analysis of skill and wage differences within education groups suggests that the basic skills of low educated workers in the US are considerably lower than those of low educated workers in any other country, but that the cross-country differences are much smaller among more educated workers. This motivated us to study the impact of skills on wages *within* education groups. We find that low educated workers would gain the most in terms of wages from having the skills of Finnish workers, but that this would also substantially increase the level of wage inequality in this group. Introducing the Finnish skill prices would reduce wages in all education groups, but to a varying extent. High education workers above the median in the wage distribution would be hit very hard relative to workers below the median, which indicates that there are very high returns to skills in the very top of the wage distribution on the US relative to Finland.

This study is structured as follows. A brief review of the literature is given in Section 2, and the Survey of Adult Skills (PIAAC) is presented in Section 3. Descriptive analysis of the distribution of skills and wages in the participating countries is presented in Section 4. The methodological framework for estimating counterfactual wage densities is outlined in Section 5, before the main results are reported in Section 6. Section 7 concludes.

2.2 Literature

In this section, we provide a brief overview of other papers that have tried to explain the international differences in income inequality by differences in skill inequality. These studies mainly use data from the International Adult Literacy Survey (IALS), which was the predecessor of PIAAC, and build on the observation that there is a positive correlation between the dispersion of cognitive skills and the dispersion of earnings.[2] In particular, the dispersion of both skills and income is exceptionally large in the US, which has also been the focal point of many studies.

Most of these studies apply decomposition methods to investigate the importance of skill dispersion for explaining the differences in wage inequality across countries. Freeman and Schettkat (2001) studied the relationship between skill and wage dispersion and employment focusing on Germany and the US, and found that the more compressed distribution of skills in Germany only explained a modest part of the difference in wage dispersion between the countries using variance decomposition. Devroye and Freeman (2002) performed a similar analysis using data on four countries and found that skill inequality only explains about 7 percent of the cross-country difference in wage inequality and that most of the cross-country differences in wage inequality stem from within skill group differences. Both of these studies find that skill prices are more important in explaining the differences in wage inequality.

Blau and Kahn (2005) continued this line of research, but performed a more comprehensive analysis using data from nine countries by applying the decomposition framework introduced by Juhn, Murphy and Pierce (1993) that decomposes the difference in wage inequality into three components: a measured characteristics effect, a wage effect and a residual effect. They found that differences in the distribution of skills explain part of the difference in wage inequality between the US and the other countries, but that the wage effect, i.e., the returns to observed characteristics in the wage equation, is more important.

[2] By matching aggregate data on test scores at an early age (such as the First and Second International Mathematics study, in 1964 and 1980, respectively) and matching these data to aggregate data on income dispersion later on in life for the same cohorts, Gregorio and Lee (2002), Bedard and Ferrall (2003) and Checchi and van de Werfhorst (2014), among others, find that educational dispersion (both in terms of test scores and years of education) is positively correlated with wage inequality.

Paccagnella (2015) was the first to study the relationship between skill and wage inequality using the PIAAC data by decomposing quantile differences using unconditional quantile regressions, a method that makes it possible to estimate the relative importance of the included covariates (Firpo, et al., 2009; Fortin, et al., 2011). His findings are in line with those of the previous studies, and he concludes that price effects dominates the composition effects, and that differences in education play a more important role than skills in explaining the cross-country differences in wage inequality.

Leuven, Oosterbeek and van Ophem (2004) took a slightly different approach by building on the work of Blau and Kahn (1996), who investigated the role of differences in supply and demand factors in explaining the differences in wage inequality across countries. Their conclusion is that differences in supply and demand for skills can explain about one third of the cross-country differences in relative wages between skill groups, when skill groups are defined using cognitive skills rather than education, and that the supply and demand framework does particularly well in explaining relative wages of low skilled workers.

It has not been the primary focus of any of the studies mentioned above to explain why there is skill dispersion between countries, although, for example, Freeman and Schettkat (2001) discussed the importance of the apprenticeship system in Germany for the lower dispersion in skills relative to the US, and Devroye and Freeman (2002) pointed out that the poor performance of low educated individuals in the US on the IALS tests is partly explained by immigrants, whose skills may be underestimated simply because of their poor English skills. Another strand of the literature has pointed out the importance of educational institutions and policies for both skill and income dispersion. For example, Checchi and van de Werfhorst (2014) showed that policies such as public pre-schooling, later school starting age and the introduction of standardized tests were correlated with lower educational inequality (both in terms of years of education and tests scores) and wage inequality, while tracking, school accountability and teacher autonomy were associated with more inequality.

While many of the IALS based studies mentioned above have noted the poor absolute and relative performance on the cognitive tests of workers with low education in the

US, this finding has not explicitly been taken into account in the analyses, however. The fact that much of the differences in skill inequality stems from the lower tail of the skill distribution indicates that skill differences are most likely to explain differences in wage inequality at the bottom of the distribution, and the effect on overall inequality may therefore be limited.

We argue that the poor skills of low educated individuals in the US may have a large effect on the wages of this group, and thereby also on the level of inequality within the group even if this does not show up in summary statistics of wage inequality. Following this argument, we extend the literature in two ways. First, we employ a reweighting technique introduced by DiNardo, Fortin and Lemieux (1996) that allows us to study the impact of changing the skill distribution on the full density of wages. Second, we study the impact of skills on wage dispersion within education groups to take into account that the cross-country differences in skills vary across education groups.

2.3 Data

The analysis relies on data from the Survey of Adult Skills, which is the outcome of the OECD Programme for the International Assessment of Adult Competencies (PIAAC) (OECD, 2013a). The survey was designed to measure the proficiency in literacy, numeracy and problem solving in technology-rich environments of the adult population in the participating countries. The survey focuses on skills that are important for meaningful participation in the labor market and in society in general. In addition to assessing the set of cognitive skills discussed above, a broad battery of background information was collected, including individual characteristics, family background, education, work life history and earnings, as well as information on skill use at home and in the work place.

In total, the survey covered about 166,000 adults aged 16-65 in 24 countries: Australia, Austria, Belgium (only Flanders), Canada, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Poland, the Russian Federation, the Slovak Republic, Spain, Sweden, the

United Kingdom (only England and Northern Ireland) and the Unites States.[3] Participation in the assessment of problem solving in technology-rich environments was optional, and four countries chose not to administer this part of the survey.[4] The data was collected between August 2011 and March 2012.[5]

The test was given in the official language or languages of each country. The survey was typically administered in the respondent's home under supervision of a trained interviewer. The assessment was mainly completed using a computer, but a paper version was available for respondents with very poor computer skills. There was no time limit on the assessment, but most participants completed the assessment in about 50 minutes and spent 30-45 minutes on the background questionnaire. The participating countries used different sampling schemes to meet national requirements, i.e., to be able to get reliable estimates for minorities or geographical regions, but the samples are weighted according to known population parameters, and post-sampling weights are supplied in the data (and used in the analysis).

The Survey of Adult Skills has a practical focus and it aims at testing skills that are important for successful participation in the labor market and useful in everyday life. The three domains of cognitive skills tested in the survey (discussed in more detail in OECD (2013a)) are:

- Literacy, defined as the ability to understand, evaluate, use and engage with written texts to participate in society, achieve one's goals, and to develop one's knowledge and potential. Literacy encompasses a range of skills from the decoding of written words and sentences to the comprehension, interpretation, and evaluation of complex texts.
- Numeracy, defined as the ability to access, use, interpret and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life. Numeracy involves

[3] Data from a nine additional countries (Chile, Greece, Indonesia, Israel, Lithuania, New Zealand, Singapore, Slovenia and Turkey) was collected in 2014, and data will be available in 2016.
[4] These countries were Cyprus, France, Italy and Spain.
[5] Except in Canada, where the data was collected between November 2011 and June 2012, and France, where the data was collected between September and November 2012.

managing a situation or solving a problem in a real context, by responding to mathematical content and concepts represented in multiple ways.

- Problem solving in technology-rich environments, defined as the ability to use digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks. The assessment focuses on the abilities to solve problems for personal, work and civic purposes by setting up appropriate goals and plans, and accessing and making use of information through computers and computer networks.

The questions asked in the survey are framed as problems that adults could face in day-to-day life at work or at home, for example retrieving information from a bibliographic search from a simulated library website, comparing price tags, managing a driver's logbook or booking a meeting room using a reservation system.

The three skill domains are designed to measure different dimensions of a respondent's set of skills but the test scores between the different domains are highly correlated. The correlation between the literacy and numeracy scores is 0.87, and the correlation between problem solving and numeracy (literacy) is 0.73 (0.77.).[6] In the analysis, we focus on numeracy skills, because these are arguably the most comparable across countries and potentially the most important in the determination of wages (Hanushek, Schwerdt, Wiederhold, & Woessmann, 2015). However, the results are similar when using literacy or problem solving test scores instead.

The test scores are reported on a 500-point scale. In the data, ten plausible values are reported for each test score. We use the average of the plausible values in the analysis, but the results are very similar if we use any single one of the plausible values.

From the background questionnaire, we obtain information on demographic characteristics, education, labor market status, experience and earnings. Information about educational attainment is obtained from a question where respondents are asked to state their highest level of qualification completed in accordance with the

[6] However, the correlation between the domains is lower than the correlation between the different domains in the International Survey of Adult Skills (IALS), which is the predecessor of PIAAC. In the IALS, the pairwise correlation between the domains was close to 0.9, making it difficult to distinguish between the three dimensions (see for example Leuven et.al. (2004), Blau and Kahn (2005)).

International Standard for Classification of Education (ISCED). Based on this information, we generate a variable that identifies individuals with low, intermediate and high education.[7]

The earnings measure that we use is PPP corrected hourly earnings excluding bonuses for wage and salary earners, expressed in USD. We use the scientific use file of the OECD that contains continuous wages for all participating countries. We remove the top and bottom one percent of the wage distribution in each country to remove outliers.

We restrict the sample to individuals aged 25-65, who reported being employed the week before the survey and working more than 30 hours because we want to restrict the sample to individuals with strong labor market attachment.[8] In addition we exclude self-employed, and individuals with missing experience and education.

In the introductory part of the analysis, we include all participating countries, except the Russian Federation because these data are preliminary and do not cover the Moscow area, and are therefore not representative of the country (OECD, 2013b). Descriptive statistics of this sample are presented in Table 1. The sample consists of 62,280 individuals in total, and the size of the country specific samples vary between 1,692 individuals in Cyprus and 11,545 in Canada. There is considerable variation in terms of both test scores and other covariates between the countries. Japan scored the highest in all three skill domains, while Italy, Spain and Poland had the lowest scores in literacy, numeracy and problem solving, respectively. In the main analysis we restrict our attention to Finland and the United States, and end up with a sample of 2,583 and 2,057 respondents from the two countries, respectively.

2.4 Decomposing differences in wage distributions

In previous empirical work, differences in the distribution of skills have been shown to explain only a modest part of the differences in wage inequality between countries.

[7] Low education is defined as having lower secondary education or less (ISCED 1, 2, 3C short, or less), intermediate education as upper secondary education (ISCED 3A-B or 3C long) or post-secondary non-tertiary education (ISCED 4A-B-C) and high education as any tertiary education (ISCED 5A, 5B or 6).
[8] Hours worked is not recorded in the Australian data, and therefore the Australian sample includes both part-time and full-time workers.

However, these studies have focused on summary measures of dispersion, which are uninformative of *where* in the wage distribution the differences in skills matter the most. This, we argue, could miss part of the story, since the largest differences in test scores between the US and the other countries occur in the lower tail of the skill distribution, and therefore, the impact of changing the skill distribution in the US may affect some parts on the income distribution more than others. We therefore use a more general framework that allows us to study the impact of skills across the entire distribution of wages.

We apply a method introduced by Firpo, Fortin and Lemieux (2009), whose method makes it possible to perform *detailed* decompositions of the difference in any distributional statistic between two groups by estimating recentered influence function (RIF) regressions.[9] In practice this decomposition method is similar to an Oaxaca-Blinder decomposition, but the outcome variable in the regressions is replaced by the RIF for the statistic of interest. We decompose the Finland-US difference in wages at 19 percentiles to study the impact of skills across the entire wage distribution. In addition, we also present decomposition results for several summary measures of wage dispersion, such as the 90/10, 90/50 and 50/10 percentile ratios, the variance and the Gini coefficient.

2.4.1 Method

To understand the method, let $v(F_Y)$ denote a distributional statistic (for example, a quantile) of the cumulative distribution of wages F_Y. When performing a decomposition of differences in $v(F_Y)$ between two groups, we divide this difference into a composition effect, which is related to the difference in observed characteristics, X, and a wage effect, which is related to the difference in the conditional distribution of wages, $F(Y|X)$. In the case of the mean, the wage effect only depends on the conditional mean of wages, but when decomposing other distributional measures, it depends on the entire wage distribution, which makes estimation more challenging.

[9] This method is also described in detail in Firpo, Fortin and Lemieux (2007) and Fortin, Firpo and Lemieux (2011). The goal of this section is only to provide a short summary of the method, and readers are advised to turn to the studies mentioned above for additional detail. The disposition of this section follows Firpo, Fortin and Lemieux (2011), who provide an instructive description of the method.

Let F_{Y_0} and F_{Y_1} be the cumulative wage distributions observed in country 0 and 1, respectively, and let *G* be a country indicator referring to the country where the worker characteristics are observed, so that $F_{Y_1|G=1}$ refers to the actual cumulative distribution of wages observed in country 1. Then, we can decompose the overall difference in the distributional statistic $v(F_Y)$ between the two countries, so that

$$\Delta_O^v = v\left(F_{Y_1|G=1}\right) - v\left(F_{Y_0|G=0}\right)$$

$$= \left[v\left(F_{Y_1|G=1}\right) - v\left(F_{Y_0|G=1}\right)\right] + \left[v\left(F_{Y_0|G=1}\right) - v\left(F_{Y_0|G=0}\right)\right]$$

$$= \Delta_W^v + \Delta_C^v$$

where Δ_W^v is the wage effect and Δ_C^v is the composition effect. $v\left(F_{Y_1|G=1}\right)$ and $v\left(F_{Y_0|G=0}\right)$ refer to the distributional statistics calculated using the actual data for each country. The challenge lies in estimating $v\left(F_{Y_0|G=1}\right)$, which is the distributional statistic of the wage distribution in a counterfactual country, where the characteristics are those observed in country 1 while the wage structure is that in country 0.

Over the years, many methods have been proposed to estimate the composition and wage effects, but they have generally not been successful in performing detailed decompositions (DiNardo, et al., 1996; Juhn, et al., 1993; Machado & Mata, 2005). Recently, Firpo, Fortin and Lemieux (2009) introduced a method that further allows us to divide the total composition and wage effects into the contributions of specific covariates using recentered influence functions (RIF). The influence function (IF) of $v(F_Y)$ represents the influence of an individual observation on that statistic, and the recentering part comes from adding the statistic $v(F_Y)$ to the influence function. This method is similar to the Oaxaca-Blinder decomposition of differences in the mean, but the outcome variable in the regression is replaced by the recentered influence function (RIF) of the statistic $v(F_Y)$. In other words, once the RIF is calculated, it is possible to run a regression of the RIF in the explanatory variables *X* for both groups, and perform an Oaxaca-Blinder decomposition where the composition effect and the wage effect can be rewritten as

$$\Delta_C^v = (\mathrm{E}[X|G=1] - \mathrm{E}[X|G=0])^{\mathrm{T}} \gamma_0^v$$

and

$$\Delta_W^v = \mathrm{E}[X|G=1]^{\mathrm{T}}(\gamma_1^v - \gamma_0^v)$$

where γ_0^v and γ_1^v are the regression coefficients from the regression of the RIF on the explanatory variables in country 0 and 1, respectively. However, Firpo, Fortin and Lemieux (2007) point out that the decomposition above may not give consistent estimates of the composition and wage effect if the conditional expectation of the RIF regression is nonlinear. In other words, the concern is that if the true relationship between wages and the observed characteristics *X* is nonlinear, but the relationship is approximated to be linear in a regression, then the regression coefficient on *X* will change if the distribution of *X* changes even if the wage setting mechanism is unchanged. What this means is that the price vectors γ_0^v and γ_1^v could be different just because they are estimated for different sets of *X*.

As a solution to this problem, Firpo, Fortin and Lemieux (Firpo, et al., 2007; Fortin, et al., 2011) suggest using a method that combines the RIF regression method with a reweighting method introduced by DiNardo, Fortin and Lemieux (1996). By reweighting the distribution of *X* in group 0 so that it is similar to that in group 1, we can construct a counterfactual country, which has the wage structure of country 0 but (approximately) the characteristics of country 1, which we use in the decomposition.[10] The details of the reweighting method are described in DiNardo et. al. (1996) and Fortin et. al.(2011), but in short, the following reweighting function is estimated as

$$\psi(X) = \frac{\Pr(X|G=1)}{\Pr(X|G=0)} = \frac{\Pr(G=1|X)\,/\Pr(G=1)}{\Pr(G=0|X)\,/\Pr(G=0)}$$

The reweighting function is then used as weights to get the counterfactual mean of the covariates $\bar{X}_{01}$, and the counterfactual regression coefficients $\hat{\gamma}_{01}^v$. Now the difference $\gamma_1^v - \gamma_{01}^v$ reflects the true change in the wage structure, since it holds the distribution of characteristics unchanged.

[10] In our case, we reweight the US to have the observable characteristics of Finland.

The total composition effect can now be rewritten as the sum of the true composition effect $\widehat{\Delta}_{C,t}^{v}$ and a specification error component $\widehat{\Delta}_{C,se}^{v}$:

$$\widehat{\Delta}_{C,rw}^{v} = (\bar{X}_{01} - \bar{X}_{0})\hat{\gamma}_{1}^{v} + \bar{X}_{01}(\hat{\gamma}_{01}^{v} - \hat{\gamma}_{0}^{v})$$

$$= \widehat{\Delta}_{C,t}^{v} + \widehat{\Delta}_{C,se}^{v}$$

The specification error is related to the fact that the unweighted decomposition only provides a linear approximation of the composition effect Δ_{C}^{v}. When the linear approximation of the composition effect is accurate, the specification error should be small. Therefore, calculating the specification error serves as a good specification test for the unweighted RIF decomposition.

The wage effect can similarly be expressed as the sum of the true wage effect plus a reweighting error:

$$\widehat{\Delta}_{W,rw}^{v} = \bar{X}_{1}(\hat{\gamma}_{1}^{v} - \hat{\gamma}_{01}^{v}) + (\bar{X}_{1} - \bar{X}_{01})\hat{\gamma}_{01}^{v}$$

$$= \widehat{\Delta}_{W,t}^{v} + \widehat{\Delta}_{W,re}^{v}$$

Where the reweighting error $\widehat{\Delta}_{W,re}^{v}$ reflects the fact that the reweighted mean $\bar{X}_{01}$ is not exactly equal to $\bar{X}_{1}$. The reweighting error should approach zero when the reweighting works well.

As Firpo, Fortin and Lemieux (2007) point out, identification of the composition and wage effect depends on two assumptions. The first assumption is ignorability, or unconfoundedness, which says that the distribution of unobserved factors that affect the wage setting is the same in the two groups, conditional on *X*. This is a strong assumption, but it is ultimately untestable. It is easy to think about reasons why this assumption would be violated. For example, the two countries differ in their degree of unionization. The second assumption is the overlapping support assumption, which says that there must be an overlap of the observable characteristics in the two groups. In other words, there must be no value in *X* that is only observed in one group. This assumption is more easily testable.

2.4.2 Estimation

In practice, the decomposition is performed as follows. We start by estimating the reweighting function $\psi(X)$. This is easily done in two steps. First, the pooled data for countries 0 and 1 is used to estimate the predicted probability of belonging to each group conditional on covariates *X* for each observation in the sample using a probit model.[11] Second, the predicted probabilities of belonging to group 0 ($\widehat{Pr}(G = 0|X)$) and 1 ($\widehat{Pr}(G = 1|X)$), together with the sample shares of each group ($\widehat{Pr}(G = 0)$ and ($\widehat{Pr}(G = 1)$, are used to calculate the reweighting function.[12]

Next, recentered influence functions (RIF) for the distributional statistics of interest (i.e., quantiles) are obtained non-parametrically as described in Firpo, Fortin and Lemieux (2009).[13,14] Once the RIFs are obtained, two Oaxaca-Blinder decompositions are performed at each quantile by replacing the outcome variable (log hourly wages) with the RIF. To get the composition effect, we compare country 0 (US) to the counterfactual country (US with the characteristics of Finland). To get the wage effect, we compare country 1 (Finland) to the counterfactual country. In addition, we perform a decomposition comparing unweighted US and Finland to be able to calculate the specification error and the reweighting error. The specification error is defined as the difference between the "total unexplained" in the unweighted decomposition and the reweighted decomposition of the composition effect. The reweighting error is similarly defined as the difference between the "total explained" in the unweighted decomposition and the reweighted decomposition of the wage effect.

We use the US as the reference country both since this is consistent with our previous analysis and since it is consistent with previous work (Blau & Kahn, 2005; Paccagnella, 2015). Our estimates are robust to choosing Finland as the reference country, and to

[11] In the probit model we interact a female dummy with controls for numeracy test scores, education, experience and experience squared.
[12] In our case, the estimated weights are multiplied by the PIAAC sampling weights.
[13] We have also estimated the RIFs parametrically using RIF-OLS, and the results are very similar.
[14] In the case of the variance and the Gini coefficient, the RIF is obtained parametrically by estimating a RIF-OLS regression where the covariates included are numeracy test score, education, gender, work experience and its square.

performing a threefold decomposition. Throughout the estimation, we use males with intermediate education as the reference group.[15]

2.5 The distribution of wages and skills

We start by presenting some descriptive evidence of international differences in the distribution of wages and skills in the PIAAC data, as well as the returns to skills from simple wage regressions.

2.5.1 Wage inequality

Figure 1 shows the 90/10, 90/50 and 50/10 log wage differentials for all countries covered by the PIAAC data. In line with previous research, we find that the level of wage inequality is higher in the US than in most other countries, although the level of inequality is slightly higher in Estonia and Korea. In contrast to what Blau and Kahn (2005) find using the IALS, we find that the level of inequality is higher in the top (90/50 difference) in the US than in the bottom (50/10 difference) of the distribution.[16] The UK is in the middle in terms of wage differences, while the Nordic countries are clustered in the low end.

To further highlight the differences in the wage distribution between the US and other countries, we estimate the wage densities of the US and Finland, Norway and the UK (albeit only England and Northern Ireland) in Figure 2.[17][18] This makes it easier to interpret the summary inequality measures in Figure 1. Figure 2 shows that the wage distribution is very wide in the US compared to the other countries, and that the left tail is especially thick. The wage distribution on the US is quite similar to that in the UK,

[15] Covariates included in the decomposition are numeracy test score, two education dummies (high and low), gender, experience and experience squared.
[16] It used to be the case that the level of inequality was higher in the bottom of the distribution than in the top, but since the 1980s, inequality has increased faster in the upper half of the wage distribution (Autor, Katz, & Kearney, 2008). Using March CPS data, Autor, Katz and Kearney (2008) find that inequality in the top exceeded that of the bottom by the mid-2000s and even earlier using the CPS May/ORG data.
[17] Finland is chosen because it is the country we use as a counterfactual in the main analysis, Norway because it has the most narrow wage distribution in the sample and the UK because it is a country with a level of inequality that is quite close to average in the sample. Also see Figure A1 in the Appendix.
[18] The densities are estimated using weighted kernel estimation as introduced by Rosenblatt (1956) and Parzen (1962). We use the Gaussian kernel with bandwidth 0.065.

which is narrower but also have a thick left tail. The wage distributions of Finland and Norway, on the other hand, lack the thick left tail and are much narrower.

2.5.2 The distribution of skills

From Table 1 it is evident that there are large cross-country differences in the level of numeracy skills, and in Figure 3 we show that there are also large differences in skill dispersion in terms of test score differentials across countries. The US is the most dispersed both in terms of overall inequality (90/10 differential), and in the top (90/50) and the bottom half (50/10) of the skill distribution. In the US, the 90^{th} percentile scored 139 points higher than the 10^{th} percentile, and 63 points higher than the median. The gap between the median and the 10^{th} percentile was 76 points. In all countries, skill inequality is higher in the bottom of the distribution than in the top. Interestingly, the Nordic countries are not concentrated in the lower end, but rather close to the middle.

The density of numeracy scores for the group of selected countries is plotted in Figure 4. There are some similarities to Figure 3, where we studied the density of wages. Again, we see that the distributions of the US and the UK are fairly similar, with thick left tails. Norway and Finland have more compressed skill distributions and do not have the thick left tail.

Thus, the pattern of high wage and skill inequality in the US compared to other countries that have been observed in the IALS data, is also present in the PIAAC data. What we find particularly interesting is that both the wage and the skill distribution in the US is characterized by a high concentration of workers in the lower end of the distribution. In an attempt to explain this phenomenon, we investigate the relationship between skills, education and wage inequality in the next section.

2.5.3 *Within* education relationship between skill and wage inequality

Wage inequality within education groups

As a first step towards explaining the high level of wage inequality in the US, we show the 90/10 wage differential among workers with low, intermediate and high education in Figure 5. There is considerable variation within education group inequality, both within and between countries. In Figure 5, the US stands out for two reasons. First, the wages in the high education group are highly dispersed compared to other countries,

while the US is placed in the middle in terms of inequality in the low education group. Second, the US has the largest gap between the high and low group. Again, we notice that the Nordic countries all have very low wage differences within each education group, while the UK seems to be in the middle also here.

Parallel to the analysis above, we estimate the wage densities of the US and Finland, Norway and the UK by education in Figure 6. The first panel is identical to Figure 2, and rescaled for comparison with the education group specific wage densities in the remaining panels. When splitting the sample by education, the most remarkable finding is the large mass of low education workers in the bottom tail of the wage distribution, relative both to the other countries and to more educated workers in the US. This implies that the thick lower bottom tail in the overall wage distribution is to a large extent explained by the low wages of low education workers. The high concentration of low education workers in the very left tail also suggests that the average wage in this group is very low, both relative that of low educated workers in other countries, and relative to high education workers in the US.[19] When looking at the distribution of wages among intermediate and high education workers, we see that the distribution is much wider and less skewed to the left, and, that the US and the UK are again more similar. Note also that the wage distribution of Finland is similar to that of Norway when we study low and intermediate education workers (although always placed to the left), but that the wage distribution of high education workers is much wider.

Skill inequality within education groups

Next we turn to skill differences within educational groups. As expected, Figure 7 shows that workers with higher education on average also have higher numeracy test scores. But it also shows that there is considerable cross-country variation in the numeracy skills of workers with the same level of education, which is a type of heterogeneity that is often abstracted from in empirical research. Most interestingly from our perspective, we find that the numeracy skills of low educated workers in the US are very low both in absolute and relative terms (relative both to more educated

[19] This second point is illustrated in Figure A2 in the Appendix, where we show the relative wages of high education workers (as compared to low education workers) by country. It shows that the gap between high and low education workers is large in a cross- country comparison.

workers in the US and to low educated workers in other countries). The average score in this group is 187, which corresponds to proficiency level 1 out of 5 (compare Figures A1 and A2 in the Appendix).[20] The performance of the intermediate and high education groups, on the other hand, does not stand out as either particularly good or bad in comparison to other countries. The US also has the largest gap in average score between high and low education (see Figure 8).

In Figure 9, we study the 90/10 skill differentials by education, and we find that the level of skill dispersion is decreasing in education in most countries, and also in the US. This is in contrast to what we saw for wages in Figure 2, where the wage dispersion was generally largest among the highly educated workers. The level of skill inequality among low educated in the US is the second highest in the sample, after the low educated in Austria. The level of skill inequality of intermediate and high education workers in the US is also in the high end of the sample.

We study the within education group distribution of skills more closely in Figure 10. Again, we show the skill distribution of all workers in the first panel, followed by the education group specific distributions. In this figure, there are many parallels to Figure 6, where we studied the density of wages by education. We see that in the intermediate and high education group, the skill dispersion in the US and the UK are quite similar, while the distribution of numeracy test scores are quite similar in Finland and in Norway. Just as in the case of wages, the largest differences appear in the low education group. From this figure we see that the level of skills in this group is far lower in the US than in the other countries. Most interestingly, however, the numeracy skills of the low educated in the US are highly dispersed. Thus, the thick left tail in the distribution of skills in the US is largely explained by the poor performance of workers with low education.

[20] The OECD describes Proficiency level 1 as follows: Tasks at this level require the respondent to carry out basic mathematical processes in common, concrete contexts where the mathematical content is explicit with little text and minimal distractors. Tasks usually require one-step or simple processes involving counting, sorting, performing basic arithmetic operations, understanding simple percents such as 50%, and locating and identifying elements of simple or common graphical or spatial representations (OECD, 2013a).

To sum up our results so far, we have seen that both the distribution of wages and skills in the US are wide compared to other countries, and that what characterizes the US is a high concentration of workers in the lower tail of both the wage and skill distribution. Low educated workers make up a large proportion of the mass in the bottom of the distribution (of both skills and wages), which is also what we could expect based on standard human capital theory. Workers with less education are less productive and therefore earn lower wages, and workers with lower education are less productive because they have lower skills. But the analysis above is of course not sufficient to draw any definite conclusions of the relationship between skills, education and wages and therefore we proceed to a more formal analysis. In the following section, we study the returns to skill in a Mincer wage regression framework, before we study the importance of differences in skill endowments and skill prices in explaining cross-country differences in wage inequality.

2.5.4 Returns to skill

Next, we consider the returns to skills and education by estimating wage regressions similar to those of Mincer. Each cell in Table 2 reflects an estimate from a separate regression. We standardize the test scores at the country level to have mean zero and standard deviation one, and the estimates can therefore be interpreted as the percentage change in hourly wages resulting from a one standard deviation increase in numeracy test scores.

The first row of Table 2, presents the raw return to numeracy skills, i.e., a simple regression of test scores on log wages, without other controls. The coefficient is statistically significant at the 1 percent level in all countries and varies between 11 percent in Sweden and 27.6 percent in the US. The Nordic countries are all in the low end, and the UK is the middle. In the pooled sample, the estimated return to a one standard deviation increase in the numeracy test scores is associated with a 19 percent increase in hourly wages. Controls for gender and labor market experience and its square are included in row 2, which in most cases only marginally affects the estimates. When we add dummies that control for the level of education (three categories) in row 3, the coefficient on numeracy test scores drop by 36 percent in the pooled sample, and within-country reductions range between 22 and 48 percent. In the US, the return is

reduced by 10 percentage points, to 17 percent, which is the second highest return after the UK. The reduction in the returns to skills when education is included in the regression reflects that a large part of the returns to skills comes the fact that individuals with higher skills also tend to have more education. However, both the coefficient on skills and education (not reported) are statistically and economically significant, which indicates that numeracy and education measure different but overlapping skills.

In the three following rows, we interact the numeracy test scores with the dummies for education to estimate whether the return to numeracy skills differ by level of education. The results imply that this is definitely the case. In most countries, the returns to numeracy skills are significant at all levels of education. While not all the estimates are significantly different from each other in all countries, the returns to numeracy skills seems to be higher among more highly educated workers. This is particularly pronounced in the US, where the return to numeracy skills ranges from 8.9 percent among individuals with low education to 21.1 percent among highly educated. Again, the Nordic countries come out with a low return.

The finding that the return to skills is high in the US even conditional on education, has important implications for the relationship between skills and wage inequality in the US. We saw in the descriptive analysis that the skills are highly dispersed in the US, and in combination with high returns to skills, this could explain the high level of wage inequality in the US. The finding that the returns to skill vary by level of education could further explain why wages of low educated Americans are compressed despite large variation in their level of skills (and why the opposite holds for highly educated workers in the US). We now take the findings from this analysis to a more formal analysis, and investigate the role of skills and skill prices in explaining wage inequality by decomposing cross-country differences in wages and wage dispersion.

2.6 Decomposing wage differences

2.6.1 Decomposition in the full sample

The main results from the decomposition of unconditional quantile differences in log hourly wages are presented in Figures 11 to 13. In the decomposition, we use the US as the reference country, and compare it to Finland in a twofold decomposition.[21] This means that the decomposition effect is calculated as the difference in mean characteristics multiplied by the price vector in the US, and the wage effect is calculated as the difference in prices multiplied by the mean characteristics in Finland. However, our results are almost identical when we use the opposite weighting scheme, and therefore we interpret the wage effect as imposing Finland's wage structure on the US. We use males with intermediate education as the base group.

In addition, decomposition results of certain summary measures of wage dispersion are shown in Table 3. For simplicity, and for coherence with our descriptive analysis and with previous research, we report results for percentile differentials (i.e., the 90/10, 90/50 and 50/10 gaps), but we also report results for the variance of wages and the Gini coefficient, which give a more nuanced picture of the results. The first panel in the table shows the decomposition in the full sample. The first two rows show the distributional measures of interest for the two countries, and the difference between them is shown in the third row. The positive number reflects that the level of inequality is higher in the US in terms of all five inequality measures. This difference is then decomposed into a composition effect and a wage effect, and we further look specifically at the effect of skills and education.

Figure 11 shows the US-Finland difference in log hourly wages across the wage distribution, as well as the decomposition of this difference into a composition effect and a wage effect. The figure shows that the difference in log wages monotonically increases across the wage distribution, reflecting the wider distribution of wages in the US. Below the 40th percentile, the difference is negative, meaning that wages are higher in Finland, which again is the results of the more slim left tail in Finland, while wages

[21] The choice of Finland is based on the reasoning that we want to compare the US to a Nordic country with low income dispersion.

above the 40th percentile are higher in the US. The composition effect is negative across the entire wage distribution, but smaller in absolute value at the bottom of the distribution. In other words, this means that imposing Finland's distribution of characteristics in the US would increase wages across the entire wage distribution, but the increase would be smaller (in absolute size) in the bottom of the distribution. As also suggested by the results in Table 3, this would increase wage inequality in the US (and thereby the gap in wage inequality between Finland and the US), especially in the lower half of the wage distribution.

Differences in the wage structure are definitely more important in explaining the log wage gap between the US and Finland, as shown in Figure 11. The wage effect monotonically increases across the wage distribution. It is negative below the 25th percentile, meaning that in the bottom of the wage distribution the wage structure of Finland pays more. In the upper part of the distribution, the returns to observable (and unobservable) characteristics are higher in the US than in Finland. Put differently, by introducing Finland's wage structure in the US, wages below the 25th percentile would increase while wages above the 25th percentile would decrease. This would dramatically reduce wage inequality in the US, as also suggested in Table 3. In fact, the level of inequality would in this case be slightly lower than in Finland, as seen from Table 3.

The specification and reweighting errors, which we discussed in Section 4, are shown in the first panel of Figure A3 in the Appendix. To recap the discussion from Section 4, the specification error reflects the fact that the unweighted RIF decomposition only provides a first-order approximation of the relationship between the outcome variable and the explanatory variables. If the specification error deviates from zero, it is an indication that the linear approximation works poorly and that the reweighted decomposition is preferable. From Figure A3, we see that the specification error is small in absolute size, but that it slightly deviates from zero below the median and in the very top of the distribution. We interpret this as an indication that the decomposition without reweighting would do a fairly good job, but that the reweighted specification is preferable. The reweighting error, which reflects how well the reweighting performs in matching the average characteristics of the counterfactual distribution to the average

characteristics in Finland, is very close to zero, suggesting that the reweighting is successful.

In Figure 12, we further decompose the composition effect to study the impact of changing the distribution of skills. Of the covariates included in the analysis, differences in the distribution of skills have the largest impact on differences in wages. The negative effect means that imposing the skill distribution of Finland in the US would increase the log wage gap between the US and Finland. This may sound counterintuitive at first, but the explanation is intuitive. Because the returns to skill are large and positive in the US, raising the level of skills would increase wages. The fact that the impact of skills is slightly smaller in the bottom of the wage distribution, reflects that the returns to skills are lower in the bottom of the distribution, and therefore a given increase in skills would have a smaller impact on wages than in the upper tail of the wage distribution. The smaller effect in the bottom also means that wage inequality would increase in the US if the distribution of skills was shifted to the level of Finland, which is also what we find in Table 3. However, the impact is quite modest. The 90/10 (50/10) differential in the US would increase by 1.7 (3.9) percent, while the US-Finland gap in the same measures would increase by 4.1 (7.6) percent. This increase is driven by the lower end of the distribution, since an improvement in skills would increase wages more at the median and the top than at the bottom.

As far as education is concerned, Finland has a slightly higher share of highly educated workers, and therefore imposing the educational level of Finland would increase US wages slightly in the higher end of the wage distribution. US workers at the lower half of the wage distribution have more work experience than their counterparts in Finland, and therefore changing the distribution of experience would reduce US wages at the bottom.

The wage effect is decomposed in Figure 13. Of all the variables that we control for in the decomposition, skill prices of skills have the largest impact (in absolute value) on the US-Finland log wage gap. Except for in the very top and bottom of the wage distribution, imposing Finland's skill prices in the US would considerably lower wages in the US and reduce the wage difference between the two countries. In fact, skill prices over-explain the log wage gap in the lower end of the wage distribution, meaning that

the log wage level would be higher in Finland in the skill prices in the US were changed. Despite the large changes in the level of wages, Table 3 shows that the impact on wage inequality expressed by the 90/10 ratio would be small. In fact, it would increase by 2.7 percent (and thereby increase the US-Finland gap). The changes in the 90/50 and the 50/10 differentials go in opposite directions, which is explained by the fact that wages are reduced more at the median than at the tails. Therefore, inequality increases at the top and decreases at the bottom of the distribution. The variance of log wages in the US would be reduced, while the Gini coefficient would increase.

Differences in the returns to education explain very little of the log wage difference between the two countries, although higher returns to education explains some of the log wage gap above the median. Put differently, US wages above the 60^{th} percentile would be slightly reduced if the education prices of Finland were imposed. This in turn would reduce wage inequality in the US and thereby the gap in wage inequality between Finland and the US, as shown in Table 3. Returns to experience are higher in the US below the 60^{th} percentile, and thereby imposing the returns to experience of Finland in the US would reduce the wage gap between the countries quite substantially.

The results so far suggest that changing the distribution and especially prices of skills would have an impact on the wage distribution in the US. Changing the distribution of skills to that in Finland would on average increase log wages in the US, while changing the prices of skills would reduce US log wages. However, the impact on wage inequality is rather small. We find that changing both the distribution and the price of skill to that in Finland would increase inequality as measured by the 90/10 ratio in the US (by 1.7 and 2.7 percent, respectively), and thereby the gap between Finland and the US. In terms of the other inequality measures studied, the results are more mixed. Changing the distribution of skills would only increase inequality in the lower end of the wage distribution (50/10), while changing the prices of skills would reduce inequality in lower end of the distribution and increase inequality at the top (90/50). The change in the variance and the Gini coefficient also go in opposite directions for the composition and the wage effect of skills. In sum, it is clear, that changing the prices

of skills has a larger effect, both on the distribution of log wages as such and on wage inequality effect, than changing the distribution of skills.

2.6.2 Within education decompositions

So far, we have seen that differences in skills only have a modest impact on the distribution of wages, and that for the most part, increasing the level of skills would actually increase wage inequality in the US. In addition, we have seen that this affects the lower tail of the distribution more strongly than the upper tail.

In the descriptive analysis in Section 5, we saw that when we split the sample by level of education, the low educated in the US stand out as being a particularly low performing group, both in terms of wages and skills. In this section, we therefore investigate the impact of changing the distribution of numeracy skills on wage inequality *within* educational groups, to see whether the overall pattern from the previous section also carries over to broadly defined education groups. Below, we study the three education groups in turn, starting with the low education group. We focus on features that differ from the main analysis in the previous section.

Low education

The decomposition of the log wage differences between low educated workers in the US and in Finland is presented in Figures 14 to 16, and the results are rather different from the analysis of the full sample. The total difference in log wages in this group is negative, which means that the wage level is higher in Finland. The total log wage gap is rather constant across the wage distribution, although slightly smaller in at the top. The second panel in Table 3 shows that in this group, the level of wage inequality is higher in the US, but the difference between the two countries is rather small compared to the overall sample.

The total composition effect is negative and substantially larger in absolute size than the composition effect in the overall sample. This means that there would be larger wage gains in this group from having the characteristics of Finnish workers with low education. As in the full sample, the composition effect is larger above the median, meaning US wages would increase more in the upper half of the wage distribution if they had the characteristics of Finnish workers. As also suggested by Table 3, this

would have a large impact on wage inequality in this group, and the difference in the 90/10 ratio between the two countries would more than double. The total wage effect is negative below the 65th percentile, and slightly positive above, meaning that introducing the wage structure of Finland would mainly have a positive impact on wages in the US, except for in the higher end of the distribution. Compared to the full sample, however, differences in the wage structure explain very little of the total wage difference.

The second panel in Figure A3 in the Appendix shows that the reweighting error is larger than in the full sample, which is partly explained by the smaller sample size. The specification error is also larger in the low education sample, which indicates that it is important to use the reweighting approach.

Figure 15 indicates that differences in the distribution of skills drive the total composition effect. Raising the level of skills of low educated Americans to the level of Finland would increase the wage level across the entire wage distribution, but wages would increase more above the median. As Panel B in Table 3 suggests, this would quite substantially increase the level of wage inequality in the US. More specifically, the 90/10 would increase by 14.5 percent, and the 90/50 and 50/10 ratios by similar amounts. The effect on the US-Finland gap in inequality is even larger. For example, the gap in the 90/10 ratio would be reduced by 81.6 percent. The composition effect of experience goes in the opposite direction compared to the full sample, meaning that low educated workers in Finland have more work experience that their US counterparts. This is partly explained by the sampling and the small sample size, which resulted in Finnish low education workers having 8 years of experience more than their US counterparts, although the average in the full sample is two years lower in Finland than in the US.

The components of the price effect are shown in Figure 16. The figure is rather noisy, which is largely explained by the small sample size. However, the figure clearly shows that imposing Finland's skill prices would increase wages of US low education workers. Table 3 shows that the difference in skill prices explain all (and more) of cross-country 90/10 and 90/50 gaps, but given the amount of noise in Figure 16, we do not want to push these results too far.

Intermediate education

The decomposition of the log wage gap in the intermediate education group is presented in Figures 17 to 19. The total log wage difference and the total composition and wage effects in Figure 17 look very similar to those for the full sample, but the slope of the total difference curve is less steep, which indicates that the two wage distributions are more similar. Again, wages below the 40^{th} quantile are higher in Finland. The composition effect is negative, and workers in the lower end of the wage distribution would gain less, relative to workers higher up in the wage distribution, in terms of wages having the characteristics of Finnish workers. Note that this comes from the fact that prices are higher at higher quantiles rather than relatively less favorable characteristics of workers in the lower tail.

Figures 18 and 19 are very similar to the corresponding graphs for the full sample, so we do not discuss them in great detail. The composition effect is driven by skills, but this effect is small relative to the wage effect of skills. The wage effect of skills exhibits more of an inverse U-shaped pattern for this group, with the largest wage reduction in the middle of the distribution. Workers at the 10^{th} percentile would experience a small increase in wages, which results in a reduction in 90/10 and 50/10 inequality. Inequality at the top of the distribution would increase.

High education

The aggregate decomposition of the log wage difference in the high education group in Figure 20 also looks quite similar to the decomposition in the full sample. Wages are higher in the US except in the very low tail, and the difference at the top of the distributions is rather large. The total composition effect is smaller in absolute size across the entire wage distribution than in the full sample, which means that changing the observable characteristics of US workers with high education would have less of an impact on their wages.

Nonetheless, Figure 21 shows that of the observable characteristics, skills matter the most, and changing the skill distribution to that of Finland would have a positive effect on wages across the wage distribution. The last panel in Table 3 shows that imposing the skill level of Finnish workers in the US would only have a marginal effect on wage inequality.

When decomposing the wage effect in Figure 22, a very interesting pattern appears. Imposing the skill prices of Finland in the US would reduce wages of US workers across the entire wage distribution, but this effect is especially large above the median. The reason is that within the group of highly educated workers in the US, there is a big spread in wages related to the skills within this groups. This indicates that the returns to skills are especially high for the very highly skilled in the US. In this case, Table 3 provides an excellent example that percentile ratios can fail to pick up changes in the wage distribution and that studying the entire distribution is important. Based solely on the aggregate measures of percentile ratios in Table 3, we would conclude that changing the skill prices would have no impact on wage inequality (or even a negative impact, meaning increasing wage inequality) in the high education group, although it is quite evident from Figure 22 that this is not the case. This is also indicated by the reduction in the variance and the Gini coefficient.

To sum up, changing the US distribution of skills to that of Finland would increase log wages in all education groups, but the size of the increase varies. We find that, on average, low educated workers would have the most to gain, which is also expected since the gap in skills between the countries is the largest in this group. However, changing the distribution of skills would increase wage inequality in all education groups, and the increase would be especially large in the low education group. Changing the skill prices would on average reduce US wages in all education groups, but the magnitude and the pattern of the change across the wage distribution varies by education group. Due to small sample sizes, there is some noise in the effects across the wage distribution, especially in the low education group. One of the more interesting patterns that emerge, is that the difference in skill prices is especially large above the median in the high education group, which suggests that there are high labor market returns to being the best of the best in the US.

Thus, our findings are in line with previous studies, in that skill prices seem to be more important in affecting the wage distribution in the US. However, in contrast to previous studies, we find that differences in skill prices cannot explain the gap in wage inequality between the US and Finland. However, we give a more nuanced picture of the role of

skills than previous studies; in part because we study the entire wage distribution rather than only summary measures of dispersion, and in part because we show that the importance of skills varies within educational groups, which is a dimension of heterogeneity that has been ignored in much of the existing literature.

2.7 Conclusion

There is a concern in the literature that basic numeracy and literacy skills deficiencies at the lower tail is causing the high degree of wage inequality observed in countries like the US and the UK, and that poorly functioning education systems may have a part in this development. The literature has studied the correlation between summary measures of wage and skill dispersion across countries, and the main finding using data for the 1990s is that differences in the skill distributions cannot really explain the large cross-country differences in wage inequality (Blau & Kahn, 2005; Devroye & Freeman, 2002; Freeman & Schettkat, 2001; Leuven, et al., 2004). However, these studies only look at various specific summary measures of wage dispersion, such as percentile ratios, and not at the full distribution of wages. Using the full wage distribution may be important since it provides us with more precise information regarding where in the wage distribution countries differ and where the skill distributions are different. Summary measures like percentile ratios provides only crude information on this.

In this paper we use data from the Survey of Adult Skills (PIAAC) collected in 2011-2012 to look at the entire distribution of wages and skills. We decompose the log wage difference between countries across the entire wage distribution into a composition component and a wage component to investigate the role of differences in skills and returns to skill. Further, we study differences in basic skills within educational groups to highlight the fact that the performance gap between countries varies across education groups. We focus on two countries, one that has high inequalities in basic skills (the US), and one that has low inequalities in basic skills (Finland). This focus permits us to consider how inequality in basic skills and in earnings can vary right across the entire distribution.

In the more descriptive part of the analysis, we confirm the findings of previous studies that the level of wage dispersion is high in the US relative to other countries. Wage inequality (measured as the 90/10 log wage differential) in the US is among the very highest in the studied countries. Closer investigation of the density of wages clearly shows that the distribution of wages is very wide, especially above the median, but that there is a relatively large mass of workers in the very bottom of the wage distribution. The distribution of numeracy skills is also very wide in the US compared to other countries, but there is more skill dispersion in the lower than in the upper tail. Here too, the lower tail is relatively thick compared to other countries.

In an attempt to explain this high concentration of workers in the bottom if the distribution, who potentially fair very poorly both in terms of wages and in terms of basic numeracy skills, we divide the sample by education, and study wage and skill dispersion *within* education groups. We find that the low education group (high school drop outs) in the US stands out as being very low-performing both in absolute and relative terms. The wage distribution of this group is not particularly dispersed but a larger concern is that the average wage level of this group is very low, both in a within and between country comparison. In addition, the left tail is very thick. The numeracy skills of the low education group are also alarmingly low compared both to those of low educated in other countries and to more educated in the US. In other words, it seems as if low educated workers contribute to the thick left tail both in the wage and the skill distribution in the US.

Next, we study returns to skill and find that the US is characterized by high returns to skill even conditional on education. High returns to skill could thereby be important in explaining the high level of wage inequality in the US. Even more interestingly, the returns to skill are higher at higher levels of education. In other words, returns to skill are low for workers with low education and significantly higher for more educated workers. This could explain why we find that wages of low educated workers are compressed despite large variation in skills (and to some extent the opposite among highly educated workers).

When using the RIF-regression based decomposition framework introduced by Firpo, Fortin and Lemieux (Firpo, et al., 2007, 2009, 2011; Fortin, et al., 2011), we find that

changing the skill distribution has a positive impact on the wage distribution in the US. Somewhat unexpectedly perhaps, wage dispersion increases slightly (the 90/10 differential increases by 1.7 percent) when the skill distribution of Finland is imposed. This happens because the positive impact on wages is smaller in the bottom of the distribution than at the top. In other words, even though the skills of low skilled workers increase relative to those of high skilled workers, the higher returns to skill make wages at the top increase more than at the bottom.

Changing the skill prices in the US would have a considerably larger impact on the wage distribution than the composition of skills. On average, imposing Finland's skill prices would reduce wages since the returns to skill are much lower in Finland than in the US. The price effect is rather constant across the wage distribution, and consequently the impact on wage inequality is small (the 90/10 differential increases by 2.7 percent). Note, however, that in contrast to the previous literature on this topic, we find that differences in skill prices cannot explain the cross-country difference in wage inequality as measured by the 90/10 log wage differential, since changing the skills prices actually slightly increases wage inequality in the US.

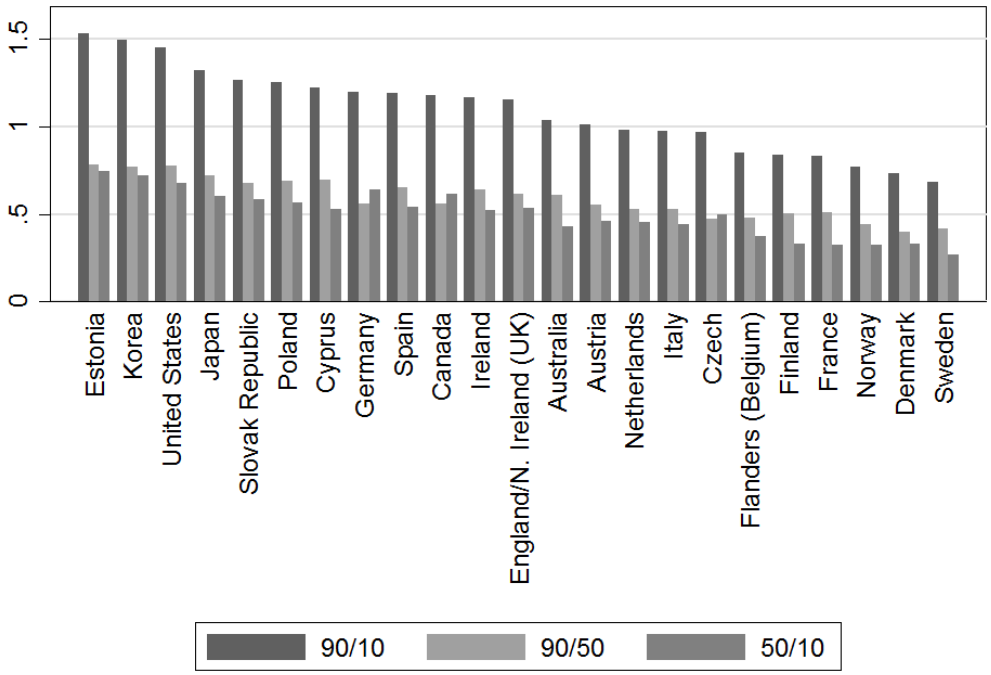
We then study the effect of skills within educational groups, and find qualitatively similar results as for the full sample, but also some differences. Low educated workers would on average have the most to gain in terms of wages from raising their skills to the level of Finnish workers. Because of the US price structure, however, the level of wage inequality would increase considerably in this group. Among highly educated workers, the composition effect of skills is small, but the wage effect is very large. Highly educated workers above the median would have their wages considerably reduced if the skill prices of Finland were introduced. This implies that there are very high returns to being among the very best in terms of skills in the US.

Our results are in line with previous work in that skill prices seem to affect the distribution of wages more than the distribution of skills. In contrast to other studies, however, we find that neither differences in skills nor skill prices can explain the large differences in wage inequality between the US and Finland. We find that both the composition effect and the wage effect of skills would increase wage inequality expressed as the 90/10 differential in log wages in the US, thus increasing the cross-

country difference. By studying the effect of skills on the entire wage distribution, we give a more nuanced picture of the importance of skills than much of the existing literature. In particular, we highlight that skills and skill prices have a significant impact on the wage distribution, although the impact on wage inequality is limited. We also show that measuring wage inequality in terms of percentile differentials is a rather crude measure that can miss important patterns in the data. In addition, we show that the impact of skills varies between education groups, which is a dimension of heterogeneity that has been ignored in much of the existing literature.

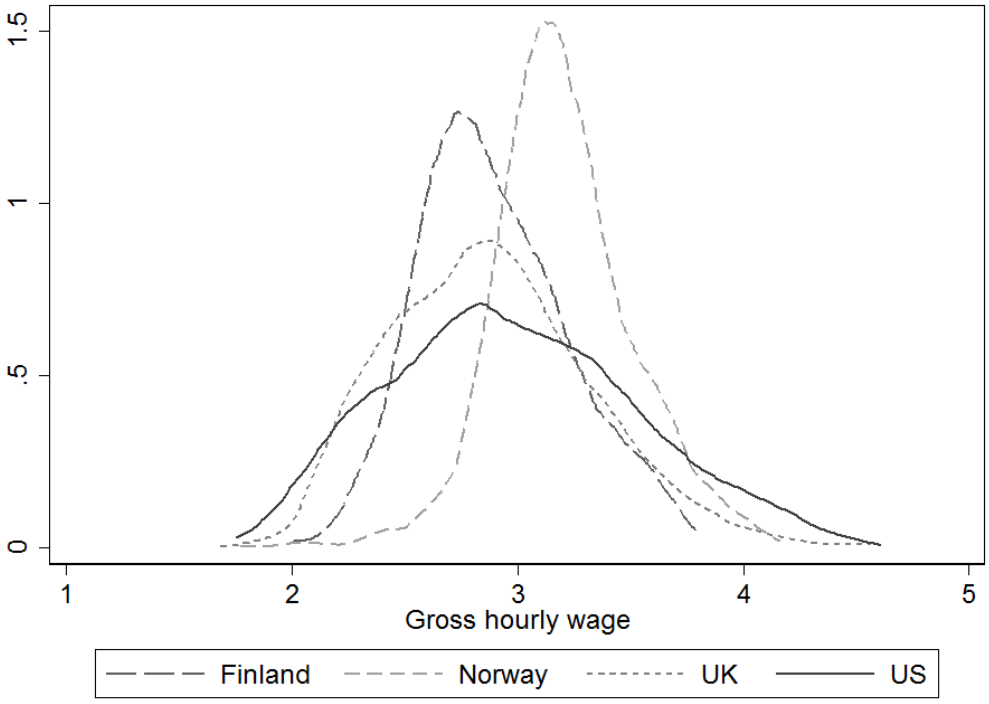
In sum, using more recent data and a more flexible approach measuring differences along the whole distribution, we do not find a big role for skill differences between the US and a Nordic high skill and low income dispersion country in explaining the difference in wage inequality. In future work, we intend to assess a broader set of factors that might be important such as institutional differences in the labor market and wage setting systems such as degree of centralization, minimum wages, and degree of unionization. There are interesting differences across countries regarding these institutions, even among the Nordic countries.

Figure 1: Log hourly wage differentials



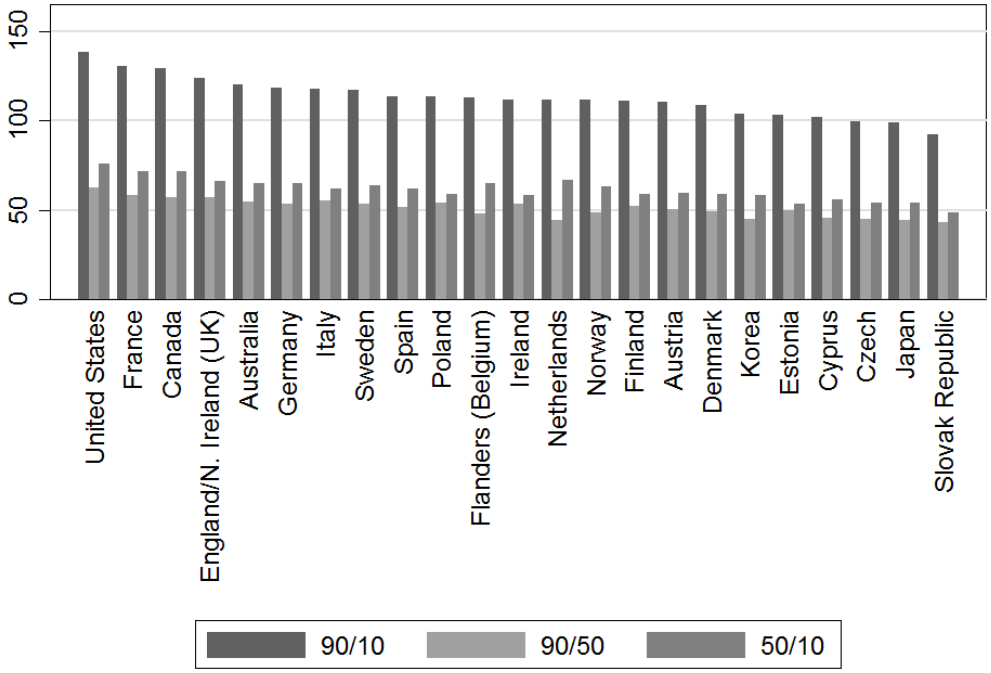
Source: Survey of Adult Skills (PIAAC) (OECD 2013a)

Figure 2: Estimated density of gross hourly wages in selected countries



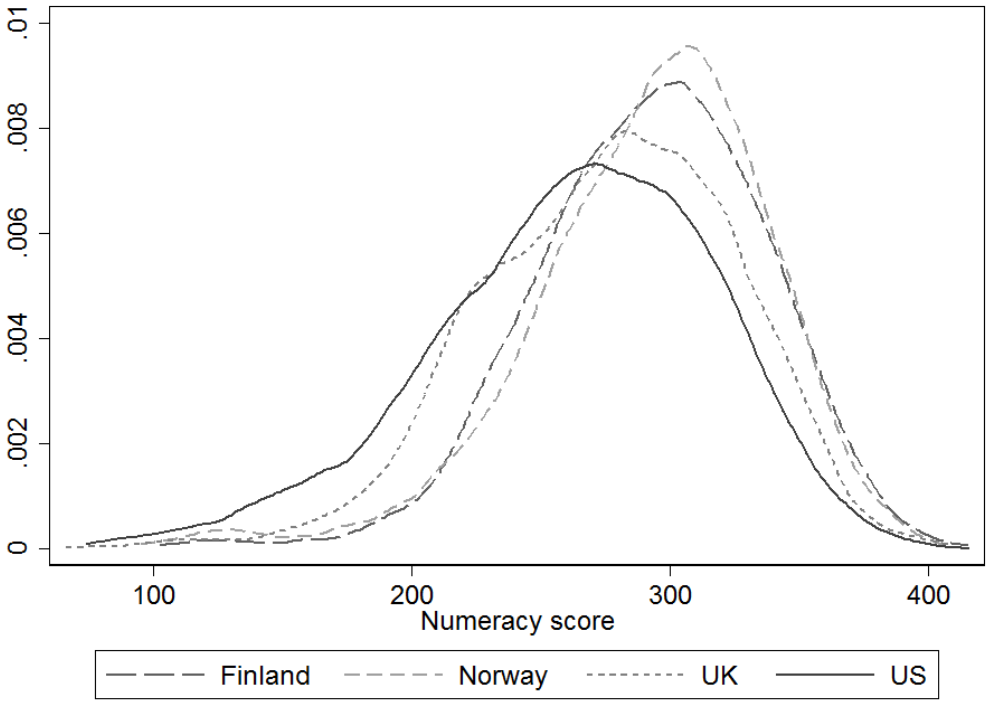
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 3: Numeracy skill differentials



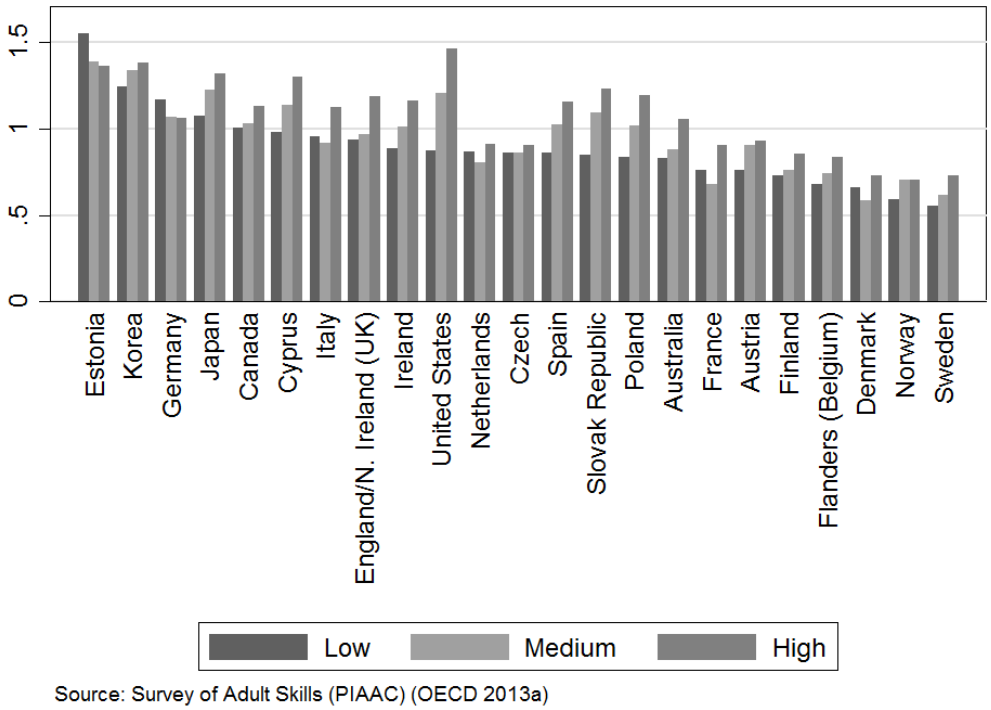
Source: Survey of Adult Skills (PIAAC) (OECD 2013a)

Figure 2: Estimated density of gross hourly wages in selected countries



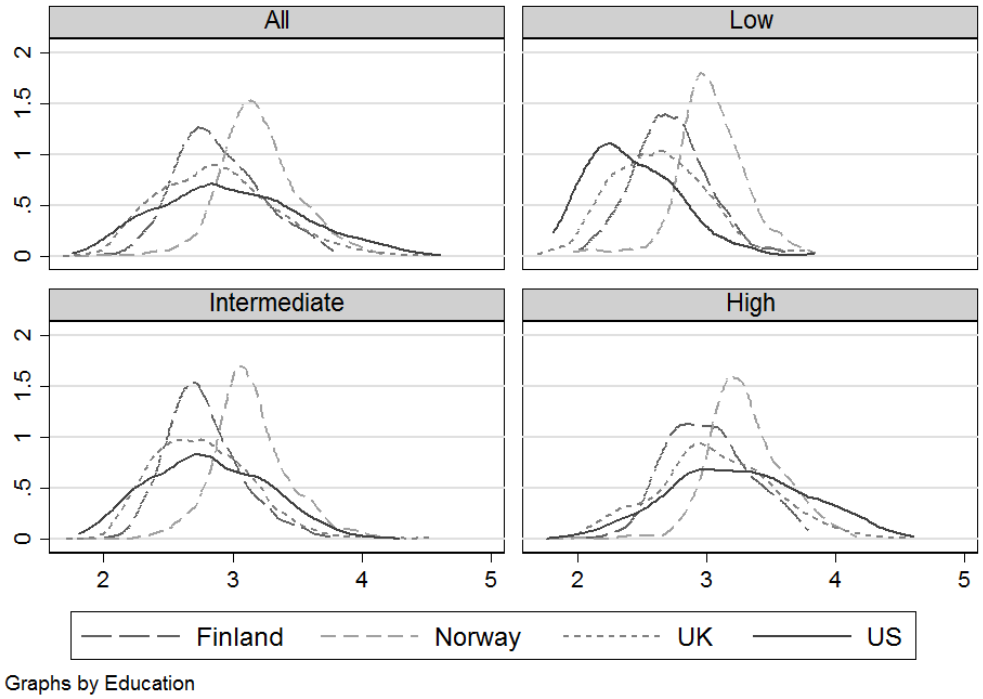
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 5: 90/10 differentials in log hourly wages by education



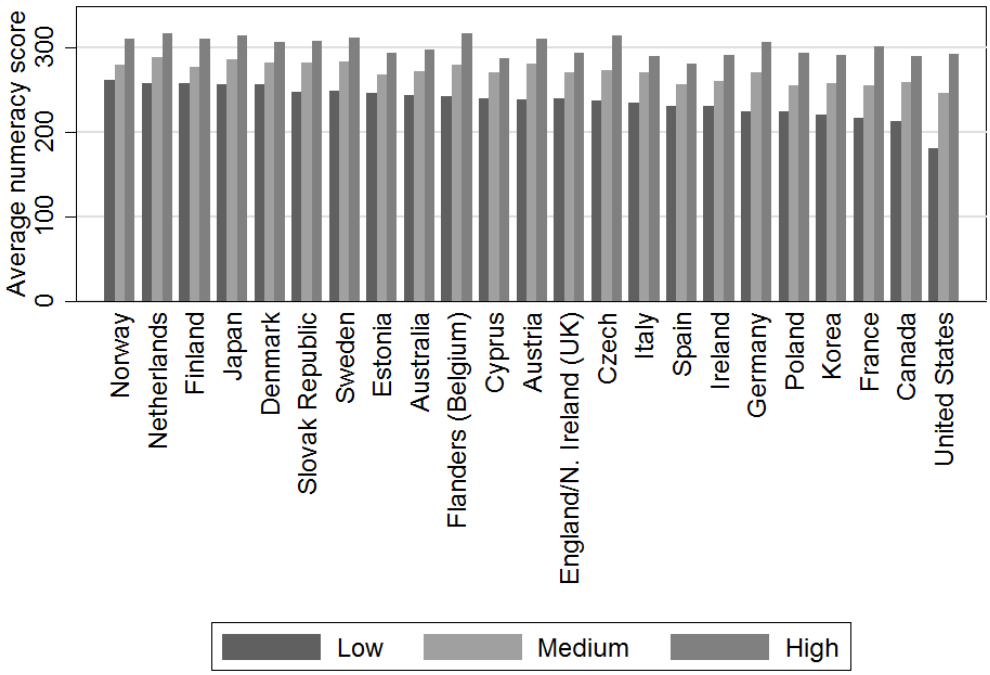
Source: Survey of Adult Skills (PIAAC) (OECD 2013a)

Figure 6: Estimated density of gross hourly wages in selected countries, all and by education



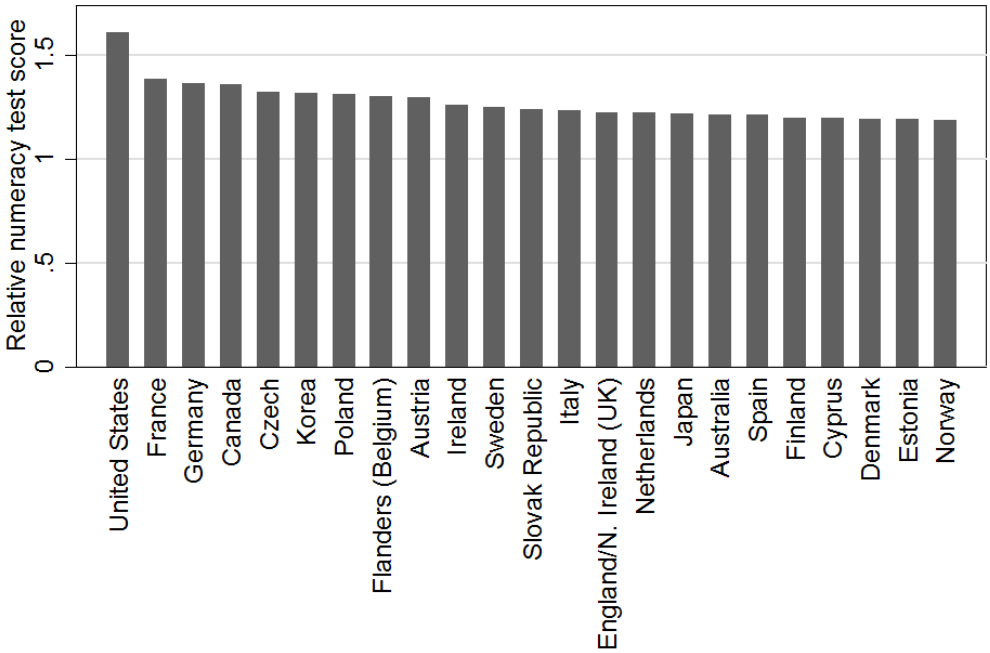
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 7: Average numeracy scores by education



Source: Survey of Adult Skills (PIAAC) (OECD 2013a)

Figure 8: Numeracy test scores of high education workers relative to low education workers



Source: Survey of Adult Skills (PIAAC) (OECD 2013a)

Figure 9: 90/10 numeracy test score differentials by education

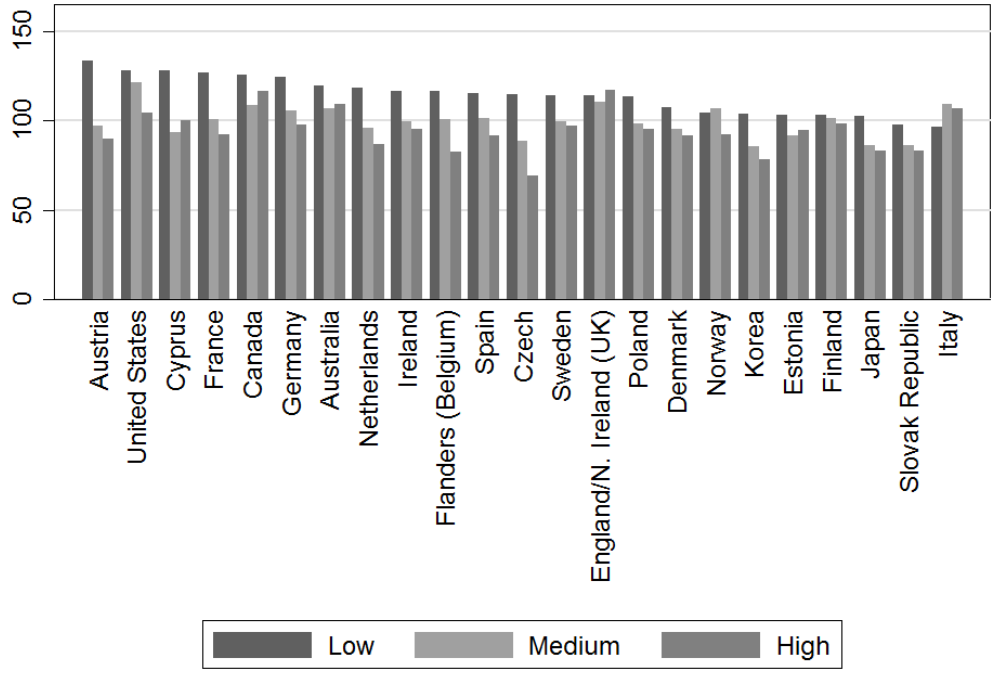
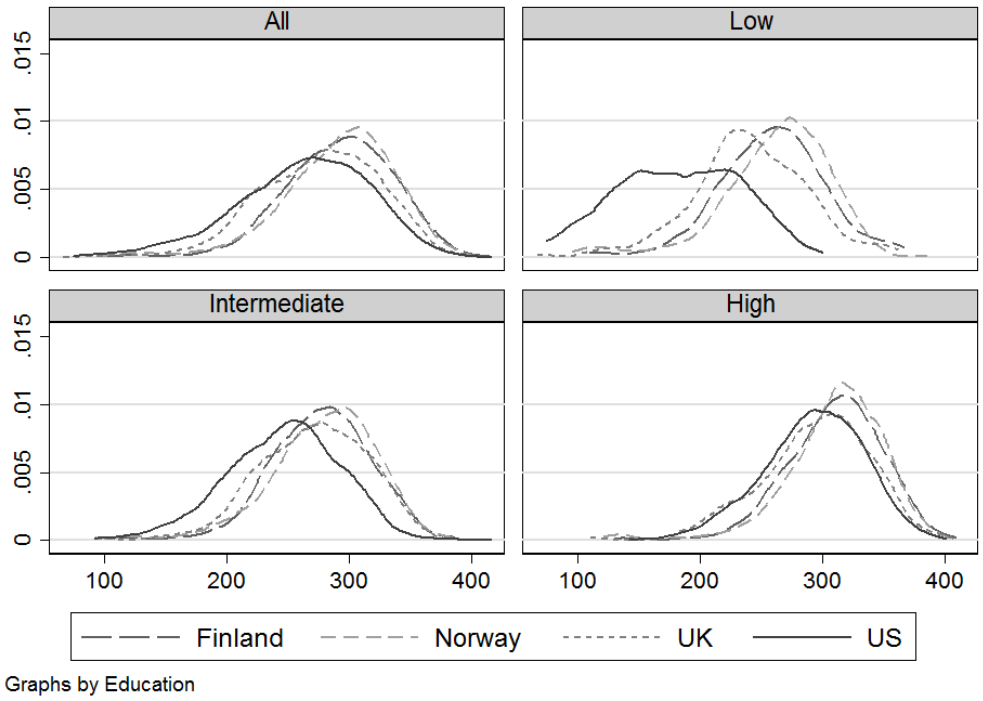
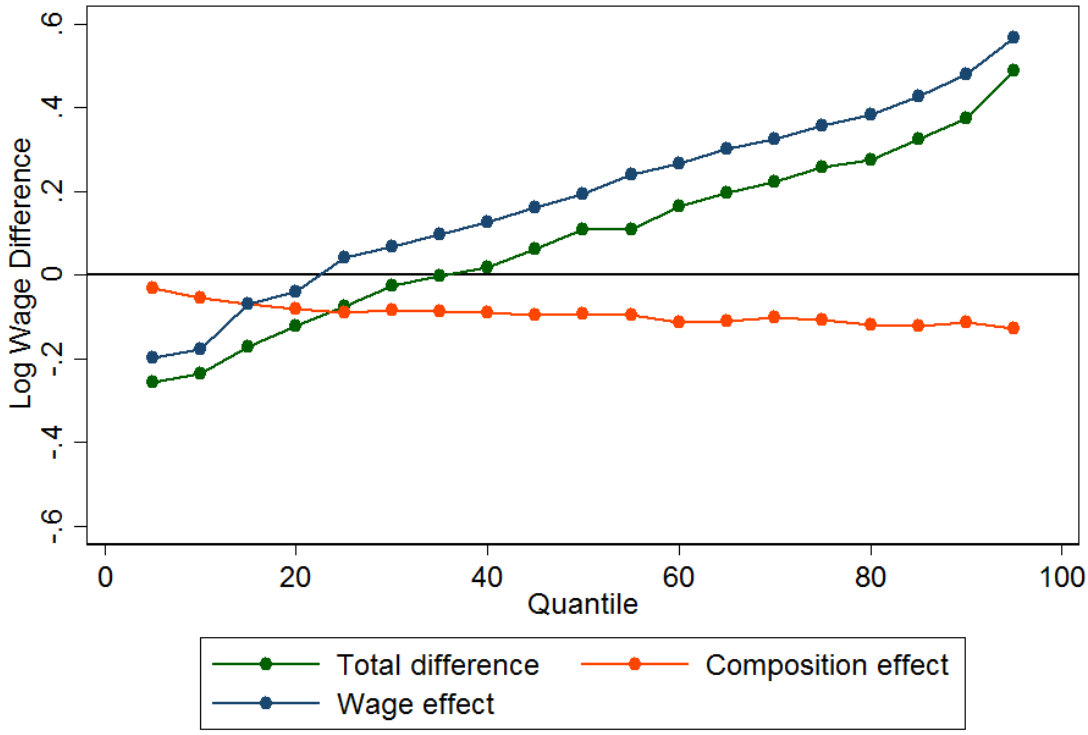


Figure 10: Estimated density of numeracy skills in selected countries, all and by education



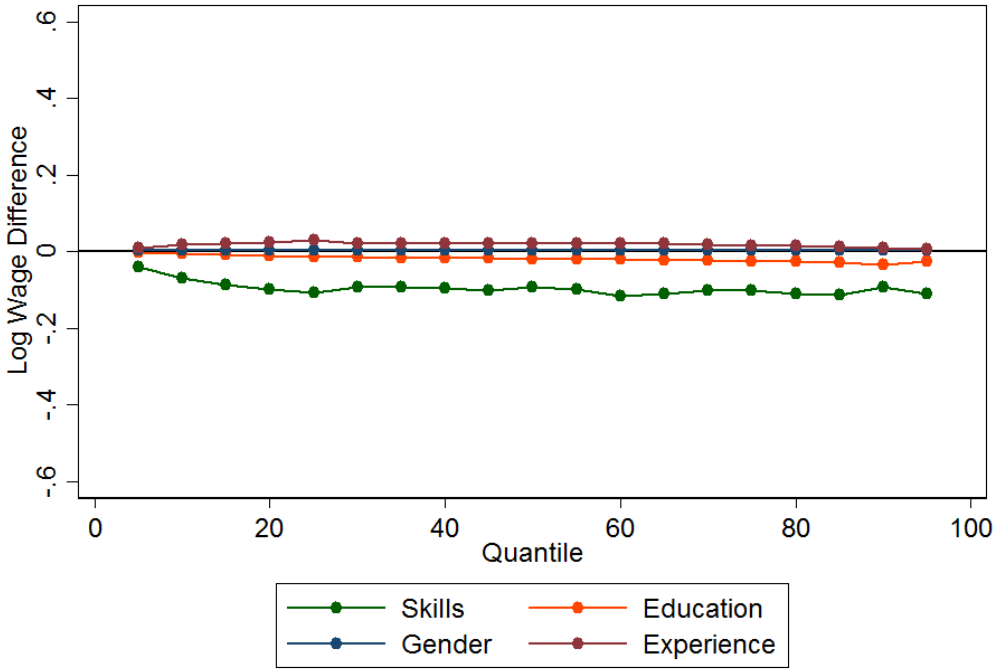
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 11: Decomposition of total log wage difference between Finland and the US into composition and wage effects[10]



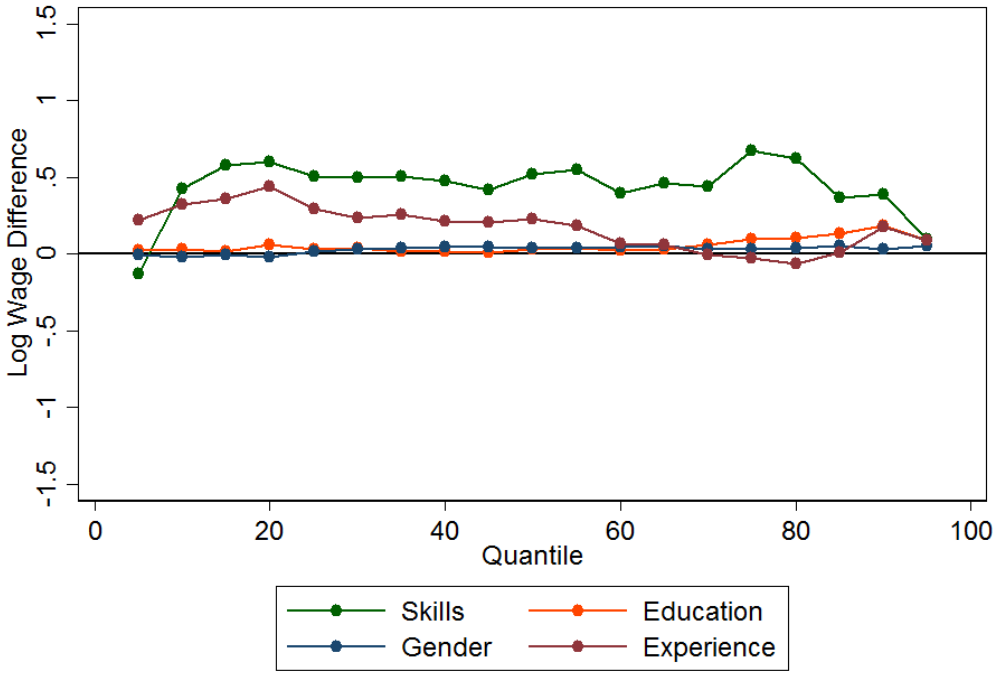
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 12: Detailed decomposition of the composition effect



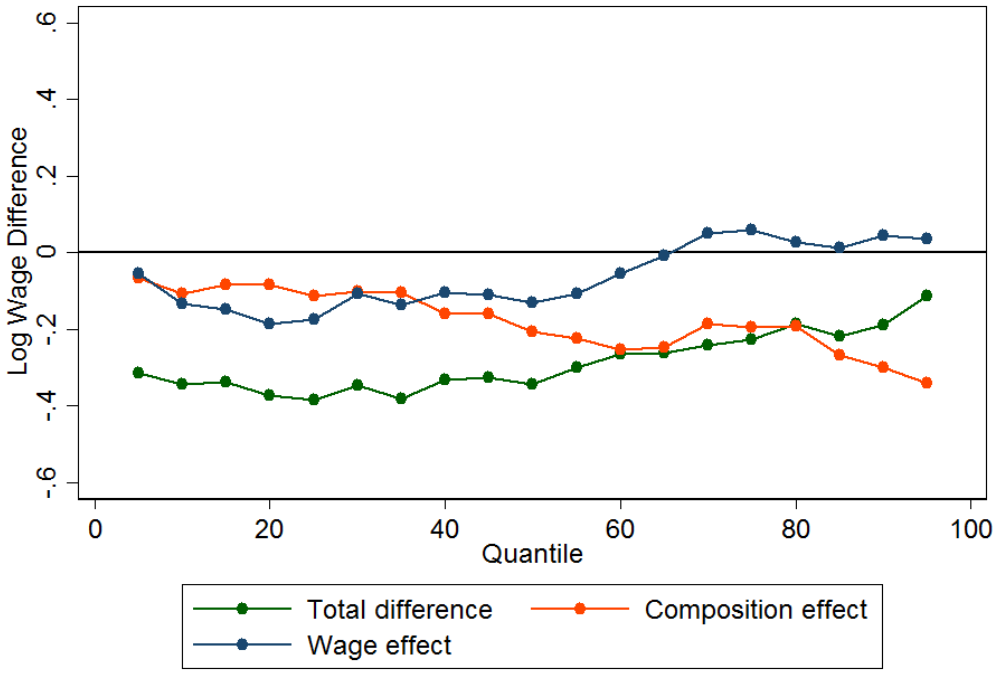
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 13: Detailed decomposition of the wage effect



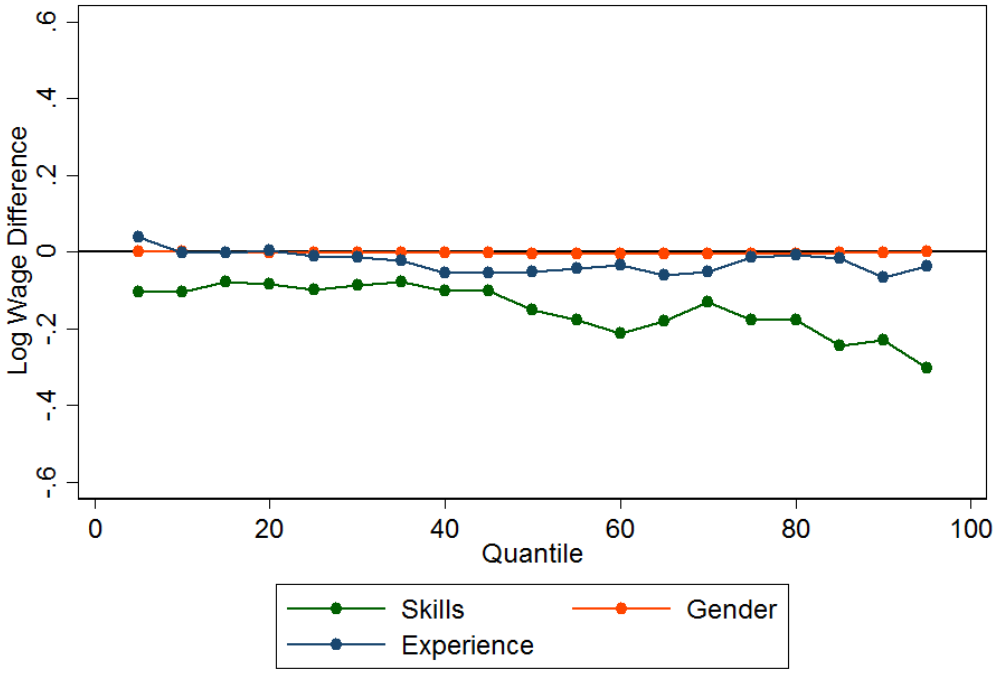
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 14: Decomposition of total log wage difference between low education workers in Finland and the US into composition and wage effects



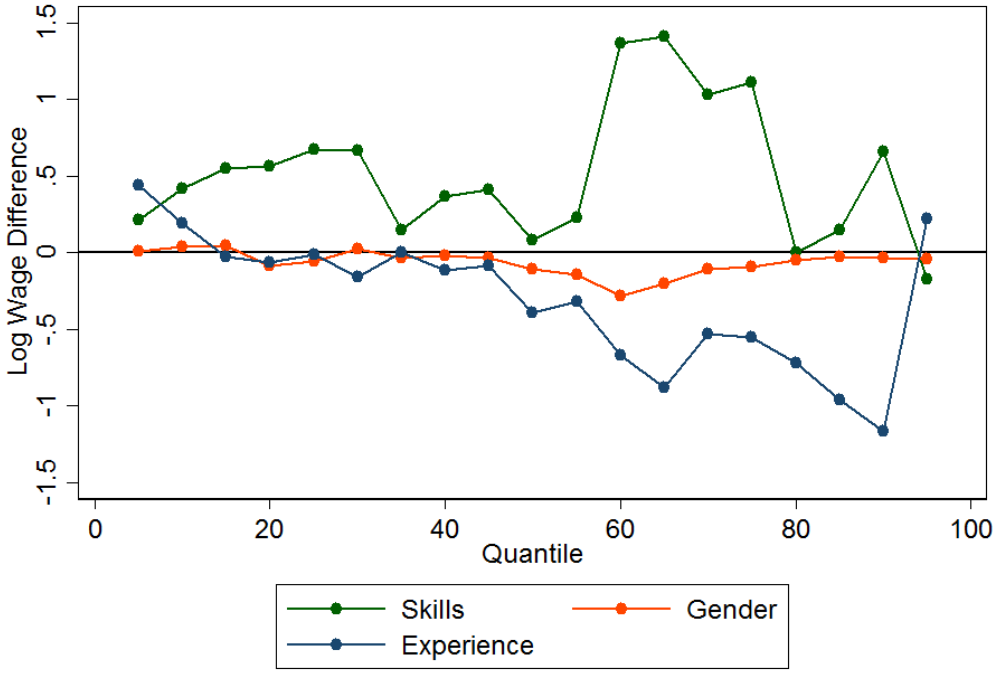
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 15: Detailed decomposition of the composition effect for low education workers



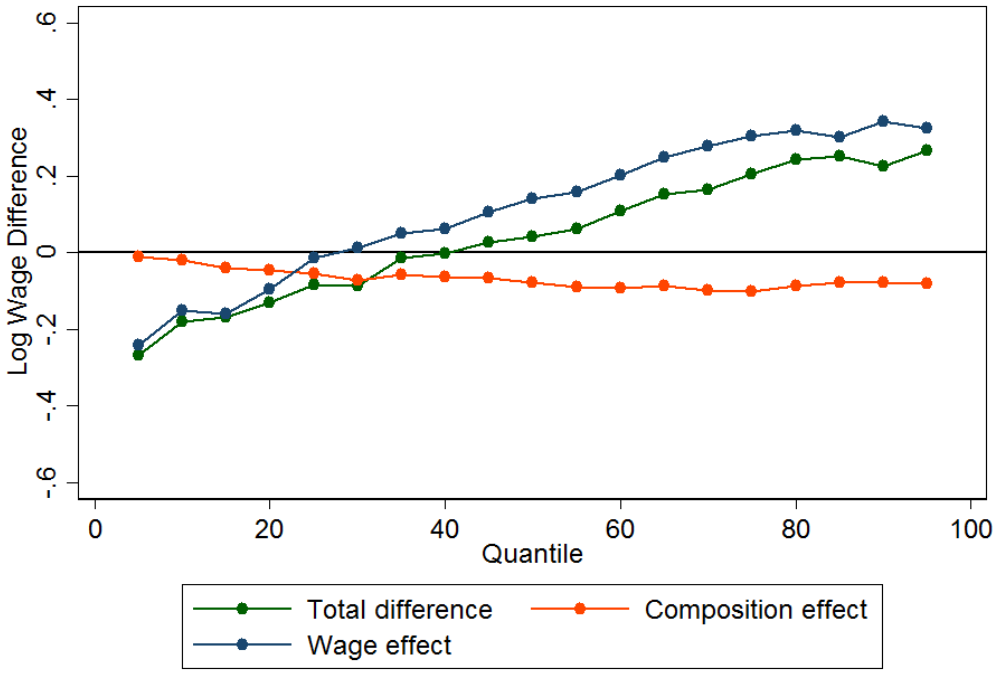
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 16: Detailed decomposition of the wage effect for low education workers



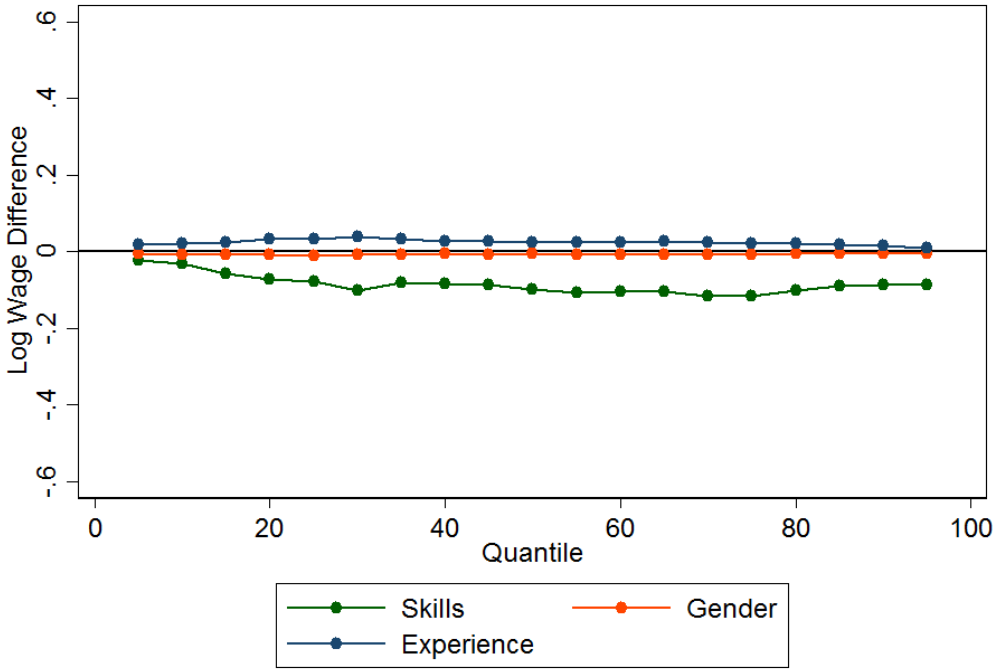
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 17: Decomposition of total log wage difference between intermediate education workers in Finland and the US into composition and wage effects



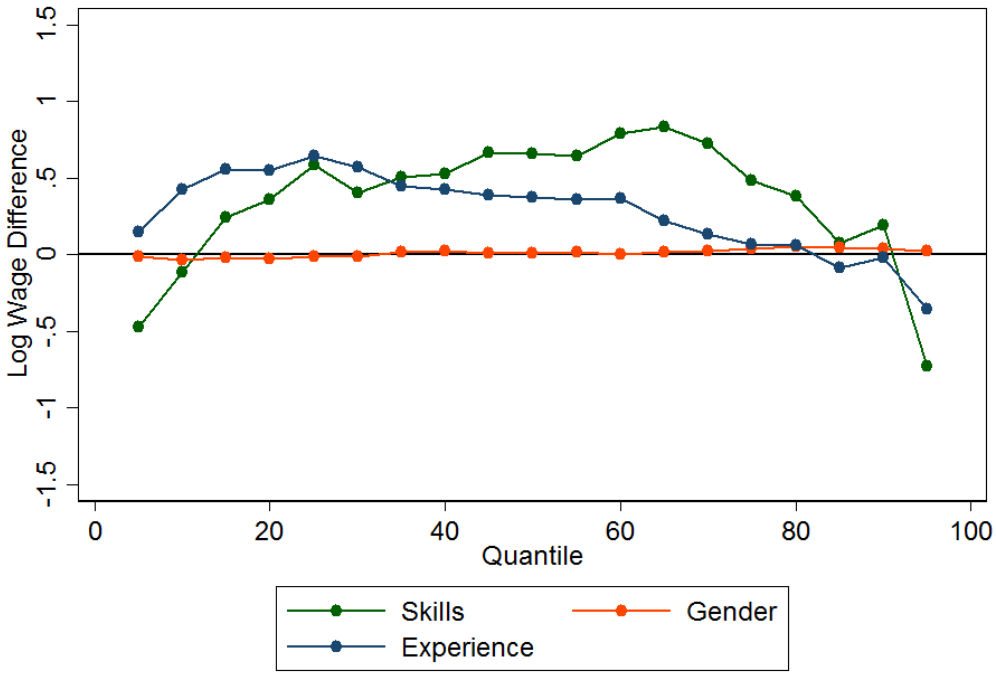
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 18: Detailed decomposition of the composition effect for intermediate education workers



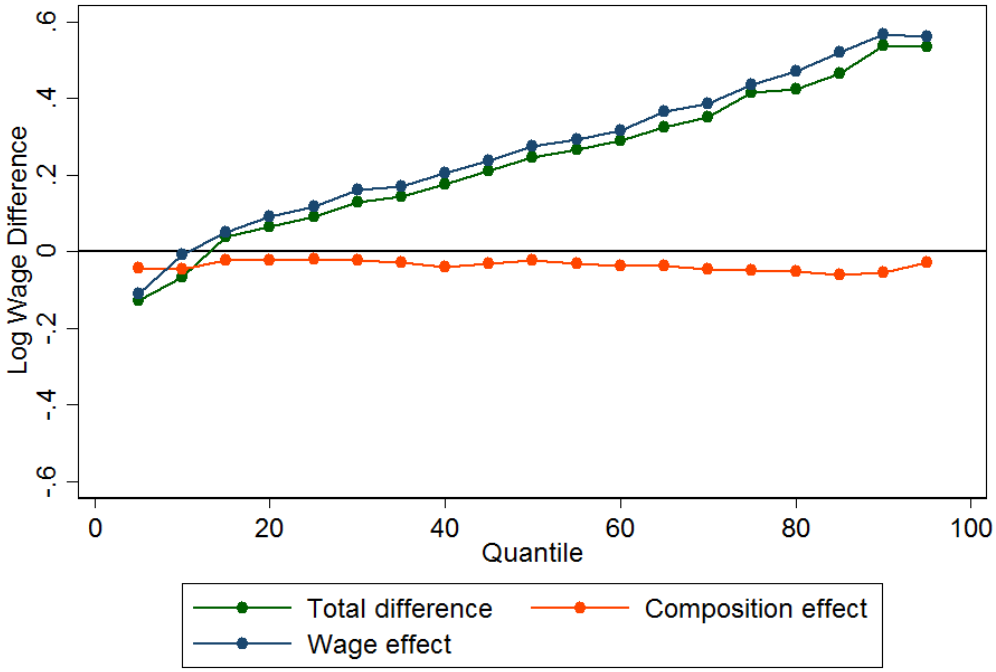
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 19: Detailed decomposition of the wage effect for intermediate education workers



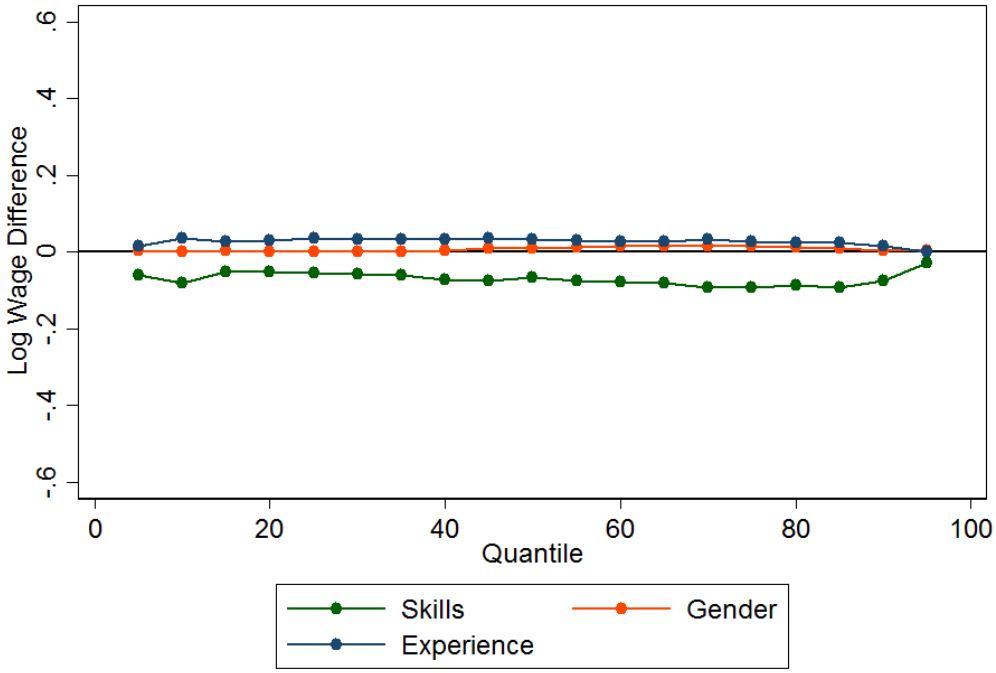
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 20: Decomposition of total log wage difference between high education workers in Finland and the US into composition and wage effects



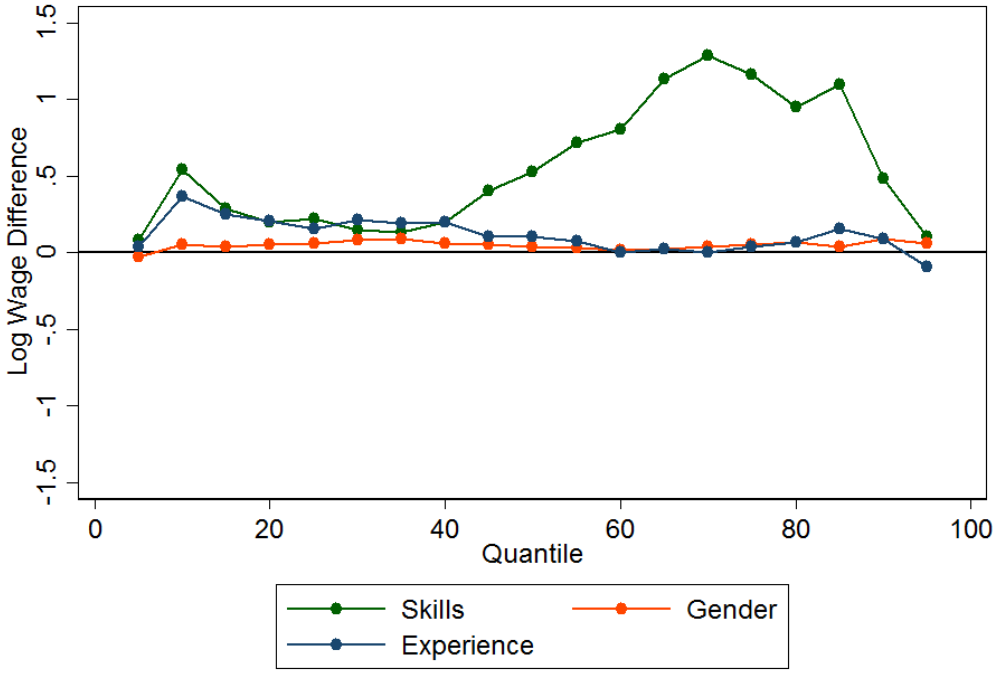
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 21: Detailed decomposition of the composition effect for high education workers



Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure 22: Detailed decomposition of the wage effect for high education workers



Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Table 1: Descriptive statistics

	Australia	Austria	Canada	Cyprus	Czech R.	Denmark	Eng./N.I.	Estonia	**Finland**	Flanders	France	Germany
Literacy	288.0	273.7	279.4	275.2	276.6	279.2	283.4	279.1	296.1	284.4	268.6	275.1
	(43.90)	(40.35)	(46.53)	(36.66)	(37.66)	(39.92)	(44.55)	(40.11)	(43.24)	(42.55)	(44.43)	(42.43)
Numeracy	276.5	281.9	273.0	273.9	279.2	289.9	276.1	277.7	292.2	292.0	264.5	281.3
	(48.58)	(45.37)	(50.50)	(41.54)	(40.30)	(43.76)	(49.43)	(40.70)	(44.62)	(44.84)	(51.13)	(45.93)
Problem Solving	291.4	285.8	284.1	.	281.8	287.5	287.9	275.6	291.3	285.4	.	284.7
	(35.60)	(34.35)	(41.74)	(.)	(42.92)	(37.28)	(40.03)	(39.71)	(38.19)	(38.80)	(.)	(40.66)
Hourly earnings PPP US$	20.05	17.36	21.71	15.75	8.753	24.61	19.92	9.483	19.05	19.90	15.21	20.24
	(9.135)	(7.067)	(9.783)	(8.832)	(3.892)	(7.762)	(10.58)	(5.897)	(6.775)	(7.107)	(6.050)	(9.284)
Experience	21.18	21.93	22.35	19.00	20.35	23.78	21.97	20.87	20.73	20.42	20.38	22.14
	(11.37)	(10.34)	(10.77)	(10.96)	(10.78)	(11.28)	(11.27)	(11.43)	(11.29)	(10.70)	(11.06)	(11.60)
Age	41.93	41.77	43.17	40.94	41.92	43.63	41.59	42.53	43.67	41.71	41.98	43.29
	(10.90)	(9.653)	(10.51)	(10.44)	(10.08)	(10.26)	(10.48)	(10.88)	(10.43)	(9.916)	(9.970)	(10.36)
Female	0.466	0.372	0.437	0.465	0.454	0.452	0.367	0.533	0.487	0.376	0.432	0.338
	(0.499)	(0.483)	(0.496)	(0.499)	(0.498)	(0.498)	(0.482)	(0.499)	(0.500)	(0.485)	(0.495)	(0.473)
Education:												
Low	21.0	12.1	8.1	11.8	6.8	14.2	17.4	9.0	9.2	10.7	18.3	6.4
Intermediate	36.0	65.4	34.0	45.3	72.1	38.6	36.8	44.8	39.2	44.6	46.2	55.4
High	43.0	22.6	57.8	42.9	21.1	47.2	45.7	46.2	51.6	44.7	35.5	38.2
Observations	3,824	1,875	11,545	1,692	1,992	3,395	2,995	3,194	2,583	1,898	2,802	2,068

Table 1: continued.

	Ireland	Italy	Japan	Korea	Nether-lands	Norway	Poland	Slovak Republic	Spain	Sweden	**US**	Pooled
Literacy	278.6	255.9	302.3	275.7	294.5	287.9	272.0	280.6	261.0	289.0	275.7	280.5
	(41.89)	(41.91)	(34.70)	(37.93)	(43.36)	(40.94)	(42.91)	(33.00)	(43.64)	(42.95)	(47.78)	(42.81)
Numeracy	271.8	258.3	297.5	268.9	293.5	292.0	267.8	286.1	258.0	289.9	262.4	279.4
	(45.38)	(46.36)	(39.44)	(41.21)	(44.72)	(46.72)	(44.66)	(38.35)	(45.05)	(47.20)	(54.76)	(46.85)
Problem Solving	283.9	.	301.2	284.1	293.5	291.0	273.2	282.1	.	290.5	280.8	286.7
	(36.53)	(.)	(39.59)	(33.80)	(37.07)	(36.19)	(44.67)	(34.13)	(.)	(40.76)	(41.73)	(39.15)
Hourly earnings PPP US$	22.76	14.27	16.79	15.01	22.06	25.76	8.735	8.371	13.75	18.88	23.10	17.64
	(10.98)	(6.806)	(10.14)	(9.758)	(8.614)	(8.394)	(4.865)	(5.410)	(6.916)	(5.560)	(14.21)	(9.730)
Experience	18.82	19.06	20.83	14.36	21.14	21.37	17.46	20.11	18.80	21.98	22.62	20.65
	(10.52)	(10.30)	(11.18)	(9.599)	(10.91)	(11.01)	(11.06)	(10.81)	(10.71)	(11.60)	(11.59)	(11.17)
Age	39.64	41.63	42.84	41.06	42.19	43.05	40.30	42.05	41.71	43.98	42.72	42.27
	(9.924)	(9.299)	(10.81)	(9.925)	(10.41)	(10.55)	(10.37)	(10.14)	(9.844)	(10.58)	(11.04)	(10.39)
Female	0.411	0.343	0.335	0.368	0.274	0.417	0.442	0.468	0.410	0.464	0.463	0.420
	(0.492)	(0.475)	(0.472)	(0.482)	(0.446)	(0.493)	(0.497)	(0.499)	(0.492)	(0.499)	(0.499)	(0.494)
Education:												
Low	14.5	43.2	9.1	12.3	21.0	15.1	5.5	8.1	34.3	13.4	7.7	13.7
Intermediate	34.4	41.6	40.2	39.1	37.4	36.2	57.8	66.7	23.9	48.3	46.4	44.9
High	51.1	15.3	50.7	48.6	41.6	48.7	36.8	25.2	41.8	38.3	45.9	41.4
Observations	1,777	1,444	2,307	2,431	1,696	2,446	2,091	2,069	1,868	2,231	2,057	62,28

Note: Means and standard errors (in parentheses) for test scores and selected variables. Sample share of individuals at different educational levels. The sample consists of individuals aged 25-65 who worked more than 30 hours per week (except for Australia, where part-time workers are included). Self-employed and individuals with missing education or experience are excluded. The wage measure is hourly earnings excluding bonuses, PPP corrected $US. In the pooled sample countries are reweighted to have equal weights. Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Table 2: Returns to numeracy skills

	Australia	Austria	Canada	Cyprus	Czech R.	Denmark	Eng./N. I.	Estonia	**Finland**	Flanders	France	Germany
Raw	0.181***	0.167***	0.196***	0.132***	0.153***	0.131***	0.226***	0.215***	0.142***	0.141***	0.150***	0.224***
	(0.00681)	(0.00817)	(0.00388)	(0.0120)	(0.00847)	(0.00521)	(0.00770)	(0.00942)	(0.00658)	(0.00734)	(0.00639)	(0.0104)
Plus controls (excl. education)	0.175***	0.176***	0.196***	0.141***	0.140***	0.133***	0.228***	0.189***	0.149***	0.164***	0.173***	0.237***
	(0.00678)	(0.00785)	(0.00381)	(0.0113)	(0.00856)	(0.00518)	(0.00771)	(0.00895)	(0.00658)	(0.00710)	(0.00619)	(0.0102)
Plus controls	0.121***	0.126***	0.151***	0.0909***	0.0873***	0.0907***	0.177***	0.147***	0.0958***	0.105***	0.104***	0.162***
	(0.00717)	(0.00825)	(0.00413)	(0.0112)	(0.00928)	(0.00538)	(0.00791)	(0.00956)	(0.00678)	(0.00801)	(0.00714)	(0.0113)
Within education												
Low	0.0768***	0.0916***	0.0928***	0.0573**	0.134***	0.0975***	0.146***	0.112***	0.0638***	0.0814***	0.0967***	0.0977***
	(0.0138)	(0.0194)	(0.0128)	(0.0268)	(0.0304)	(0.0127)	(0.0177)	(0.0284)	(0.0197)	(0.0190)	(0.0137)	(0.0368)
Intermediate	0.113***	0.143***	0.121***	0.0439**	0.0838***	0.0672***	0.155***	0.121***	0.0804***	0.106***	0.0987***	0.138***
	(0.0123)	(0.0105)	(0.00710)	(0.0171)	(0.0106)	(0.00866)	(0.0131)	(0.0142)	(0.0103)	(0.0113)	(0.0107)	(0.0146)
High	0.156***	0.108***	0.178***	0.155***	0.0773***	0.108***	0.208***	0.181***	0.116***	0.118***	0.119***	0.220***
	(0.0110)	(0.0184)	(0.00542)	(0.0175)	(0.0231)	(0.00795)	(0.0118)	(0.0139)	(0.00952)	(0.0137)	(0.0132)	(0.0191)
Observations	3,824	1,875	11,545	1,692	1,992	3,395	2,995	3,194	2,583	1,898	2,802	2,068

Table 2: continued.

	Ireland	Italy	Japan	Korea	Netherlands	Norway	Poland	Slovak R.	Spain	Sweden	**US**	Pooled
Raw	0.212***	0.126***	0.192***	0.230***	0.170***	0.143***	0.203***	0.212***	0.208***	0.110***	0.276***	0.190***
	(0.0113)	(0.0102)	(0.0103)	(0.0110)	(0.00884)	(0.00625)	(0.0104)	(0.0116)	(0.0105)	(0.00593)	(0.0105)	(0.00225)
Plus controls (excl. education)	0.225***	0.136***	0.190***	0.236***	0.191***	0.139***	0.210***	0.211***	0.220***	0.115***	0.269***	0.198***
	(0.0108)	(0.00964)	(0.00963)	(0.0104)	(0.00859)	(0.00612)	(0.0102)	(0.0113)	(0.0103)	(0.00582)	(0.0103)	(0.00222)
Plus controls	0.150***	0.0801***	0.142***	0.121***	0.122***	0.0990***	0.114***	0.139***	0.119***	0.0892***	0.170***	0.127***
	(0.0118)	(0.0102)	(0.0104)	(0.0116)	(0.00898)	(0.00648)	(0.0105)	(0.0116)	(0.0108)	(0.00630)	(0.0123)	(0.00237)
Within education												
Low	0.184***	0.0370**	0.113***	0.0178	0.118***	0.0672***	0.0116	0.0504	0.0955***	0.0804***	0.0894**	0.101***
	(0.0286)	(0.0153)	(0.0285)	(0.0275)	(0.0158)	(0.0150)	(0.0412)	(0.0368)	(0.0163)	(0.0151)	(0.0383)	(0.00549)
Intermediate	0.136***	0.103***	0.109***	0.0858***	0.101***	0.101***	0.0913***	0.136***	0.124***	0.0943***	0.155***	0.120***
	(0.0189)	(0.0155)	(0.0165)	(0.0181)	(0.0143)	(0.0105)	(0.0137)	(0.0139)	(0.0223)	(0.00942)	(0.0169)	(0.00350)
High	0.149***	0.145***	0.177***	0.199***	0.154***	0.110***	0.169***	0.190***	0.144***	0.0873***	0.211***	0.150***
	(0.0174)	(0.0273)	(0.0148)	(0.0178)	(0.0163)	(0.00950)	(0.0176)	(0.0252)	(0.0177)	(0.0100)	(0.0196)	(0.00385)
Observations	1,777	1,444	2,307	2,431	1,696	2,446	2,091	2,069	1,868	2,231	2,057	62,280

Note: Weighted least square estimates using sampling weights. The dependent variable is log hourly wages excluding bonuses, and the covariate of interest is standardized numeracy test score. The numeracy test score is standardized within country to have mean zero and standard deviation one. Additional control variables are education (low, intermediate, and high), gender, experience and experience squared. Standard errors in parentheses. Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Table 3: Reweighted RIF-regression based decomposition of US-Finland differences in wage inequality

	90/10	90/50	50/10	Variance	Gini
A. Full sample					
US	1.4540	0.7767	0.6773	0.3048	0.1053
Finland	0.8420	0.5088	0.3332	0.1096	0.0647
Difference	0.6121	0.2680	0.3441	0.1952	0.0406
Composition effect	-0.0596	-0.0200	-0.0396	-0.0191	-0.0005
Skills	-0.0253	0.0008	-0.0261	-0.0077	0.0016
Education	-0.0278	-0.0123	-0.0155	-0.0081	-0.0008
Specification error	0.0156	0.0025	0.0132	0.0126	0.0025
Wage effect	0.6553	0.2851	0.3701	0.2016	0.0387
Skills	-0.0394	-0.1308	0.0913	0.0350	-0.0169
Education	0.1491	0.1527	-0.0036	0.0425	0.0023
Reweighting error	0.0008	0.0004	0.0004	0.0001	0.0000
B. Low education					
US	0.8753	0.5432	0.3321	0.1289	0.0818
Finland	0.7197	0.3886	0.3312	0.0831	0.0593
Difference	0.1556	0.1546	0.0010	0.0458	0.0225
Composition effect	-0.1919	-0.0914	-0.1006	-0.0095	-0.0031
Skills	-0.1270	-0.0793	-0.0476	-0.0101	-0.0020
Specification error	0.1682	0.0724	0.0958	0.0228	0.0108
Wage effect	0.0070	0.1734	0.0047	0.0348	0.0154
Skills	0.2374	0.5702	-0.3327	-0.1413	-0.0587
Reweighting error	0.0011	0.0001	0.0010	-0.0024	-0.0006
C. Intermediate education					
US	1.1666	0.6214	0.5452	0.2068	0.0921
Finland	0.7612	0.4390	0.3222	0.0836	0.0580
Difference	0.4054	0.1824	0.2230	0.1232	0.0341
Composition effect	-0.0571	0.0023	-0.0594	-0.0139	-0.0013
Skills	-0.0545	0.0108	-0.0653	-0.0116	-0.0002
Specification error	-0.0297	-0.0202	-0.0096	-0.0002	0.0003
Wage effect	0.4916	0.1999	0.2917	0.1372	0.0350
Skills	0.3023	-0.4670	0.0002	0.0326	-0.0172
Reweighting error	0.0006	0.0004	0.7694	0.0002	0.0000

Table 3: continued.

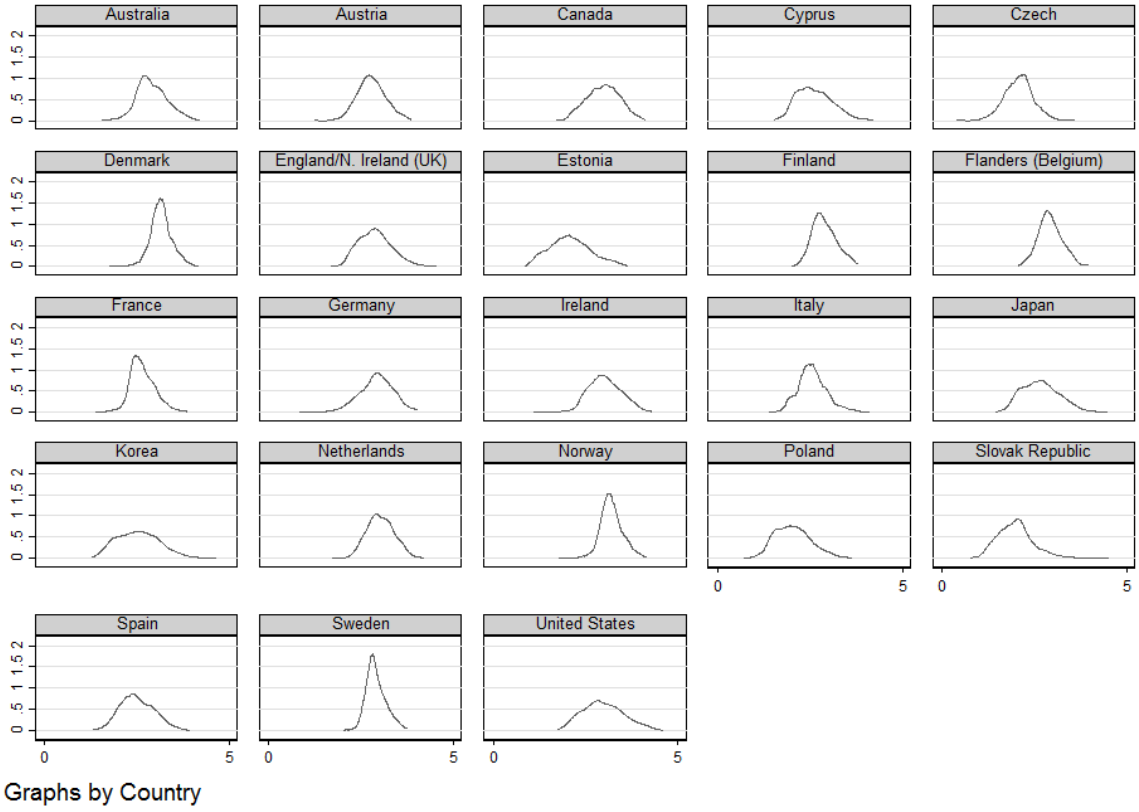
D. High education					
US	1.4617	0.7778	0.6839	0.2847	0.0937
Finland	0.8569	0.4839	0.3729	0.1042	0.0610
Difference	0.6049	0.2939	0.3110	0.1805	0.0328
Composition effect	-0.0087	-0.0322	0.0235	-0.0035	0.0002
Skills	0.0076	-0.0067	0.0144	-0.0008	0,0015
Specification error	0.0357	0.0315	0.0042	0.0013	0.0002
Wage effect	0.5770	0.2938	0.2832	0.1825	0.0322
Skills	-0.0624	-0.0420	-0.0204	0.1486	0,0219
Reweighting error	0.0008	0.0007	0.0001	0.0002	0.0001

Note: The dependent variable is log hourly wage. The control variables included are numeracy test score, education dummies (intermediate education omitted), gender, experience and experience squared. US is the base country. The decomposition is conducted by first estimating influence functions for the percentiles, the variance and the Gini coefficient separately for the two countries. Then a standard Oaxaca-Blinder decomposition if performed by replacing the dependent variable by the estimated influence function for the statistic of interest. To obtain the decomposition of the percentile differences, the decomposition at the percentiles in question are compared.
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Appendix

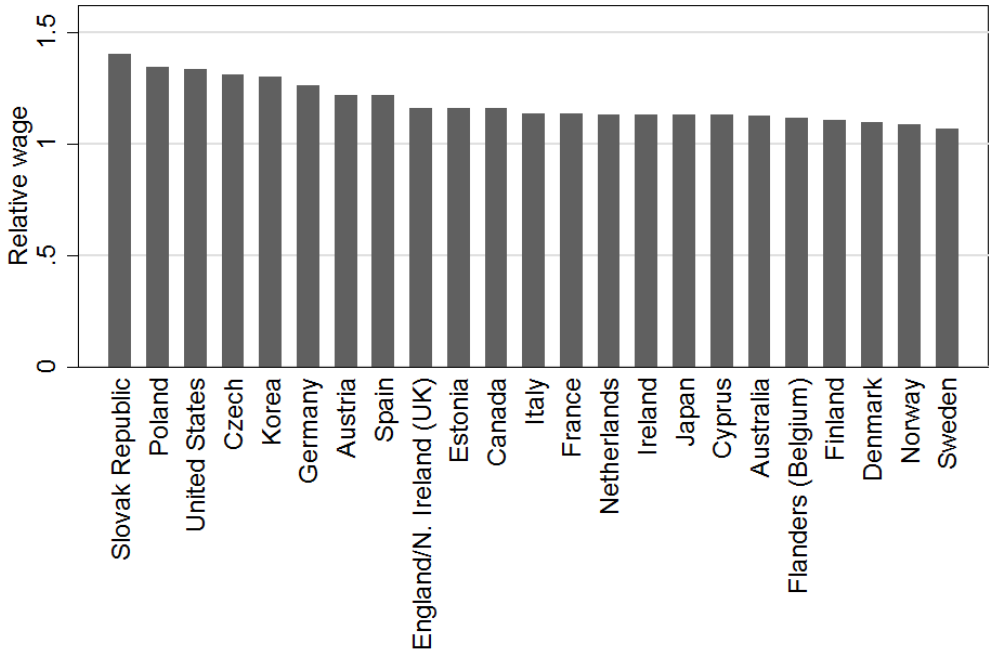
2.A Figures

Figure A1: Estimated log wage densities by country



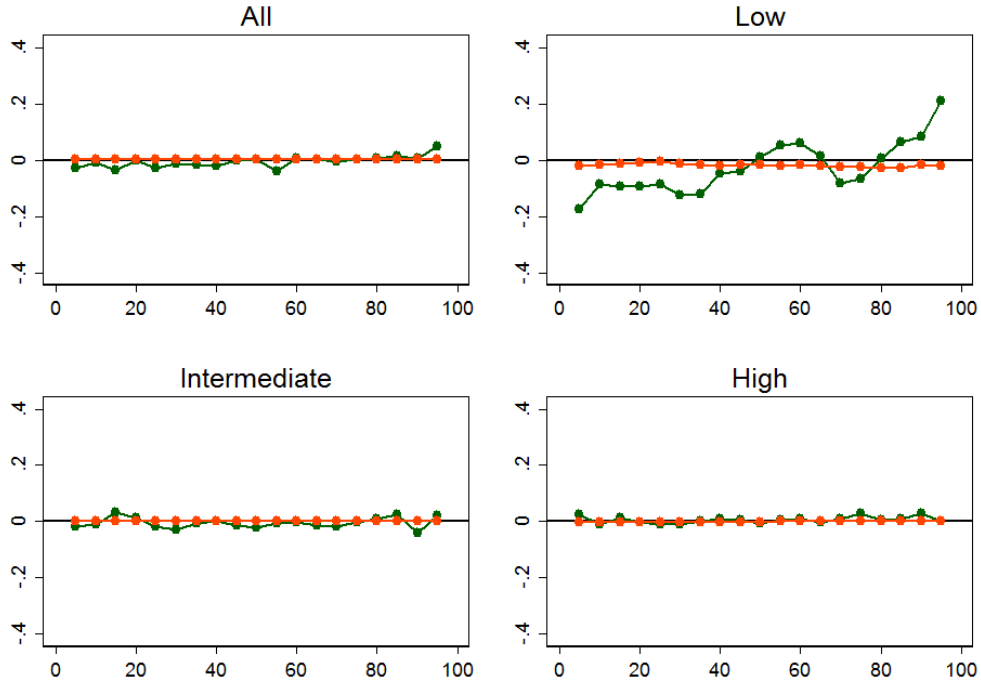
Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Figure A2: Relative log wages of high to low educated workers



Source: Survey of Adult Skills (PIAAC) (OECD 2013a)

Figure A3: Specification and reweighting error in full sample and by education



Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

2.B Tables

Table A1: Percentage of workers scoring at each proficiency level in numeracy

Proficiency level	Below 1	1	2	3	4	5
Score	0-176	176-225	226-275	276-325	326-375	376-500
Australia	2.78	11.07	33.17	37.87	14.26	0.86
Austria	2.34	7.83	31.94	41.66	15.40	0.84
Canada	3.48	14.65	30.48	37.18	13.48	0.74
Cyprus	2.05	9.83	36.70	42.53	8.70	0.19
Czech	1.03	8.31	35.62	44.03	10.25	0.77
Denmark	1.30	6.12	26.11	46.17	19.10	1.19
England/N. Ireland (UK)	2.68	13.52	30.47	37.43	15.08	0.83
Estonia	1.17	9.10	37.00	41.61	10.62	0.50
Finland	1.15	5.58	27.93	41.73	22.09	1.52
Flanders (Belgium)	1.51	6.67	24.30	44.67	21.59	1.26
France	5.72	16.00	34.09	33.86	9.83	0.49
Germany	1.18	11.38	30.91	40.00	15.69	0.84
Ireland	2.76	11.85	36.96	37.34	10.76	0.34
Italy	4.30	20.02	38.06	31.54	6.00	0.08
Japan	0.50	4.26	22.07	49.13	22.84	1.20
Korea	2.45	11.40	38.71	40.99	6.34	0.11
Netherlands	1.47	7.04	22.07	46.69	21.85	0.88
Norway	2.11	5.91	24.06	44.99	21.58	1.35
Poland	2.82	14.27	38.80	34.78	8.81	0.52
Slovak Republic	1.06	5.65	30.33	49.59	12.83	0.55
Spain	4.86	17.44	41.31	31.45	4.89	0.05
Sweden	1.91	7.19	25.64	42.98	20.61	1.67
United States	6.97	17.06	32.53	32.50	10.26	0.68

Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Table A2: Percentage of workers scoring at each proficiency level in numeracy

	Proficiency level					
	Below 1	1	2	3	4	5
Finland						
All	1.15	5.58	27.93	41.73	22.09	1.52
Low	3.61	17.71	46.19	26.86	5.64	0.00
Intermediate	1.47	7.85	38.82	40.45	10.91	0.49
High	0.47	1.69	16.39	45.35	33.52	2.58
United States						
All	6.97	17.06	32.53	32.50	10.26	0.68
Low	47.21	30.46	20.96	1.37	0.00	0.00
Intermediate	6.80	24.94	42.04	23.32	2.56	0.34
High	0.40	6.84	24.83	47.01	19.78	1.14

Source: Survey of Adult Skills (PIAAC) (OECD, 2013a).

Bibliography

Autor, D. H., Katz, L. F., & Kearney, M. S. (2008). Trends in US wage inequality: Revising the revisionists. *The Review of Economics and Statistics, 90*(2), 300-323.

Bedard, K., & Ferrall, C. (2003). Wage and test score dispersion: some international evidence. *Economics of Education Review, 22*(1), 31-43.

Blau, F. D., & Kahn, L. M. (1996). International Differences in Male Wage Inequality: Institutions versus Market Forces. *Journal of Political Economy, 104*(4), 791-837.

Blau, F. D., & Kahn, L. M. (2005). Do Cognitive Test Scores Explain Higher U.S Wage Inequality? *The Review of Economics and Statistics, 87*(1), 184-193.

Checchi, D., & Van de Werfhorst, H. G. (2014). Educational policies and income inequality. *IZA Discussion Papers No. 8222*.

Devroye, D., & Freeman, R. (2002). Does Inequality in Skills Explain Inequality of Earnings Across Advanced Countries? *National Bureau of Economic Research Working Paper Series, No. 8140*.

DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica, 64*(5), 1001-1044.

Firpo, S., Fortin, N., & Lemieux, T. (2007). Decomposing Wage Distributions Using Recentered Influence Function Regressions. *mimeo*.

Firpo, S., Fortin, N., & Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica, 77*(3), 953-973.

Firpo, S., Fortin, N., & Lemieux, T. (2011). Occupational Tasks and Changes in the Wage Structure. *IZA Discussion Papers No. 5542*.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition Methods in Economics. In A. Orley & C. David (Eds.), *Handbook of Labor Economics* (Vol. Volume 4, Part A, pp. 1-102): Elsevier.

Freeman, R., & Schettkat, R. (2001). Skill Compression, Wage Differentials, and Employment: Germany vs the US. *Oxford Economic Papers, 53*(3), 582-603.

Gregorio, J. D., & Lee, J. W. (2002). Education and Income Inequality: New Evidence From Cross-Country Data. *Review of Income and Wealth, 48*(3), 395-416.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review, 73*(C), 103-130.

Juhn, C., Murphy, K. M., & Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 410-442.

Leuven, E., Oosterbeek, H., & Ophem, H. v. (2004). Explaining International Differences in Male Skill Wage Differentials by Differences in Demand and Supply of Skill. *The Economic Journal, 114*(495), 466-486.

Machado, J. A. F., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics, 20*(4), 445-465.

OECD. (2013a). OECD Skills Outlook 2013: First Results from the Survey of Adult Skills. *OECD Publishing*.

OECD. (2013b). Technical Report of the Survey of Adult Skills (PIAAC). *OECD Publishing*.

Paccagnella, M. (2015). Skills and Wage Inequality: Evidence from PIAAC. *OECD Education Working Papers, No. 114*.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics, 33*(3), 1065-1076.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics, 27*(3), 832-837.

Chapter 3

Gender Segregation in the Welfare State

Industrial Segregation in Norway 1970-2009

Abstract

The aim of this study is to provide an overview of the trends in industrial gender segregation in Norway between 1970 and 2009. This is motivated by the observation that the Norwegian labor market is considered to be one of the most gender equal in the world, while at the same time one of the more gender segregated. This paper is purely descriptive, but it relates the observed changes in segregation to changes in institutional and economic conditions. It uses data from 1970-2009, covering the period when female labor force participation increased the most and when the public service sector was developed. This study is also one of the few that covers the whole economy, and it uses industry data reported on a more detailed level than previous studies. This allows me to more precisely identify the industries that have contributed to changes in segregation. I find that the level of segregation, as measured by the dissimilarity index (Duncan and Duncan, 1955), has remained fairly stable over time. Decompositions of the changes over time suggest, however, that the gender composition within industries has become more equal over time, which contributed to reducing segregation, while the relative size of segregated industries increased, which contributed to increasing segregation. In particular, employment growth in female dominated industries such as health care, child care and care for the elderly and disabled has been important drivers of segregation over time. When the sample is divided into groups based

on educational attainment, I find that the observed trends in segregation differ. In particular, workers with secondary education experienced a fast increase in the level of segregation after 1990, while the level of segregation among workers with tertiary degrees decreased rapidly especially in the 1970s and 1980s. In all groups, there were signs of slower desegregation or increasing segregation from the late 1990s onwards, as within industry improvements in gender composition subsided and employment in gender segregated industries increased.

3.1 Introduction

The Norwegian labor market is among the most gender equal in the world. For instance, Norway has repeatedly scored very well on the UN's Gender Inequality Index,[1] and the female labor force participation rate is one of the highest in the world. Still, the Norwegian labor market has long been among the most gender segregated, meaning that i) Norwegian men and women work in different industries and occupations to a large extent and ii) women are less likely to hold managerial and other influential positions.

In the literature, the parallel existence of these two phenomena has been referred to as the welfare state paradox, since it is a pattern that is common to many advanced welfare states. Mandel and Semyonov (2006) argue that welfare states are successful in increasing female labor force participation, but that they create sheltered labor markets for women that enable them to combine family responsibilities and working life, but prevent them from competing with men for powerful positions. Family-friendly policies may discourage employers from hiring women in important positions while at the same time influencing women's employment preferences towards family-friendly occupations with convenient working conditions.

The aim of this study is to provide an overview of the changes in gender segregation in Norway between 1970 and 2009, and to assess them in the light of the emerging welfare state. The expansion of the public service sector started in the 1960s, and was extended to cover health care, care for the elderly, education, and social security among other things. This development had an enormous impact on the Norwegian labor market, and especially on the employment of women. This paper contributes to the literature in four ways. First, I use a

[1]The Gender Inequality Index measures gender inequalities along the dimensions of reproductive health, empowerment and labor market status (UNDP, 2010, 2011, 2013, 2014)

longer time frame than any previous Norwegian study, and can thereby identify trends in a longer perspective. Second, the study also uses more detailed data on industry than previous studies, while it still covers the whole economy. In doing so, I can identify more narrowly defined industries that have been important in driving changes in segregation than before. Third, I study whether the level and trend in segregation differ between skill groups, which also helps to understand the changes in segregation in Norway better. As in most industrialized countries, educational attainment started to increase rapidly in the 1970s, especially among women. Norwegian females have increased their formal qualifications considerably relative to males, and this has changed women's opportunities to compete with men in the labor market. I study whether this also changed the patterns in gender segregation, which is an under-researched area in the literature. Fourth, I discuss segregation in the light of the historical context. This study is purely descriptive, and the aim is not to explain why there is gender segregation in the labor market in the first place. Rather, the objective is to study the observed trends in segregation in Norway using the historical context, and to assess how institutional factors may have contributed to changes in the observed level of segregation.

While most of the segregation literature has focused on occupational segregation, I study industrial segregation. One motivation for this is that I link the findings to the changes in the industry structure and in particular to the expansion of the public sector. The drawback of studying industrial instead of occupational segregation is that industrial segregation is likely to hide some degree of occupational segregation if males and females within the same industry have different occupations, i.e., males are physicians and females are nurses. Nonetheless, studying industrial segregation is informative, and as long as there is industrial segregation, there is also occupational segregation. Furthermore, by studying industrial segregation within skill groups, it is possible to draw some suggestive conclusions about occupational segregation.

I start the analysis by documenting the increase in female labor force participation, the expansion of the public sector and the change in the industry structure that followed, the introduction of the generous family policies as well as the changes in educational attainment. I then study the trends in gender segregation in Norway between 1970 and 2009, using Norwegian Census and register data that contains detailed information on employment industry for the Norwegian population. I focus on horizontal segregation, which means the extent to which males and females work in different sectors, industries or occupations.

This is not to be confused with vertical segregation, which is concerned with how males and females are placed in different places in the hierarchy of organizations. I mainly rely on the dissimilarity index (Duncan and Duncan, 1955), and decompose the changes in this index to identify the relative importance changes in within-industry gender composition and changes in the relative size in industries and identify the specific industries that have contributed the most to the changes in segregation over time. I then divide the sample into four groups based on educational attainment and study segregation within each of these groups.

I find that throughout the period, males and females became more evenly distributed within industries, although this slowed down after 1990 when the female labor force participation rate stabilized. Further, changes in the industry composition played an important role in the development of gender segregation over time. The expansion of the health and welfare sector was very important in driving segregation, as it absorbed almost 50 percent of all female labor market entrants since 1970, and employed almost 40 percent of all female workers in 2009. The expansion of child care services and care for the aged and disabled were the main the drivers of between sector segregation, especially after 1990.

The story of segregation in the Norwegian labor market is primarily a story about female workers, but men have also played an important role. In the 1980s, downsizing of male dominated industries, such as agriculture and manufacturing, counteracted the upward pressure that the expansion of female dominated service industries put on segregation. In later years, male employment increased in male dominated industries such as business activities, mining and quarrying and construction, which lead to more segregation.

The extent of gender segregation also varied between educational groups. Workers with secondary education experienced a rapid increase in the level of segregation after 1990. This was to a large extent driven by relative employment growth in segregated industries such as child care, care for the aged and disabled, construction and transportation. Workers with tertiary degrees (both short and long) experienced a large reduction in the level of segregation between 1970 and 1998, and it was mainly driven by within industry changes in gender composition. Short tertiary degree holders were more concentrated in public sector industries than other educational groups, while public sector industries were slightly less important for the changes in segregation in the long tertiary education group. In the 2000s, there were signs of increased segregation among workers with long tertiary degrees, mainly because the female (male) employment share increased in

a number of female (male) dominated industries, such as general somatic hospitals, higher education and veterinary services (software consultancy and supply).

The paper is structured as follows. I start in Section 3.2 by defining the concept of segregation and discussing its causes and consequences, before I review the existing literature on gender segregation in Norway in Section 3.3. I describe the data and discuss the methods used in the empirical analysis in Sections 3.4 and 3.5. In Section 3.6, I present the institutional background. In particular, I discuss the industry composition in the Norwegian economy, the expansion of the pubic sector, changes in the education system and educational attainment and changes in labor force participation. The main part of the empirical analysis is presented in Section 3.7, and in Section 3.8, I study whether different skills groups experienced differential trends in segregation over time. Finally, in Section 3.9, I summarize the findings and conclude.

3.2 Segregation - definition, causes and consequences

In this section I discuss the definition of segregation, as well as the causes and consequences of segregation that are often put forward in the literature. There is an enormous literature that discusses the causes and consequences of segregation and the purpose of this section is not to give a complete overview of this literature, but simply to give a brief introduction of the concept of segregation.[2]

3.2.1 Definition

The concept of gender segregation has two dimensions, horizontal and vertical segregation. By horizontal segregation we mean the tendency for males and females to work in different industries.[3] When males and females are employed in completely different industries, there is complete segregation and when the share of female workers in every industry is identical to the share of females in the labor force, there is perfect integration.

[2]For a more detailed discussion, see for example Anker (1998); Reskin and Bielby (2005). For a summary of the empirical evidence of the causes and consequences of segregation, see Blau, Ferber, and Winkler (2013).

[3]Since the focus of this study is industrial segregation, the following discussion will only focus on industries, but the word industry could be replaced with occupation or sector throughout the discussion.

The term vertical segregation refers to the distribution of male and female workers in the hierarchy of jobs. Males tend to be overrepresented in jobs with higher status, such as positions with leadership responsibilities (see e.g., International Labour Organization (2015)).

3.2.2 Causes of segregation

Economists and sociologists have debated the causes of sex segregation in the labor market since the 1970s. The explanations are often divided into supply side explanations that focus on the choices and characteristics of individuals, and demand side explanations that focus on the employer side of the labor market and institutions.

Occupational and industrial segregation in the labor market is closely linked to gender differences in educational attainment. Historically, gender differences in educational attainment have been large, but in the last few decades, the gender patterns in educational attainment have changed. Since the 1970s, female educational attainment in particular has increased rapidly and females have overtaken males in human capital investments in many countries. But while the gender gap in educational attainment is now small, males and females still differ in their choices of field of study. Males are overrepresented in STEM fields,[4] which are associated with higher earnings potentials, and females are overrepresented in softer sciences, such as humanities and arts, teaching and health related subjects. These differences have implications for the occupational choices of males and females, and contribute to gender segregation in the labor market.

Human capital theory explains gender differences in educational attainment and occupational choice as a result of rational utility maximizing behavior. One prediction is that females invest less in human capital than males because they expect shorter working careers with more and longer periods out of the labor force (Polachek, 1981; Blau et al., 2013). This explanation was more helpful in explaining why women were less likely to invest in higher education in the past, but is less useful today when the educational attainment of females exceeds that of males in many countries. These days, the gender difference in education is largely a question of field of study, and in this case Polachek (1978) offered a more relevant explanation. He suggested that females, who expect more time out of the labor force, avoid fields such as engineering and science where scientific

[4]STEM is an acronym referring to the academic disciplines of science, technology, engineering, and mathematics.

and technological progress is fast, because this makes breaks in the working life more costly. Another more traditional explanation was offered by Becker (1985), who suggested that females choose careers that are more flexible with respect to effort and working hours because they are easier to combine with child bearing and household work.

Economists tend to believe that decisions regarding human capital investments are based on individual preferences and voluntary choices, without further discussion of where these preferences come from. Sociologists, on the other hand, tend to go further in analyzing the formation of preferences. In particular, sociologists often highlight the importance of socialization processes, or societal discrimination, in preference formation (Blau et al., 2013). Socialization can take many forms, but are in general social influences from family, friends or society that lead females (and males) into associating themselves with gender typical traits and making gender typical choices with respect to education, career or family situation that will adversely affect their prospects in the labor market. Socialization often starts early in life, for example through encouraging boys and girls to play with different toys or through encouraging different kinds of behavior (England, 1992; Blau et al., 2013).

The list of supply side explanations of segregation is long but a few more explanations are worth mentioning. For instance, Blau et al. (2013) discuss access to on-the-job training and psychological attributes, such as attitudes towards negotiating and risk and competitiveness (see also Bertrand (2011)).

On the demand side, discrimination is probably one of the most important explanations for segregation in the labor market. Labor market discrimination means that two individuals with equal qualifications or skills are treated differently only because of their gender, age or race (Becker, 1957; Arrow, 1973). Discrimination can take many forms but I focus on taste-based and statistical discrimination. Taste-based discrimination means an employer's, co-worker's or customer's unwillingness to associate with a particular group (Becker, 1957). Examples of taste-based discrimination are employers who are unwilling to employ female construction workers while not having a problem employing a female secretary, or females who are reluctant to buy lingerie, but not electronics, from a male sales clerk. If the reluctance to interact with females is instead based on the perception that women are on average less qualified to pursue a particular job, it is a case of statistical discrimination (Phelps, 1972; Aigner and Cain, 1977). The theory of statistical discrimination relies on the assumption that employers make

hiring decisions in environments with imperfect information and therefore base their decisions on average characteristics of certain groups such as gender or race.

Today, anti-discrimination laws prohibit open discrimination against workers with certain characteristics. However, discrimination against women is still likely to exist in less obvious or conscious forms. Blau et al. (2013) discuss subtle barriers to education and certain jobs. For example, the lack of female role models might prevent females from entering certain study programs or professions. Similarly, mentor-protégé relationships are more likely to form between two males or two females, and this may adversely affect career prospects of females in male dominated industries. Informal networks in workplaces can be important sources of information, skills and support. In male dominated work places, female workers might be excluded from these networks, which could affect their skill acquisition, chances of promotion or similar.

Institutional settings can also cause segregation in the labor market. Access to child care, parental leave schemes and income taxation rules are examples of institutional settings that may affect the employment choices of men and women differently.

While both the supply and demand explanations have been found to have empirical support (Blau et al., 2013), the literature on segregation has been criticized for taking a too narrow approach to explaining segregation as it has often focused on only one potential explanation at a time. Not very often have supply and demand explanations been studied at the same time, or in a changing environment. Blackburn et al. (2002) argue that in order to explain the patterns of gender segregation, one must study the larger context, social and economic, in which the present situation has developed. In particular, the authors point out three processes that have influenced gender segregation and inequality: the expansion of the education system, changes in the occupational structure and the increase in female labor force participation, especially among married women with children.

3.2.3 Consequences of segregation

Researchers and policy makers mainly care about segregation because of its impact on the inequality between males and females. Most importantly, gender segregation is an important predictor of the gender wage gap. Numerous empirical studies have shown that there is a negative relationship between the female share in an occupation and and the wage in an occupation (Boraas, 2003; Bayard et al.,

2003; Levanon et al., 2009; Blau et al., 2013).

However, the problem stretches further than the gender wage gap, as there are a number of nonpecuniary consequences (Anker, 1998; Bettio and Verashchagina, 2009; Jensberg et al., 2012; Blau et al., 2013). There are generally better career prospects in male dominated occupations, and male dominated occupations often have a larger degree of responsibility and autonomy. Part-time work (both voluntary and involuntary) and non-standard contracts are more common in female dominated industries, such as health care.

Persistent gender segregation is likely to reinforce stereotypes of male and female preferences, skills, roles and status in society. This may in turn make it even harder for females, and males, to enter occupations or industries that are gender atypical (Anker, 1998). In addition, a lot of talent is likely to be lost if females are limited to doing typically female work and vice versa.

A high level of segregation also makes an economy less efficient in adjusting to changes in the industry structure, especially if the changes mainly affect highly segregated industries (Anker, 1998). An example is a situation where the demand for manual manufacturing workers (mainly males) decreases while the demand for service workers in sales or health care (typically females) increases. Readjustment to such changes in the industry structure is easier if there is less gender segregation, as the threshold to enter gender atypical industries is lower.

3.3 Segregation in Norway - a literature review

Segregation is a frequently reoccurring issue in the public policy debate in Norway. The general perception is that the Norwegian labor market is highly segregated, but that it is an under-researched topic. Requests for more systematic and comprehensive studies are repeatedly expressed in policy documents and literature reviews (Teigen, 2006; NOU, 2012; Jensberg et al., 2012). In this section, I give a brief overview of the existing literature. More comprehensive reviews of the literature are offered by Teigen (2006); NOU (2012); Reisel and Brekke (2013).

The majority of segregation studies that use Norwegian data focus on more or less narrowly defined occupational groups or sectors of the economy. These groups are generally large occupational groups that are easily defined in time series data, such as occupations in construction, crafts, manufacturing, health care and education. As Teigen (2006) also points out, these are among the most gender segregated occupations, and by focusing on these groups important changes

in other occupations may be overlooked. The impression from existing research is that segregation has remained high in many of these traditionally male and female dominated occupations, but that there are signs of desegregation in some less segregated occupations (Håland and Daugstad, 2003; Teigen, 2006).

A recent report by Jensberg et al. (2012) is more comprehensive than previous studies. The aim of the report is to study the degree of both horizontal and vertical gender segregation in the Norwegian labor market in the period 1990-2010. The analysis of horizontal segregation covers both occupational, industrial and sectorial segregation and the measures used are mainly the distribution of males and females, as well as the female employment share, across categories. The authors conclude that there were no large changes in horizontal segregation over time. When studying occupational segregation, they found a slight decrease in segregation over time, but the pattern was mixed. The female share in some female dominated occupations decreased, but so did the female share in some male dominated occupations. Sectoral segregation increased as the public sector has become more female dominated over time, while there were no changes in industrial segregation. This report is important as it provides a very broad overview of the trends in segregation in Norway, but it uses very aggregated data. In most of the analysis the authors operate with 10-20 categories (occupations or industries). This high level of aggregation could miss important patterns that take place at more detailed levels.

Existing research points in the direction of more equal distribution of males and females in the hierarchy of jobs over time. In a comparison of the gender gap in access to managerial positions in Norway and the US in 1997, Birkelund and Sandnes (2003) found that the gender gap in access had decreased since the beginning of the 1980s. In his study of white collar workers in manufacturing in 1980-1997, Olsen (2004) found that both vertical and horizontal segregation decreased considerably. The female employment share among white collar workers increased, and women entered positions on all hierarchical levels. But still in 1997, there were very few women in top positions. Håland and Daugstad (2003) confirmed that the share of female managers increased from 1982 to 2002, but show that women were most frequently represented in middle management. Jensberg et al. (2012) also found that vertical segregation decreased in 1990-2010, partly as a consequence of fewer male managers.

As mentioned earlier, the Norwegian labor market is generally perceived as highly segregated compared to other European countries. This perception is

largely based on a number of comparative studies that were conducted in the 1990s that showed that Norway, along with the other Nordic countries, had the most segregated labor markets in Europe.[5] More recent studies suggest, however, that Norway is no longer at the extreme. A report funded by the European Commission found that Norway experienced a decline in occupational segregation between 1997 and 2007, and concluded that Norway no longer could be classified as a high-segregation country (Bettio and Verashchagina, 2009). Norway did not, however, experience a decline in sectoral/industrial segregation, and was still classified as a country with a high level of sectoral segregation. When interpreting these findings, it is important to notice that many countries were added in the last years of the study, and that many of the new countries were Eastern European countries with high levels of segregation. Therefore, as also pointed out by Solheim (2012), the fact that Norway no longer stands out in the ranking is in part explained by the addition of new countries with higher levels of segregation. In addition, structural changes such as changes in female labor force participation might have changed the ranking of the countries (Solheim, 2012).[6]

Hallden (2014) also found that Norway did not stand out as a particularly segregated country in a comparison of 22 countries in 2010. However, her study also includes a number of highly segregated Eastern European countries, which shifts Norway towards the middle of the distribution of countries. She found that the ranking of countries depends on the segregation measure used but that Norway remains slightly below the European average and therefore qualifies as a medium segregated country.

Educational choice is a strong predictor of occupational choice in the labor market. Therefore, attention has also been directed towards segregation in educational attainment, and in the policy debate there is a wish for policies that encourage gender atypical educational choices in hope to affect gender segregation in the labor market (NOU, 2012). Reisel and Brekke (2013) provide an overview of the gender composition at different levels of education and a discussion of the processes that affect the educational choices of boys and girls. In secondary education, there are mixed patterns. The gender distribution is close to equal in general education, but many vocational tracks are highly segregated (NOU, 2012). The share of female students who choose vocational tracks has decreased over time (Høst and Evensen, 2009), which has increased the gender segregation in the vocational tracks further (Reisel and Brekke, 2013). The number of students

[5] See for example Charles (1992) and discussion in Hansen (1995).

[6] cf. Appendix A, where the advantages and disadvantages of the IP-index are discussed.

in higher education and the share of female students has increased since the 1970s. The female share has increased in most subjects, but there is still a strong gender bias in engineering, natural sciences and programs related to health care and social work. Overall, there is a stronger tendency towards gender equality in traditionally male dominated study programs than in female dominated programs (NOU, 2012).

Based on the existing literature, the Norwegian labor market does not seem quite as segregated as the public debate sometimes claims. In recent years, there have been signs of more gender equality both in terms of vertical segregation and horizontal occupational segregation. The evidence on industrial segregation is more mixed, but indicates stability over time. Despite some signs of improvements, the level of segregation in the Norwegian labor market remains high.

However, many of the available studies focus on particular segments of the labor market or shorter time periods (with the exception of Jensberg et al. (2012)). In particular, there is limited evidence on how gender segregation was affected in the 1970s and 1980s when the public sector expanded the most and the inflow of female workers was the highest. This study adds to the existing literature by studying changes in segregation over four decades, from 1970 to 2009, especially focusing on the impact of the growing public sector. In addition, this study offers a more formal analysis than some of the existing studies, which focus on very simple measures of segregation (such as female employment shares) and highly aggregated data.

3.4 Data

This study relies on register data from two sources: Census data for years 1970 and 1980, and register data covering years 1986-2009 (2010). I focus on employed individuals aged 18-66. I define individuals whose earnings are above twice the National Insurance Scheme basic amount in a given year as employed, which is identical to what e.g., Havnes and Mogstad (2011a) have done.[7] I also exclude individuals for which employment industry is missing, either because they are not employed or because of misreporting.

Norwegian females work part-time to a larger extent than females in many

[7]The basic amount is used for calculation of most of the benefits in the National Social Insurance Scheme. It is adjusted by the Parliament every year and corresponded to USD 12,500 in December 2014.

other European countries. When the female employment rate started to increase in the 1970s most females worked part-time. Full-time work only started to become more common in the 1980s (Ellingsæter and Gulbrandsen, 2007). In 2014, around 38 percent of all working females still worked part-time according to official statistics (Statistics Norway, 2015b). While this is an important issue, I do not address it here in more detail. I include both part- and full-time workers in the analysis.

The main variable of interest is employment industry. In 1970-2010, the classification of industries changed no less than five times. This poses some challenges for the analysis since a consistent definition of industries is important if one reliably wants to study the time trends in segregation. If the classification of industries changes over time, it can be hard to tell whether a change in segregation is real or whether it merely reflects a regrouping of the data. In order to get around this problem, Bertaux (1991) suggests three possible solutions. The first is to use all current industries, in which case it is possible to compare years for which the industry classification is the same. In this case, this would result in the following subperiods: 1970-1980, 1986-1998, 1999-2001, 2002-2009 and 2010. Second, one can use a sub-sample of industries that are consistently defined over time. In many cases, however, this would lead to a small and non-representative sample of industries. Third, it is possible to aggregate the data or to use crosswalks to obtain a consistent number of industries. How well this works depends on how different the classifications are and how many new industries are added over time. I use a combination of the first and the third strategy, and below I describe the steps taken to end up with the sample of the analysis.

The Standard Industrial Classification from 1978 (Statistics Norway, 1978; Vassenden, 1987) and its 1983 version (Statistics Norway, 1983), which were used in 1970-1980 and 1986-1998, respectively, are very similar, and they can easily be harmonized using crosswalks.[8] However, industry is reported on the 3-digit level in 1970 and on the 5-digit level in the following years. In addition, since the 1970 data was originally collected using an earlier classification, not all 3-digit industry codes were used in 1970. In order to have a comparable number of industries across time, I group these industries with similar existing industries.[9] The results

[8]Both the Standard Industry Classification from 1978 and 1983 are based in the International Standard Industrial Classification of all Economic Activities (ISIC) Rev 2.

[9]One "new" industry worth mentioning separately. In 1970, the industry sector called Welfare Institutions was not observed in the data. It was one of the larger sectors in 1980, employing 2.07 percent of the sample. When comparing 1970 and 1980, it is merged with Medical, dental, other health and veterinary services which was the single largest industry employing 8.47 percent

are not sensitive to these changes. In years 1980 and 1996-1998 a small number of individuals were removed from the sample because industry was reported at a less detailed level than the rest of the data (less than 0.5 % of the sample in each year).

In 1999-2001 industry was reported using the 1994 version of the Standard Industrial Classification (SN1994) (Statistics Norway, 1994), and in 2002, the following Standard Industrial Classification (SN2002) (Statistics Norway, 2002) was released.[10] It was very similar to its predecessor, and I harmonize the industry data in 1999-2009 using crosswalks. A small number of industries with a very small number of employees (no employees in some years) were grouped with other industries to obtain a consistent number of industries over time. This affected 1.1 % of the individuals in the sample and did not impact the results. The total number of consistently defined industries was 570 in this period.

In the last year available, 2010, yet a new version of the Standard Industrial Classification (SN2007) (Statistics Norway, 2007) was introduced.[11] This was very different from the earlier versions and harmonizing it with the previous years would not weigh up for the value of adding one more year of data to the sample.

This leaves me with two subperiods for which I have consistent data on industry: 1970-1998 and 1999-2009. In the main analysis I use the most detailed level available since measuring industry at higher levels of aggregation can tend to wash out some segregation by mixing industries with different gender composition. The drawback of using very detailed data is that male dominated industries are often more narrowly defined than female dominated industries, which can also affect the segregation measures (Anker, 1998; Rubery et al., 1999).

It is important to remember that trends in segregation can only be properly interpreted within periods when industries are consistently defined, as the changes in industry classifications can affect the level of the segregation index as such. As an additional feature of the analysis, I create an industry sector variable that is consistently defined in 1970-2009 and that has 17 categories.

In order to study segregation within educational groups, I divide the population into four groups based on their level of education. In compulsory education, I include individuals with compulsory education (up to 10 years of education), and individuals with up to two years of secondary education, who have not graduated

of the sample in 1980. When comparing 1980 with later years, however, Welfare Institutions occur as an industry of its own.

[10] SN1994 and SN2002 were based on NACE Rev.1 and NACE Rev 1.1, respectively.

[11] SN2007 is based on NACE Rev. 2.

from high school. The second group includes individuals who have graduated from secondary education (13-14 years of education), and from Folk high schools (Folkehøyskole), who make up for a very small share of this group.[12] I divide individuals with tertiary degrees into two group depending on duration. I define degrees that have a duration of 2-4 years of tertiary education as short tertiary degrees and degrees with a duration of more than five years of tertiary education as long tertiary degrees. The latter group is very small compared to the other education groups, but there is a reason for not including it with the short tertiary degrees. The educational composition of the two groups is very different. Among the short tertiary degree holders, there is a high concentration of females in health related degrees, and this is likely to drive the trends in segregation in this group. Among long tertiary degree holders, the distribution of individuals across field of study is less segregated. The individuals with long tertiary degrees are presumably among the most career oriented, and it is therefore interesting to study the trends in segregation in this highly skilled group separately.

Summary statistics of the sample are presented in Table 3.1. It shows the female share in all samples, as well as the sample size and the number of industries.

3.5 Methods

There is a long debate in the literature about the best way to measure segregation, and the only fact that is widely agreed upon is that there is no single best method for measuring segregation.[13] While potentially cumbersome in practice, good practice is to report several measures of segregation. This is useful in checking the robustness of the findings, but different measures also tend to pick up slightly different dimensions of segregation, which can enrich the analysis. Most of this analysis relies on the dissimilarity index (also known to as the ID-index) developed by Duncan and Duncan (1955), but I supplement the analysis with other measures of segregation. In particular, I divide industries into male and female dominated industries following Hakim (1993). In Appendix A I also use the IP-index (Karmel and MacLachlan, 1988) as an alternative to the dissimilarity index. In this section I start by describing the dissimilarity index, before I discuss its advantages and

[12]This definition of compulsory and secondary education is different from that of Statistics Norway, who define individuals with two years of high school as high school graduates. The definition used here is consistent with that in international work, see e.g., Education at a Glance (OECD, 2014) or the Survey of Adult Skills (OECD, 2013).

[13]The literature is reviewed and discussed in detail by Grusky and Charles (1998), Watts (1998), Anker (1998) and Bridges (2003) among others.

disadvantages relative to other measures of segregation. The dissimilarity index is calculated as:

$$S_t = (0.5) \sum_i \mid m_{it} - f_{it} \mid \tag{3.1}$$

where m_{it} (f_{it}) is the percentage of male (female) workers employed in industry i in year t. This index takes a value between 0 and 100 and is interpreted as the percentage of women (or men) who would have to change industries in order to reach an equal distribution of males and females across industries. An index of zero means that the female share in each industry is identical to the overall female share in the labor force (i.e., full integration), while an index of 100 means that men and women work in different industries (i.e., that there is complete segregation).

The dissimilarity index is by far the most commonly used measure of segregation in the literature, which is partly explained by its simplicity. It is easy to calculate and to interpret, but it has some drawbacks. As many segregation indices, it is sensitive to the number of industries in the analysis. The level of the index mechanically increases when the number of industries increases.[14] The relationship between the level of the dissimilarity index and the number of industries is not linear, however. Anker (1998) shows that the increase in the dissimilarity index is much smaller when moving from 2-digit to 3-digit data than when moving from 1-digit to 2-digit occupational data.

Carrington and Troske (1997) show that the dissimilarity index is sensitive to the size of the industry cells and to the minority share (in this case the female share of total employment). When the industry cells are small, random allocation of workers can in itself lead to considerable deviation from evenness. This leads to a segregation index that is higher than the true level of segregation. Similarly, Carrington and Troske also show that the dissimilarity index tends to be too high when the female share of total employment is low. Therefore, one has to be careful when using the dissimilarity index to study time trends in segregation in time periods when female labor force participation increases. To work around these issues, the authors propose calculating an alternative dissimilarity index that measures the distance from randomness rather than evenness and compare this to the standard dissimilarity index. If the two differ substantially, using the

[14] To illustrate: if everybody worked in one industry, the index would be zero, and if everybody worked in their own industry the segregation index would be 100.

standard dissimilarity index is problematic.[15]

The main critique of the dissimilarity index is that changes in the dissimilarity index over time can stem either from changes in the within industry gender composition, or from changes in the relative size of industries. Watts (1998) and Grusky and Charles (1998), among others, have argued that a good measure of segregation should only pick up changes in segregation that stem from the changes in the distribution of males and females within industries.

To get around this critique, many researchers (Blau, Simpson, and Anderson, 1998; Blau, Brummund, and Liu, 2013) have used a decomposition technique introduced by Fuchs (1975) that decomposes the change in the dissimilarity index between two points in time into a sex component and an industry mix component.[16] The sex component refers to the part of the change in the index that happened because the distribution of men and women within industries changed keeping industry composition fixed. The industry mix component is interpreted as the part of the change in the segregation index that is explained by changes in the relative size of industries keeping the distribution of males and females within industries constant. The decomposition is easily computed by first noting that M_{it} (F_{it}) is the number of males (females) in industry i in year t and that $T_{it} = M_{it} + F_{it}$, and by rewriting Equation 3.1 as

$$S_t = (0.5) \sum_i \left| \frac{q_{it} T_{it}}{\sum_i q_{it} T_{it}} - \frac{p_{it} T_{it}}{\sum_i p_{it} T_{it}} \right| \tag{3.2}$$

where $q_{it} = M_{it}/T_{it}$ ($p_{it} = F_{it}/T_{it}$) is the proportion of males (females) in each industry. The sex composition and industry mix components are then defined as:

$$S_{sex} = \left[(0.5) \sum_i \left| \frac{q_{i2} T_{i1}}{\sum_i q_{i2} T_{i1}} - \frac{p_{i2} T_{i1}}{\sum_i p_{i2} T_{i1}} \right| \right] - S_1 \tag{3.3}$$

[15] Female labor force participation increased rapidly in the first half of the period of this study (see Figure 3.2). To study whether this is a problem, I calculated the alternative measure of segregation proposed by Carrington and Troske. The results suggest that there is very little bias in the dissimilarity index. In other words, neither changes in the female participation rate nor in the size or number of industries is a problem in this setting. Nonetheless, I have also calculated the dissimilarity index at different levels of aggregation, and by excluding industries with industries with less than 50 and 100 individuals to make sure that the results are not driven by small industries. The results are robust to all of these changes.

[16] An alternative to decomposing the dissimilarity index is to size-standardize it, which in practice means giving the same weight to all industries. The changes in the standardized index indicates how the level of segregation would have changed if there had been no change in the relative size of the industries (Jacobs, 1989). This method is not very commonly used in the literature since it has the drawback that it gives large weight to small industries (Anker, 1998).

$$S_{ind} = S_2 - \left[(0.5) \sum_i \left| \frac{q_{i2} T_{i1}}{\sum_i q_{i2} T_{i1}} - \frac{p_{i2} T_{i1}}{\sum_i p_{i2} T_{i1}} \right| \right] \tag{3.4}$$

where S_1 and S_2 denote the dissimilarity index as defined in Equation 3.2 in the start and end years, respectively. As pointed out by Blau et al. (1998), this decomposition has two weaknesses. First, the results may depend on the set of weights used, and second, in order for the two effects to the sum to the total effect inconsistent weights must be used. Alternatively, one can use consistent weights and allow for an interaction term.

Bertaux (1991) noted that since the sex and industry mix components (S_{sex} and S_{ind}) are simply the sums of the industry specific components, it is possible to study the influence of the specific industries on the dissimilarity index by disaggregating the sex and industry mix components. I use this approach to study which industries contributed the most to changes in segregation over time.

When interpreting the sex and industry mix components of specific industries, it is helpful to be understand a few things about how the decomposition of the segregation index works in practice.[17] First, in any given year, an industry where the female share of employment is equal to the female employment share in the labor force will have a sex component of zero. The sex component of an industry increases in absolute size as the deviation from the overall female employment share increases. Further, an increase in the sex component can occur for two reasons; either because the female share in a female dominated industry increases, or because the female share in a male dominated industry decreases. It is important to note that in periods where the female share of employment increases, the point of comparison for the dissimilarity index is moving. If the female share in a female dominated industry increases by less than the female share in the labor force, the sex component will be negative as the distance to the comparison point decreases. Second, the contribution of an industry to the dissimilarity index, and to the sex and industry mix components, depends on its size. When calculating the sex component, industry size is held constant, but a given change in the female employment share will be assigned more weight if the initial size of the industry is larger. The same is true for the industry mix components.

Based on the discussion above, it is clear that the dissimilarity index has its flaws, but so have many other indices of segregation. I have chosen to use

[17] This is also explained in Blau et al. (1998).

the index of dissimilarity as the main measure of segregation in this analysis despite the issues mentioned above for four reasons. First, the fact that is it so commonly used makes the results comparable to those in other studies. Second, it is very easy to calculate and interpret. Third, while some researchers argue that is it a weakness that the dissimilarity index depends on changes in the industry structure, I find this useful. Since the 1970s, the industry structure has undergone large changes, as will be clear in subsequent sections, and ignoring this would not give a truthful picture of what happened in this period. Fourth, while the dissimilarity index is not very well suited for studying time trends in segregation in periods where female labor force participation is increasing, my view is that at least it does better than the IP-index, which is sometimes used as alternative to the dissimilarity index.[18] The problem with the IP-index is that the maximum value of the index in a given year is dependent on the female share of employment, and so the index has a tendency to show an increasing trend in segregation in periods of increasing female labor force participation simply because the maximum value of the index is increasing. This is explained in more detail in Appendix A and also shown in Figure 3.A.1.

3.6 Institutional background

Norway has undergone tremendous changes that have affected all spheres of society since the end of World War II. In the aftermath of the war, the political focus was on rebuilding the economy. The initial goal was to build a strong manufacturing sector that would generate jobs and economic security. As the economy started to recover, resources were also targeted at building a welfare system to help individuals that could not provide for themselves. As a means of reaching this goal, it was necessary to build up a large public sector with responsibility for education, health care, social security and care for the elderly.

In this section, I briefly discuss some of the developments that are important for understanding the changes in employment and gender segregation that are analyzed in Sections 3.7 and 3.8.[19]

[18] For example, the IP-index is the preferred index of the European Commission and it is used to monitor gender equality within the European employment strategy (Bettio and Verashchagina, 2009).

[19] A more detailed description of the development of the Norwegian economy and welfare state is provided by Halvorsen and Stjernø (2008).

3.6.1 Industry composition

When World War II ended, the political focus was on rebuilding the Norwegian economy. In this quest, the manufacturing sector (especially heavy industries such as metal and mineral industries) played an important role, partly because of close government involvement (Hodne and Grytten, 2002). Employment in manufacturing peaked in the early 1970s, when it employed about 22 percent of the work force (Statistics Norway, 2015a,d). Since then the manufacturing sector has been restructured. Spurred by the oil findings on the Norwegian continental shelf and technological change, Norwegian manufacturing became more specialized. The demand for low skilled manual workers decreased and the demand for high skilled labor increased. The employment share in manufacturing decreased steadily, and in 2014 the manufacturing sector employed less than 10 percent of all workers (Statistics Norway, 2015b).

In the 1970s, Norway started to transform into a service economy. The employment share in the service sector (public and private) increased from 53 percent in 1970 to 73.5 percent in 2000 (Hodne and Grytten, 2002). Most of the increase came from the public sector, which expanded from 20.6 percent of total employment in 1970 to 33.7 percent in 1993, whereafter it has stabilized at 32 percent (Figure 3.1). But the private service sector also increased. On the one hand, employment in restaurants, entertainment and personal services increased. On the other hand, the demand for professional and technical services increased along with the technological development. In particular, the demand for engineering services increased as the manufacturing sector was restructured and the oil industry developed (Hodne and Grytten, 2002).

The restructuring of the economy had important implications for female workers. First, as physical strength became less important, their relative disadvantage relative to males decreased. Second, many of the new (service) jobs were considered especially suitable for females, and new labor markets opened up. Third, the increased demand for skilled labor favored women who increased their educational investments relative to men.

3.6.2 The public sector

As the Norwegian economy recovered and welfare increased (boosted also by the oil findings on the Norwegian continental shelf in 1969), the Norwegian welfare system developed. Among the first steps towards the welfare state was building

a social security system that covered a wide range of benefits, such as child benefits, sickness insurance, unemployment insurance, disability pensions and old age pensions. But soon, the public sector expanded its responsibilities in providing education, health and child care. In that process, the public sector became the largest employer in Norway. The public sector employment share increased from 10.8 percent in 1950 (Statistics Norway, 2015d) to 20.6 and 31.7 percent in 1970 and 2014, respectively (Figure 3.1). The public sector has a wide range of responsibilities, such as general government and administration, national insurance, health care, child care and education, but Figure 3.1 shows that the increase in public sector employment over time was almost entirely driven by expansions in the health and social work sector, that increased from 7.6 to 21.1 percent of total employment in 1970-2014.[20]

The health care sector expanded in the whole post-war period, but it took off after 1970 when the Hospital Act was implemented. The law shifted the responsibility for planning, building and managing hospitals from the municipalities to the county level, and the range of care provided was extended. Also, the responsibility for the care for the elderly was moved from municipal to county level and was integrated in the health care system and the government made large investments in building nursing homes (NOU, 1997; Daatland, 1997; Otterstad, 2013). These changes had a large impact on the employment structure in Norway, as the demand for nurses and other professionals in health and social work increased massively.

Another development that greatly changed public sector employment was the expansion of child care that started in the mid 1970s. Between 1963 and 1991 the number of children in formal day care increased from 8,500 to 150,000 and the child care coverage among children under age seven increased from 2 to 38 percent (Statistics Norway, 1995). The expansion has continued, and the number of children in day care increased further from 187,600 to 287,200 between 2000 and 2013, when coverage reached 90 percent. Between 2000 and 2014 the number of employees in kindergartens increased from 52,673 to 93,814 (Statistics Norway, 2015c).

[20]Note that this figure refers to both public and private providers of health care services, but that private actors only provide a small share of the total.

3.6.3 Education

In the years following the war, increasing the educational level of the population was an important political goal. The main motivation was to boost economic growth through human capital investments, but the social aspect was also emphasized. Access to education should not be dependent on social of geographical background (Jrgensen, 1997). Since the 1960s, a number of educational reforms have been implemented at all levels in the educational system and they have been accompanied by large increases in educational attainment.

The first important expansion started in the 1960s after the Norwegian Parliament legislated a mandatory schooling reform that increased compulsory schooling from 7 to 9 years. The new system was gradually implemented, and was not fully implemented until 1972 (Black et al., 2008). The change in compulsory schooling lead to changes in both secondary and tertiary education. Secondary education was restructured to improve the balance between general education and vocational education (Jrgensen, 1997).

Before the late 1960s, higher education was only for the chosen few. But at the end of the decade, an expansion of the higher education system started, partly because the large baby boom cohorts demanded more education, and partly because the demand for high skilled workers increased as the private sector became more knowledge intensive and as the public sector expanded (Jrgensen, 1997). A period that has been referred to as the era of the student explosion started in the mid 1980s and the number of students in higher education increased from 101,000 to 169,000 in years 1986-1994 (Try, 2000). Since then the number of students has continued to increase, but at a much slower pace.

The changes in educational attainment over time are discussed in more detail in Section 3.8.1. The increased educational attainment has greatly changed the labor market prospects of the Norwegian work force, and women have improved their formal qualifications relative to men. But a closer look at the distribution across field of study reveals that men and women to a large extent sort into different fields of study, and the pattern is surprisingly stable, especially among men. Thus, given these large and persistent gender differences in specializations, some level of segregation is likely to persist in the labor market for some time to come.

3.6.4 Labor force participation

Figure 3.2 shows the labor force participation rate in the population aged 15-74 by gender in 1972-2014 as reported by Statistics Norway. The female labor force participation rate was only 45 percent in 1972 but it started to increase in the mid 1970s, reaching 63 percent in the late 1980s. From the 1990s onwards, the increase in labor force participation was more modest, and stabilized around 69 percent. The increase in female employment was mainly driven by married women and mothers, and especially by mothers of young children (Havnes and Mogstad, 2011a). The male labor force participation rate was high, between 70 and 80 percent, but slightly decreasing throughout the period. The gender gap in labor force participation almost closed, and was only five percentage points in 2014. The female share of total employment increased from roughly 36 percent in 1972 to 45 and 48 percent in 1990 and 2014, respectively.

The increased labor force participation among females was a consequence of many forces in society. Traditional family values have historically been very important in Norway, but this started to change in the late 1960s. The student revolt of the late 1960s helped change the idea of the ideal family from a male breadwinner and a housewife to a dual-earner family, although females still mostly worked part-time (Leira, 1992; Halvorsen and Stjernø, 2008). The expansion of the public sector also helped as it created a shortage of labor, and married women were the only large labor reserve that could be mobilized.

The changes in family structure also shaped the welfare state. Publicly provided child care was discussed early in the post-war era, but it was not prioritized until the 1970s. The main motivation for introducing formal child care was to support child development, but survey evidence also pointed towards increasing demand for formal child care among mothers (Leira, 1992; Havnes and Mogstad, 2011a). It is often argued that public child care was implemented to stimulate maternal labor supply, but Leira (1992) argues that it was rather the other way around. Before formal child care was available, mothers developed informal care networks in order to work. But as the female employment rate increased, the supply of informal care decreased, and the demand for formal day care increased. This is also supported by Havnes and Mogstad (2011b) who find that the causal effect of child care on maternal employment was very small. Instead, publicly subsidized child care mainly crowded out informal child care.

Another family policy that potentially had an impart on female labor force supply was the introduction of paid maternity leave in 1977. Through this reform,

the maternity leave entitlements were extended from covering 12 weeks of unpaid leave to 4 months of paid leave and 12 months of unpaid leave. The main mechanism through which maternal employment was stimulated was through improved job protection that allowed women to return to their old jobs after one year of absence. However, Carneiro et al. (2015) find no effect on maternal employment or income 2 and 5 years after the child was born.

Even though the empirical studies find no significant effect of either child care or maternity leave on maternal employment, these interventions might still be important for the employment choices of mothers. First, it is possible that both policies affected labor force participation in the long run through changes in social norms or similar. Another possibility is that the policies changed the degree of labor force attachment (i.e., hours worked) or that women entered different types of sectors or occupations due to the reforms.

The expansion of the public sector was probably one of the most important drivers of female labor force participation. As the welfare state developed, the number of service jobs exploded. These were generally jobs in nursing and child care - jobs that were considered suitable for women, and that were very similar to the tasks that they had performed outside the labor market before. The public sector was also an attractive employer for mothers, as working hours were flexible and work was easy to combine with family life. The downside of this development was that the public sector in general, and the health and social work sector in particular, became very female dominated. Still, in 2013, 81 percent of all persons employed in the health and welfare sector were females (Statistics Norway, 2015b).

The increasing investments in higher education also helped to increase female labor force participation, as it is a well-established fact that labor force attachment is positively correlated with educational attainment. But educational attainment also affected female employment more indirectly as discussed by Blackburn et al. (2002). Before the boom in higher education started, most of the women in the labor force were young unmarried women but when these women spent more years in education, part of the labor force went "missing". This allowed older women/mothers to enter the labor market in their place.

3.7 Empirical analysis

In this section, I present the results from the empirical analysis. In Section 3.7.1, I discuss general trends in employment by gender, before I turn to analyzing the

trends in gender segregation as expressed by the dissimilarity index in Section 3.7.2. In order to study the relative importance of within and between industry changes, the changes in the dissimilarity index are decomposed into sex and industry mix components in Section 3.7.3. Motivated partly by the observation that the changes in gender segregation over time are driven by industries that do mainly employ medium and low skilled workers, I study the trends in segregation by level of education in Section 3.8.

3.7.1 General employment trends by gender

As mentioned in Section 3.6, the labor force participation rate of Norwegian females started to increase rapidly in the 1970s. This had an enormous impact on the Norwegian work force. Table 3.1 shows that the work force, as defined in Section 3.4, increased by about 820,000 individuals between 1970 and 2009 and about 77 percent of these "new" workers were female. The female share of the work force increased from 24 percent in 1970, to 34.9 percent in 1980 and 44.5 percent in 1990, where after it slowly increased to 47.1 percent in 2009, as shown in Table 3.1.

The large inflow of female workers affected the industry composition of the Norwegian economy, as shown in Table 3.2. Manufacturing was initially the largest industry sector but its share of total employment declined steadily throughout the period. Agriculture, construction, transportation and wholesale and retail trade also decreased in relative size. Public sector industries, and especially health and social work, increased, as did the employment share in real estate, renting and business activities. Employment shares by gender are plotted in Figure 3.3. It shows that manufacturing, construction and wholesale and retail trade have been the most important industries employing males, and that health and social work has been the most important employer of females. Over time, however, the the male share in manufacturing has decreased, and males have become more evenly distributed over industries. Females, on the other hand, have become more concentrated in health and social work over time, and in 2009, almost 40 percent of all employed women were working in the health and social work sector. The female employment share increased in most industries as shown in Figure 3.4. Interestingly, however, the female employment share did not change in the two most segregated sectors, construction and health and social work, where the female share has remained below ten percent and above 80 percent, respectively.

This analysis already suggests that there have been many changes in the

Norwegian labor market in the past decades that have implications for gender segregation. However, the facts that there is no change in the most segregated industries, that females are becoming more concentrated in public service jobs, and that the health and welfare sector is still increasing indicate that segregation is still an issue in the Norwegian labor market.

3.7.2 Trends in segregation

The overall trend in gender segregation, as expressed by the dissimilarity index, is depicted in Figure 3.5. The dissimilarity index is computed at three different levels of aggregation. I focus on the most detailed level of data (3-digit level in 1970-1980 and 5-digit level in 1986-1998 and 1996-2009) and use the more aggregated data as sensitivity checks.

Given the large changes in the female labor force participation and industry composition discussed in the previous section, the level of industrial segregation has been surprisingly stable. Using industry data reported at the 5-digit level, the dissimilarity index varied around 50 percent, meaning that in order to obtain an equal distribution of males and females within industries, about 50 percent of all workers should change industries. Between 1970 and 1980, segregation decreased when using 3-digit level data. In the 1980s, there was a decrease in the dissimilarity index measured at all levels of segregation, followed by a number of years characterized by stability. From the mid 1990s, there were signs of increasing levels of segregation, although the increase was very small. These results are in line with those of Jensberg et al. (2012), who concluded that the period 1990-2010 was characterized by stability, and that if anything industrial segregation was on the rise.

3.7.3 Decomposition of trends

The advantage of expressing the level of segregation as one single number is that it is easy to interpret, and trends are easy to overview. The downside is that a lot of information is lost. Based on Figure 3.5, one would conclude that there was very little change in the level of segregation between 1970 and 2009. But from Section 3.5, we know that the dissimilarity index is the sum of within industry composition of males and females and the industry mix in the economy. From Section 3.6, we also know that there have been significant changes both in the gender composition within sectors and in the industry composition. In this section,

I therefore decompose the changes in the dissimilarity index to investigate the relationship between these two forces, and I study which industries contributed to the changes in segregation. I split the data into four periods: 1970-1980, 1980-1990, 1990-1998 and 1999-2009.

In addition, I apply a method that was introduced by Hakim (1993), which is helpful in studying whether the changes in segregation over time stem from male or female dominated industries, or both. More specifically, I divide industries into three categories based on the female share of employment in each industry. An industry is defined as male (female) dominated if its female share of employment is more than 10 percentage points lower (higher) the overall female share of employment. In 2009, for example, the female share of total employment was 47.7 percent. Then it follows that industries with a female share below 37.7 percent were defined as male dominated, while industries with a female share above 57.7 percent were female dominated. The rest were classified as integrated industries.

1970-1980

Between 1970 and 1980, the dissimilarity index decreased by 0.48 percentage points. Table 3.3 shows that the decrease was the sum of two opposing forces. On the one hand, the negative sex component suggests that the within industry change in gender composition alone would have led to a 2.07 percentage point decrease in segregation in this period had the industry composition remained at its 1970 level. On the other hand, the industry mix component was positive, meaning that the relative employment share of segregated industries increased, thus increasing segregation by 1.59 percentage points.

In Table 3.4, I investigate the impact of different industries on the dissimilarity index, following Bertaux (1991). I calculate the sex and industry mix components separately for all industries and aggregate the industry specific components to the 1-digit level to facilitate the display of the results. Most sex components were negative, indicating that males and females became more equally distributed within industries, either because an originally male dominated industry experienced faster female employment growth than in the overall economy, or because a female dominated industry experienced slower growth than average growth in female employment. There were, however, two exceptions: the agriculture, hunting and forestry sector became more male dominated, while the wholesale and retail trade sector became even more female dominated, increasing segregation by 1.41 and 0.80 percentage points, respectively.

The industry mix component of the health and welfare sector was by far the largest. Alone, it suggests that the employment growth in the health and welfare sector (from 4.7 to 10.9 percent of total employment) lead to a 5.74 percentage point increase in segregation in this period. Manufacturing, wholesale and retail trade and transport, storage and communication industries had the largest negative industry components because they were all fairly segregated industries that decreased in relative size.

The industry specific sex components are plotted against the industry mix components in Figure 3.6, which allows me to identify the industries that contributed the most to the changes in the dissimilarity index. The figure confirms the findings in Table 3.4, but also further highlights that the gender distribution became more equal in most industries, as only three industries (agriculture, hunting, fishing and forestry, retail of food beverages and tobacco and food manufacturing) had significantly positive sex components, i.e., became more segregated.

In Table 3.5, I study whether the changes in segregation came from changes in male or female employment, or both. Following Hakim (1993) and Blau et al. (1998), all industries were divided into male dominated, integrated and female dominated industries based on their gender composition as discussed above. By holding the category of each industry fixed over time (within subperiod) in Panel A, it is possible to track the flows of workers between the three categories. In other words, by tracking the flow of male and female workers between male dominated, integrated and female dominated industries between 1970 and 1980 we can learn about the underlying changes in the sex component. Similarly, the change in the distribution of total employment is informative of the industry mix component. In Panel B, the industries are re-categorized based on their current year gender composition and this shows how the reallocation of workers affected the categorization of industries. The distribution of industries is tabulated in Panel C.

Panel A suggests that the negative sex component in Table 3.5 is driven by male employment. Males moved from male dominated industries, such as agriculture, fishing and manufacturing to integrated and female dominated industries. Women, on the other hand, became more concentrated in female dominated industries, which increased segregation. Panel B suggests that some initially integrated industries became male dominated as a consequence of the inflow of male workers and outflow of female workers.

Table 3.5 also indicates that the positive industry mix component in Table 3.3 was driven mostly by employment growth in female dominated industries. In Panel B, there are some signs of polarization, meaning that the employment shares in female and male dominated industries increased at the expense of integrated industries.

1980-1990

Between 1980 and 1990, the dissimilarity index decreased by 3.4 percentage points. Table 3.3 shows that this was the result of men and women becoming more equally distributed within industries, and of relative employment growth in integrated industries. In particular, Table 3.4 indicates that reductions in employment in male dominated industries like agriculture and manufacturing, and female dominated wholesale and retail trade had a lowering impact on segregation, and that employment growth in the health and welfare sector did not have as strong an impact on segregation in the 1980s as in the 1970s. The within industry changes in gender composition were very similar to those in the 1970s, with two exceptions: the female employment share in the agriculture sector increased (while total employment decreased), leading to less segregation, while public administration became more female dominated, leading to an increase in segregation.

Figure 3.7 shows that local government administration had a particularly large impact on segregation, as it became more female dominated and increased in relative size. Expansions of child care and municipal social service offices, which were highly female dominated, and national defence, which was male dominated, also put upward pressure on segregation. At the other extreme, downsizing in male dominated industries such as agriculture, fishing, shipbuilding, ocean transport and construction helped to decrease segregation. Ocean transport became less male dominated partly because downsizing mostly affected male workers, and general somatic hospitals became slightly less female dominated, which helped to reduce segregation.

Table 3.5 shows that similarly to the 1970s, male employment increased in female dominated and integrated industries at the expense of male dominated industries. Among female workers, employment increased in integrated industries, and decreased in both male and female dominated industries. Total employment increased in integrated and female dominated industries, but a comparison with the numbers of Panel B suggests that the gender composition of some initially integrated industries changed to being either male or female dominated in this

period. Therefore, the effects of male and female worker flows on segregation are somewhat unclear.

1990-1998

The dissimilarity index increased by 0.87 percentage points between 1990 and 1998. Table 3.3 shows that the increase was driven by changes in the industry mix (+1.03), as the sex component was very small (-0.16).

The small sex component was partly explained by the fact that the industry specific sex components were smaller in absolute size than in previous periods (notice the difference in the scale on the axes in Figures 3.6 to 3.9). In other words, there was a general tendency towards smaller within industry changes in gender composition. The fact that the components were smaller could be related to the slower increase in the female labor force participation rate. Still, some industries stand out. First, the sex component of the health and welfare sector was only -0.32, which was largely explained by a slower defeminization of public service sectors, such as child care activities and social welfare services for the aged (see also Figure 3.8). Fast female employment growth in (primary) education also increased segregation. Local government administration and municipal social service offices decreased in relative size in the 1990s, and the female share of employment decreased in these industries, resulting in negative sex and industry mix components. The further expansion of child care and social welfare services for the elderly, which mostly employed females, continued in the 1990s, putting upward pressure on segregation. The only private sector industry that stand out in Figure 3.8 is telecommunications. The employment share in telecommunications decreased while the female employment share decreased, resulting in a rather large positive sex component.

In the 1990s the flows of male and female workers between integrated, male and female industries were smaller than in previous decades, but the directions of the worker flows were essentially the same as in earlier periods. Both male and female employment decreased in male dominated industries and increased in female dominated industries, suggesting that male workers moved in a way that reduced segregation while the opposite was true for female workers. The total employment numbers suggest that the positive industry component was driven by employment in female dominated industries.

1999-2009

Between 1999 and 2009, the dissimilarity index increased by 1.10 percentage points. The decomposition in Table 3.3 shows that the sex component was negative and the industry mix component was positive, but the relative size of the two components changed compared to earlier periods. Within industry changes contributed less to reducing segregation (-0.61), while employment growth in segregated industries contributed more to the change in segregation (+1.70).

In Table 3.4 and Figure 3.9, where the industry specific components are displayed, we see that most sex and industry mix components were small in absolute size compared to earlier periods. Many industry mix components in Figure 3.9 were positive, but it was mainly employment growth in child care activities and nursing and caring for the aged that drove the increase in segregation. Feminization of primary and secondary education also continued in the 2000s. This period was characterized by good economic conditions in Norway, and as a consequence the construction industry expanded while its female employment share decreased, resulting in a rather large and positive impact on gender segregation.

A few additional points in Figure 3.9 are worth mentioning. The expansion of child care was the main force driving the increase in the dissimilarity index in this period, although the female employment share in the sector decreased somewhat. While not labelled in the figure, the female employment share in telecommunications (hidden behind primary and lower secondary schools) decreased sharply in this period, which contributed to an increase in the dissimilarity index. At the other end of the sex component scale, we find national post activities, which went from female dominated to gender balanced as the employment share decreased. Provision of personnel went from being female dominated to slightly male dominated.

Table 3.5 shows a pattern that is very similar to previous periods, although in 1999-2009 both male dominated and integrated industries lost workers to female dominated industries. The changes were, however, quite small. Comparing Panels A and B, it looks as if some of the initially female dominated industries became integrated and some of the initially integrated industries became male dominated as a consequence of the worker flows.

Summary

The analysis above suggests that Figure 3.5 hid a lot of action. Throughout the period, males and females became more evenly distributed within industries, although this slowed down after 1990, when the female labor force participation rate stabilized. Further, the analysis suggests that changes in the industry composition played an important role in the development of gender segregation over time. The expansion of the health and welfare sector was very important in driving segregation, as it absorbed almost 50 percent of all female labor market entrants since 1970, and employed almost 40 percent of all female workers in 2009. The impact of the health and welfare sector on segregation was at its largest in the 1970s, but it also affected the level of segregation considerably after 1990. The expansion of child care services and care for the aged and disabled were the main the drivers of between sector segregation over time.

The story of segregation in the Norwegian labor market is primarily a story about female workers, but men have also played an important role. Especially in the 1980s, downsizing of male dominated industries, such as agriculture and manufacturing, counteracted the upward pressure that the expansion of female dominated service industries put on segregation. In later years, male employment increased in business activities, and to some extent also to mining and quarrying and construction, which lead to more segregation as these industries are male dominated.

3.8 Segregation by education

In this section, I study segregation within educational groups. Educational attainment increased considerably after 1970, especially among women, and this has had a large impact on the composition of the work force. As female workers have increased their human capital investment more than male workers, there may be interesting changes in employment patterns over time, both within and between educational groups.

The results in the previous section indicated that different skill groups may have contributed differently to changes in segregation over time. For instance, it was shown that the changes in segregation were driven by a relatively small number of industries. Among these industries, there was child care, care for the elderly, agriculture, manufacturing industries, and to some extent education, public administration, business activities and construction. These industries have

different demand for skills, and many mostly hire low and medium skilled workers. Therefore, it is possible that different educational groups experienced different trends in segregation.

There is also a close connection between educational and occupational attainment. In the absence of good data on occupations, the combination of information on educational attainment and industry can serve as a proxy for occupation. For example, the majority of employees at general somatic hospitals with a long tertiary degree are physicians while holders of short tertiary degrees are nurses. Therefore, it is possible to think of the results in this section as suggestive evidence on occupational segregation.

In this part of the analysis, I divide the sample into four groups based on educational attainment as discussed in Section 3.4. These groups are compulsory education, secondary education, short tertiary degrees (2-4 years of higher education), and long tertiary degrees (five or more years of higher education).

3.8.1 Educational attainment

Before turning to the analysis of segregation within educational groups, I review the changes in educational attainment. Table 3.6 shows the sample by level of education. The share of workers with only compulsory education decreased from 75.5 to 28.1 percent between 1970 and 2009, and the decrease was about equally large for males and females. The share of workers holding a secondary education degree increased from 12.6 to 34.5 percent, and the share was higher among males. The share of the work force with a short tertiary degree, increased from 9 percent to 28 percent between 1970 and 2009 and the share was higher among females. In 2009, 34.6 and 28.1 percent of the female and male workers, respectively, had completed a short tertiary degree. The explanation for the high share of female workers with a short tertiary degree already in 1970, is that nursing requires a degree from nursing college. The share of the work force that had a long tertiary degree was small relative to the other educational categories, but it increased rapidly over the last decades. In 1970, only 3.1 percent of the work force had a long tertiary degree (3.7 percent of the males and 1.1 percent of the females) but increased to 9.5 percent (10.8 and 8.0 percent of the males and females, respectively).

The female share increased at all levels of education, but the increase was the fastest among long tertiary degree holders, and it reached 40 percent in 2009. Throughout the period the highest share of female workers was in the group of

short tertiary degree holders, where the female share increased from 38.5 to 59.2 percent.

Figure 3.10 shows the distribution of males and females with upper secondary and tertiary education across field of study. Two things are worth noting. First, there is a surprisingly high stability in the choices of fields of study over the period, especially among men.[21] Second, there was a high level of segregation in the education system. Males were overrepresented in natural sciences, vocational and technical subjects at all levels of education, while women were much more likely to choose education in health, welfare and sports, teaching and business and administration. Thus, given these large and persistent differences in specializations of male and female workers, some level of segregation is likely to persist in the labor market for some time to come.

3.8.2 Within education trends in segregation

The trends in segregation for the four educational groups are shown in Figure 3.11.[22] There were large differences in both the levels and trends in the dissimilarity indices of the different educational groups. In general, there was a negative correlation between level of segregation and level of education. Tertiary degree holders experienced reductions in segregation until the late 1990s, and workers with secondary education experienced a rapid increase in segregation especially after 1990.

In the remainder of this section, I decompose the education group specific changes in segregation and investigate which industries were driving the observed patterns. I plot the industry specific sex components against the industry mix components, which allows me to identify the industries that contributed the most to changing the level of segregation. Note that the axis scales of the plots differ since the components varied considerably in size depending on time period and educational group.

[21] Over time the share of both males and females in general programs decreased, which coincided with the decrease in secondary education.

[22] In the 1970s, only a small share of the population had higher education, especially among females. To make sure that small cell sizes are not driving the results, I calculated the dissimilarity index that uses the distance from randomness rather than from evenness as proposed by Carrington and Troske (1997), and found that this is not a problem in the analysis.

3.8.3 Compulsory education

Workers with compulsory education experienced small changes in the dissimilarity index compared to the other educational groups. In the 1970s, the level of segregation increased slightly, but it decreased again between 1980 and 1992. Thereafter, the level of segregation started to slowly increase again. Table 3.7 reveals that the sex component was negative in all periods, but decreasing in absolute size over time, and that the industry mix component was negative in all periods except in the 1990s, and increasing in absolute size over time. In other words, both within and between industry changes contributed to the changes in the dissimilarity index, but the relative importance on the sex (industry mix) component decreased (increased) over time.

In 1970-1980 and 1980-1990, the decompositions of the dissimilarity index for the compulsory education group look very similar to the overall decomposition, which is explained by the fact that this group covered 54-75 percent of the work force in this period. The largest difference between the compulsory education group and the other educational groups (and the work force as a whole) in 1970-1980 was that the level of segregation increased in the compulsory education group while it decreased in the others. This was largely explained by a larger sex component (in absolute size) on agriculture, hunting, fishing and forestry and a smaller sex component (in absolute size) on health and welfare services.[23]

The compulsory education group experienced a smaller increase in segregation in the 1990s than did the overall economy, and this was mainly driven by a more favorable development in within industry gender composition. Compared to the overall economy (Figure 3.8), workers with compulsory education were not affected by the employment growth in primary education to the same extent as workers in other educational groups, and this is part of the explanation for the smaller change in segregation in this period.

Between 1999 and 2009, the dissimilarity index increased more in the compulsory education group than in the overall economy. This was the result of a smaller sex component and a larger industry mix component (in absolute size). A comparison of Figures 3.9 and 3.B.4 reveals that the smaller sex component was in part explained by a decrease in the female employment share in construction, and the larger in industry mix component was in part explained by fast relative employment growth in child care activities.

[23]The agriculture, hunting, fishing and forestry sector mainly employed workers with compulsory education and therefore mainly affected the dissimilarity index through this group of workers.

Throughout the period, the expansion of the health and welfare sector put upward pressure on segregation as it mostly employed female workers. This was counteracted by increasing female employment in a number of manufacturing industries, as well as in other parts of the economy, in the 1970s, and in the 1980s by downsizing in several large and male dominated industries such as agriculture, fishing, shipbuilding and construction. From the 1990s onwards, there was less restructuring in male dominated industries, both in terms of gender composition and relative size. Therefore, there growth in female dominated service industries became more dominant in changing in the dissimilarity index.

Secondary education

Workers with secondary education experienced very different changes in segregation than workers in other educational groups. The level of segregation initially decreased slightly, but increased rapidly after 1990. Table 3.7 shows that the dissimilarity index increased by 5.2 and 4.4 percentage points in 1990-1990 and 1999-2009, respectively. Within industry gender composition improved in the 1970s and 1980s, but got more uneven especially in the 1990s. The increase in the dissimilarity index was, however, mainly driven by between industry changes. The size of the industry mix component increased in all periods, reaching 4 percentage points in 2009.

Figure 3.B.5 shows that the negative sex component in the 1970s was driven mainly by education, but also business activities, manufacturing industries and local government administration which all became less female dominated. In the full sample (covering all educational groups), the female employment share in health and welfare increased faster than in the overall female employment share, contributing to lower segregation, but in the secondary education group, the female employment share in health and welfare increased at about the same rate as the female share in the compulsory group on average. Therefore, the sex component was smaller and did not contribute to lowering segregation in the compulsory education group.

In the 1980s, there was almost no change in the dissimilarity index because the negative sex component and positive industry mix component cancelled each other out. Many of the industries stood out in Figure 3.7 reemerge in Figure 3.B.6, but the coefficients are of slightly different size. In particular, the downsizing of ocean transport had a larger negative (lowering) impact on the dissimilarity index in the secondary education group as total employment decreased and the female

employment share increased. Employment growth and female employment growth were larger in retailing of food, beverages and tobacco and local administration, which had a positive impact on the dissimilarity index. The child care sector did not have a particularly large impact on the level of segregation in this period.

Downsizing in male dominated industries such as agriculture, fishing and construction that was observed to affect the dissimilarity index of the overall economy and in particular in the compulsory education group, did not affect the level of segregation as much in the secondary education group, and had part in explaining why the dissimilarity index did not decrease more in this period.

The increase in segregation in the 1990s was the sum of a positive sex component (+1.56) and industry mix component (+3.66). There was very little variation in the sex component. National defence stood out as it became less male dominated, and operation of banks decreased in relative size. The large and positive sex and industry components were driven by positive but fairly small components in a large number of industries, rather than by large components in a small number of industries as in earlier periods. The industries with the largest industry mix components were all in health and social work, construction and transportation, all of which were highly segregated (with a minority share below 20 percent).

The increase in segregation continued almost as strong in the 2000s, but now it was mainly driven by relative employment growth in segregated industries. The large industry mix component was driven particularly by day and night nursing, day nursing and caring for the aged and disabled and child care activities. But all the industries with the largest industry mix components (larger than +0.2), are industries in either construction or health and social work.

Taken together, relative employment growth in health and welfare was important for the evolution of segregation in this group. Interestingly, the health and welfare sector did contribute less to the change in the dissimilarity index in the 1970s and 1980s than in other educational groups, and more in the 2000s, when relative employment increased, especially in nursing. The financial sector stood out in the analysis of workers with secondary education. In 1970-1990, total employment and female employment increased rapidly, contributing to an increase in segregation. After 1990, the relative employment in the financial sector decreased, contributing to a reduction in segregation.

Short tertiary education (2-4 years)

Short tertiary degree holders experienced a large decrease in the level of segregation between 1970 and 2009. Between 1970 and 1980, the dissimilarity index decreased by 4.5 percentage points, and in 1980-1998, the index decreased by another 10.7 percentage points, and the decrease was particularly large in the 1980s. Between 1999 and 2009, the decrease was much more modest, only 1 percentage point.

Table 3.7 suggests that the reduction in segregation was largely driven by within industry changes, meaning that males and females became more evenly distributed within industries. The absolute size of the sex components did, however, decrease over time. The industry mix components were relatively small, and positive only in 1970-1980 and 1999-2009. The employment patterns of the short tertiary education group were very different from those in the overall economy. In general, the public sector was more important for the changes in segregation in this group.

In the 1970s, the health and welfare sector had a large negative sex component and a slightly larger positive industry mix component (resulting in a small net effect of this sector). The employment share of the health and welfare sector increased from 17 to 24 percent, while the female employment share actually decreased, from 94 to 87 percent. This sector was the second largest employer of short tertiary degree holders and especially females, only the education sector employed a larger share of this group. Between 1970 and 1980, the employment share of the education sector decreased from 36.3 to 30.6 percent of the group, resulting in a negative industry mix component.

Only one industry, education, had a significantly positive sex component, which indicates that the gender balance became more equal in most industries. In education, the female employment share increased more rapidly than the overall female employment share in the group, thus leading to more segregation. Health and welfare, central and and local government administration and real estate and business services were among the industries that contributed the most to the reduction in segregation.

The integration of males and females continued in most industries in the 1980s. Again, basically all sex components were negative except that of combined primary and junior high schools and upper secondary schools, which became more female dominated. The industry that contributed the most to the reduction in segregation in terms of the sex component was general somatic hospitals,

where the female employment share decreased from 92 to 91 percent and the total employment share decreased from 11 to 9 percent. Primary schools and general somatic nursing homes also contributed to the negative sex component. The downsizing of male dominated industries that was an important driver of segregation in the overall economy had little impact on this educational group, mainly because the employment share of these industries was low and also because downsizing was less likely to affect white collar workers than blue collar workers. Local government administration also increased in relative size in this group, but its impact on the changes in the dissimilarity index was relatively small compared to the overall economy.

The dissimilarity index continued to decrease in the 1990s but at a slower pace. The sex components of three industries are worth noting. Primary education became even more female dominated, leading to a positive sex component, while general somatic hospitals and local government administration became less female dominated, thus contributing to the negative sex component. Apart from these industries, within industry changes in gender composition were small. Increased relative employment in social work related industries lead to the positive industry mix components, while relative employment in education, general somatic hospitals and local government administration decreased.

In the 2000s the decrease in segregation subsided. In part, this was explained by smaller within industry changes in gender composition, and in part, it was the result of increases in relative employment growth in segregated industries, mainly in social work. While the scales in Figures 3.9 and Figure 3.B.12 are different, the decomposition of the change in the 2000s look very similar to that in the 1990s. Software consultancy and supply was among the few private sector industries that stood out, and contributed to increasing segregation as it became more male dominated and increased in relative size.

What characterized the short tertiary degree holders was the high concentration in public sector industries. In all four decades, all the industries that had a large impact on segregation were in the public sector. Interestingly, however, different parts of the public sector had different impacts on segregation. In general, the within industry sex composition became more equal, with the exception of primary (and lower secondary) education. General somatic hospitals helped reduce segregation as both the female employment share and the total employment share decreased. Employment in child care and welfare services such as nursing for the elderly and disabled increased throughout the period, and the latter especially

after 1990, but the relative importance of the employment growth in these industries was smaller than in the compulsory and secondary education groups. Public administration was also important for short tertiary degree holders, but the impact varied over time. Public administration was initially male dominated, and in the 1970s, both central and local government administration helped reduce segregation as they became less male dominated (while total employment decreased slightly). In the 1980s especially local government administration increased segregation as both its female and total employment share increased rapidly. In the 1990s and 2000s both relative total employment and female employment in local government administration decreased, and thus it helped reduce segregation.

Long tertiary education (more than 5 years)

Figure 3.11 shows that the level of segregation was considerably lower among the long tertiary degree holders than in any other group in all of the period 1970-2009. The dissimilarity index decreased by 6.2 percentage points and by 10 percentage points in 1970-1980 and 1980-1998, respectively, and was stable in 1999-2009. Table 3.7 indicates that most of the decrease was due to within industry changes in gender composition. The sex component decreased in absolute size after the 1980s and turned positive in 1999-2009. The industry mix components were mainly small relative to the sex components, except in the 2000s. Figures 3.B.13 to 3.B.16 show that the industries that were driving the changes in segregation were different from the other groups, but that public sector industries were important also among the most highly educated.

Figure 3.B.13 shows that the negative sex component was to a large extent driven by three industries in the 1970s although the majority of the sex components were negative. Central government administration and education both became less male dominated, which resulted in large negative sex components. The third industry that stands out is "other retailing", which became less female dominated in this period, while it decreased in relative size.[24] Real estate and

[24]The industry category "other retailing" contains retailing of books and stationary, retailing of drugs and pharmaceuticals, retailing of cosmetics, retailing of flowers and plants, retailing of household fuel and retailing of commodities not elsewhere classified. In 1980, two out of three workers in this category were employed in retailing of drugs and pharmaceuticals, and it is reasonable that a similar concentration of the workers in other retailing were employed in this industry also in 1970. If so, one can assume that the one of the main reasons that other retailing stands out in the analysis is because of changes in retailing of drugs and pharmaceuticals, which is an industry that also stands out in the decompositions in later periods.

business services, had the largest positive sex component because the female share increased at a slower pace than the average female share in the group. Both total employment and the female employment share in the health and welfare sector increased. Note that in among long tertiary degree holders, the health and welfare sector was initially male dominated, because the majority of physicians were males. Central government administration and education were both large industries in terms of total employment, but decreases in relative size resulted in negative industry mix components.

In the 1980s, the sex component was larger in absolute size, and most sex components were negative. Upper secondary schools had the largest negative sex component, since the female employment share increased. Universities, central government administration related to industrial activities, primary schools and retailing of drugs and pharmaceuticals also became more integrated. Of the industries with the largest positive sex components, there were three branches of central government administration (two of the hidden under engineering, architectural and technical services in the figure), where the female employment share increased faster than in the group as a whole. The female employment share in engineering, architectural and technical services only increased from 6 to 9 percent, resulting in a positive sex component. Relative employment growth in national defence and crude petroleum and natural gas production, both male dominated, contributed to the positive industry mix component as these were both highly male dominated industries.

In the 1990s, there was very little variation in the industry mix components, suggesting that changes in the industry composition did not affect segregation much. Most sex components were still negative, but smaller in absolute size. In this period, the female employment share in engineering, architectural and technical services increased from 9 to 19 percent, which contributed to lower segregation. Upper secondary school became less female dominated, while research and scientific institutes and crude petroleum and natural gas production became less male dominated. In universities, primary schools and municipal social service offices, the female employment share increased, contributing to more segregation in this group.

The decomposition of the change in the dissimilarity index between 1999 and 2009 looks quite different from the one between 1990-1998. In this period, a larger share of the sex components were positive, meaning that men and women became more segregated. The female employment share increased particularly

fast in general somatic hospitals (from 34 to 48 percent), general higher education (from 34 to 45 percent) and in veterinary activities (31 to 67 percent). In software consultancy and supply and other technical consultancy services the female employment share remained around 15 percent. Most industry mix components were small, and the industries that contributed the most to reducing the dissimilarity index were (male dominated) national defense and (female dominated) administration of health care, education etc..

Private sector industries did play a larger role in the decomposition of the changes in the dissimilarity index among workers with long tertiary degrees, but still, there was a relatively high concentration workers with long tertiary degrees in public sector industries. The level of segregation as expressed by the dissimilarity index was considerably lower in this group than in any other group, and in addition the level of segregation decreased massively especially in the 1970s and 1980s. After 1999, however, the concentration of female workers increased in a number of industries, such as education, general somatic hospitals, veterinary services and retailing of drugs and pharmaceuticals. This development was in part responsible for the increase in the dissimilarity index in the 2000s. Similarly, highly educated males were concentrated in industries related to engineering.

Summary

There were large differences in both the level and trend of segregation between educational groups. Among workers with tertiary degrees, the level of segregation initially decreased rapidly, but the decrease subsided in the 2000s. Workers with secondary degrees, on the other hand, experienced a rapid increase in segregation after 1990 and among workers with compulsory education, the level of segregation was fairly stable throughout the period.

The gender composition within industries became more equal over time, but there was a tendency towards smaller sex components over time. In the secondary education and long tertiary education groups, the within industry segregation even increased towards the end of the period. The differential trends between the education groups were partly explained by differences in the industry mix components. In particular, there was a tendency towards larger industry mix components over time among workers with compulsory and secondary education, suggesting that relative employment growth in segregated industries, such as health and social work and construction, became more important in explaining the increase in the dissimilarity index.

The analysis showed that the impact of the public sector on the observed changes in the dissimilarity index was large, but that its impact differed between educational groups and over time. In particular, relative employment growth in welfare related industries such as child care and care for the aged also put large upward pressure on segregation among workers with foremost secondary education but also compulsory and short tertiary education throughout the period. Relative employment in education decreased in all educational groups, but increased female employment in education among workers with short tertiary degrees contributed to the slowdown in the decrease in segregation in this group. Employment changes in public sector administration had a large impact on the development in segregation in all educational groups, but the direction of the change varied between groups and over time. The gender composition in general somatic hospitals improved dramatically among holders of both short and long tertiary degrees.

3.9 Conclusion

The Norwegian labor market is considered to be one of the most gender equal in the world. The female labor force participation rate is among the highest in international comparisons, and public institutions facilitate labor force participation of women, and especially mothers, by providing universal child care and offering flexible arrangements for workers with children. At the same time, the Norwegian labor market is one of the more gender segregated in the world, meaning that i) Norwegian men and women work in different industries and occupations to a larger extent and ii) women are less likely to hold managerial and other influential positions than in many other countries.

The parallel existence of these two phenomena has been referred to as the welfare state paradox, since it is a pattern that is common in developed welfare states (Mandel and Semyonov, 2006; Ellingsæter, 2013). Mandel and Semyonov (2006) argue that welfare states are successful in increasing female labor force participation, but that they create sheltered labor markets for women that enable them to combine family responsibilities and working life, but prevent them from competing with men for powerful positions. Family-friendly policies may discourage employers from hiring women in important positions while at the same time influencing women's employment preferences towards family-friendly occupations with convenient working conditions.

One implication of this hypothesis is that one of the main differences between

high segregation welfare states and low segregation countries with less developed welfare systems is that care work is internalized in the formal labor market in the first type of countries, while these services are provided outside the labor market in the latter type.

Today, Norway undoubtedly falls into the first category of countries, but this has not always been the case. The expansion of the public service sector and the labor market entry of females only took off in the 1970s. Before that, the social service sector was small, and labor market participation among women was relatively low. In other words, Norway was quite similar to many low segregation countries of today. This motivates the question of how segregation has changed since the 1970s. If the argument of Mandel and Semyonov holds also in a longitudinal context, one would expect an increase in segregation in Norway over time.

In the first part of the analysis, where I study segregation in the labor market as a whole, I find that the level of industrial segregation, as expressed both by the dissimilarity index (Duncan and Duncan, 1955) and the IP-index (Karmel and MacLachlan, 1988), remained fairly stable over time, but that the increase in the relative size of the health and welfare sector did indeed contribute to a higher level of segregation over time.

When segregation is studied within educational groups, however, two distinct patterns emerge. On the one hand, workers with higher education experienced a sizable reduction in the level of segregation, especially in the earlier part of the period. Part of the explanation for this lies in the fact that the concentration of women in female dominated industries (mainly education) decreased over time. Instead, female employment increased in less segregated industries, which helped to reduce segregation in this group.

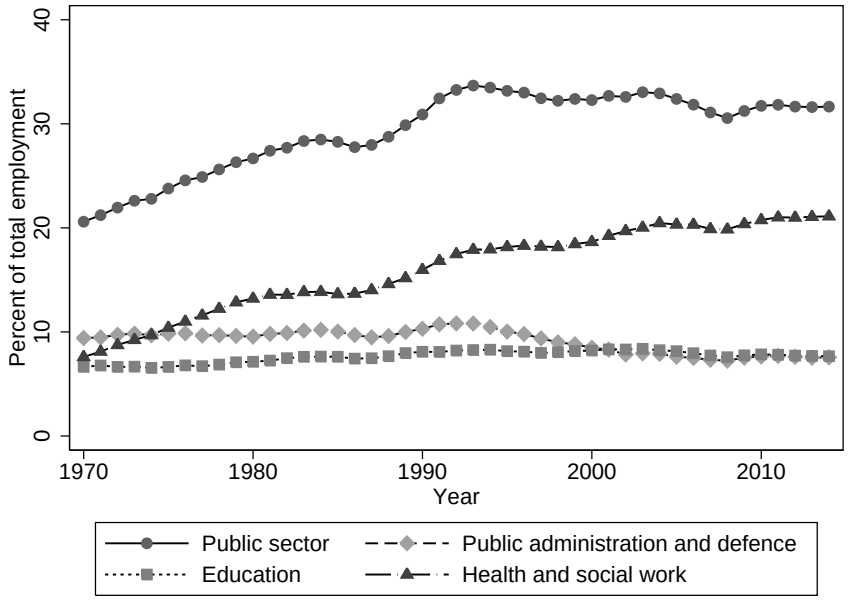
On the other hand, the level of segregation increased massively among workers with secondary education, especially after 1990. Closer analysis reveals that this was mainly driven by rising relative employment in female dominated service industries, but to a smaller extent also by more within industry segregation. This is largely in line with the welfare state paradox lined out above. The late start of the increase in segregation is explained partly by the fact that in the the 1970s and 1980s it was counteracted by increasing female employment in basically all sectors of the labor market and by downsizing in many male dominated industries, which helped to reduce the level of segregation.

Taken together, these findings can at least in part be reconciled with the

hypothesis that the expansion of the welfare state spurs increasing gender segregation as women are channeled into the female dominated service sector. This is what happened especially among workers with secondary education. But gender segregation was reduced among workers with higher education, indicating that other forces were also at play. The findings in this study imply that the boom in educational attainment was very important for the employment of Norwegian women. In the last four decades women overtook men in educational investments. This opened up for more competition and more integration in the labor market. But despite these important improvements among highly educated workers, the level of segregation remains fairly high as there is a strong preference for gender typical specializations among students in higher education.

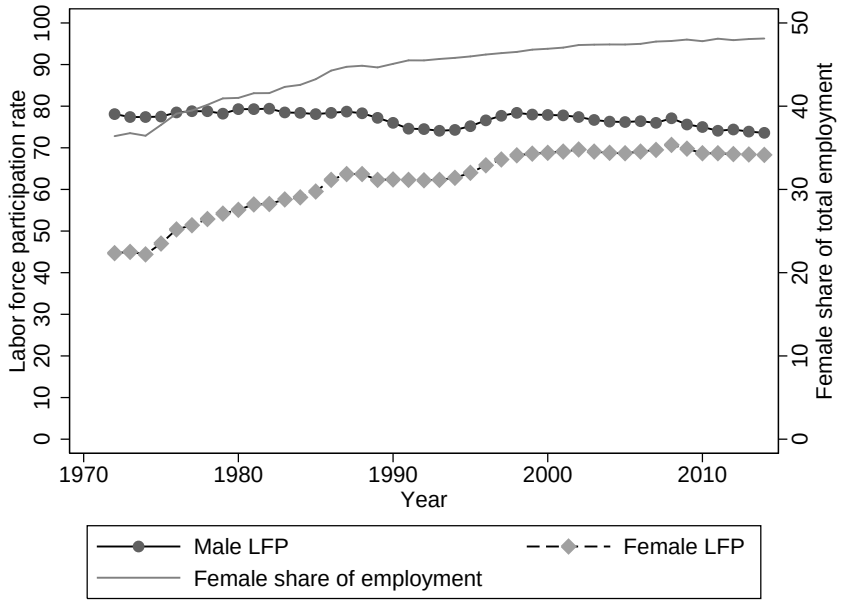
While this discussion has focused on the importance of developments and policies that are more or less directly connected to women, it is important to note that the observed pattern in segregation was also affected by other developments in society, which were unrelated to the gender issue. For example, restructuring of the manufacturing and agriculture sector and business cycles were important for the development of gender segregation in Norway. This serves as an important reminder that segregation is an extremely complex concept. The pattern in segregation over time is the sum of many processes in society, and is therefore hard to explain within narrow theoretical frameworks.

Figure 3.1: Employment in public sector industries 1970-2014



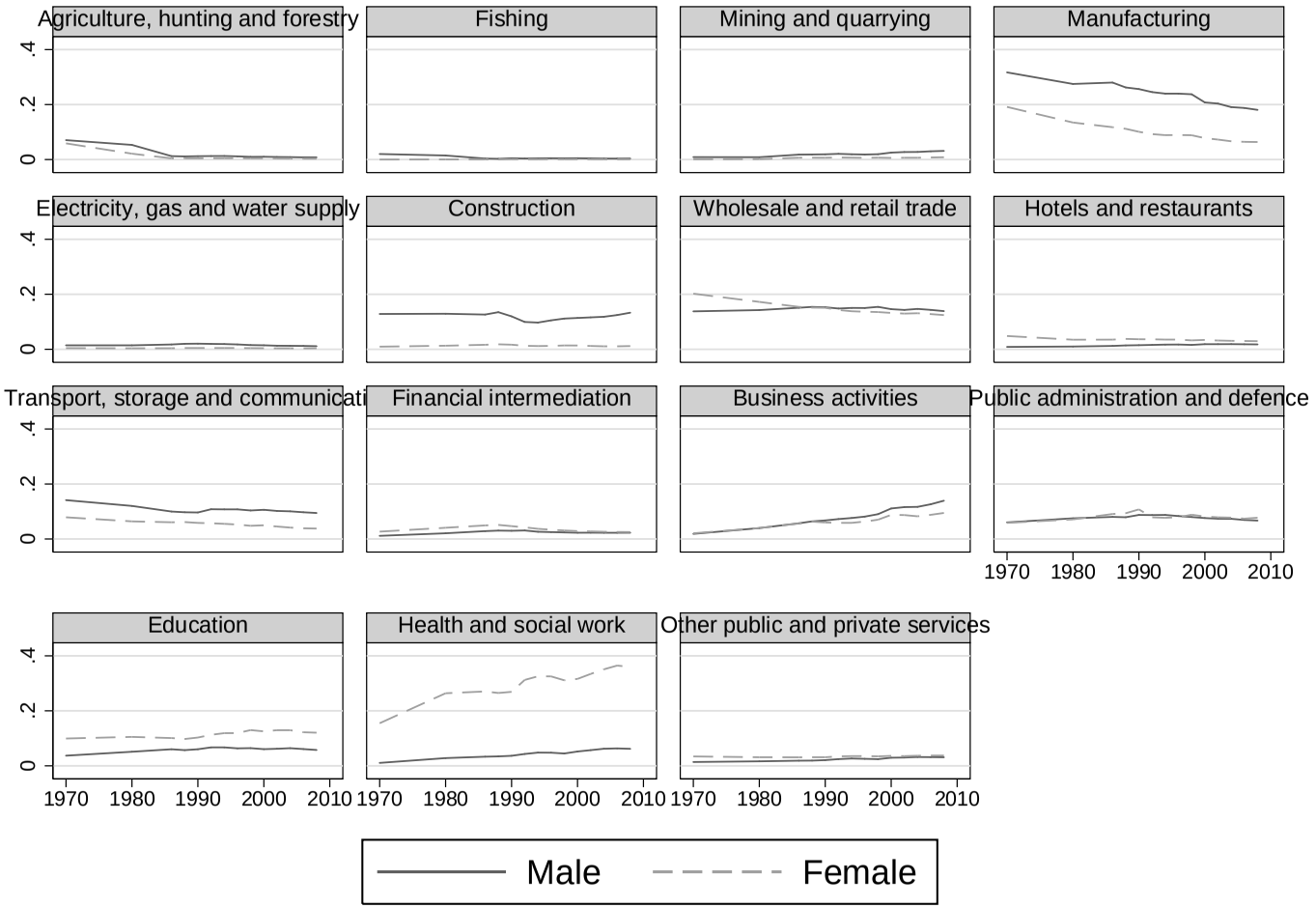
Note: The figure plots the percentage of total employment that is employed in the three largest branches of the public sector, public administration and defence, education and health and social work, as well as the sum of these three industries, which is called Public sector. Note that the figure also includes private sector workers in these industries, but they are only a small share of the total. Source: Statistics Norway.

Figure 3.2: Labor force participation 1972-2014



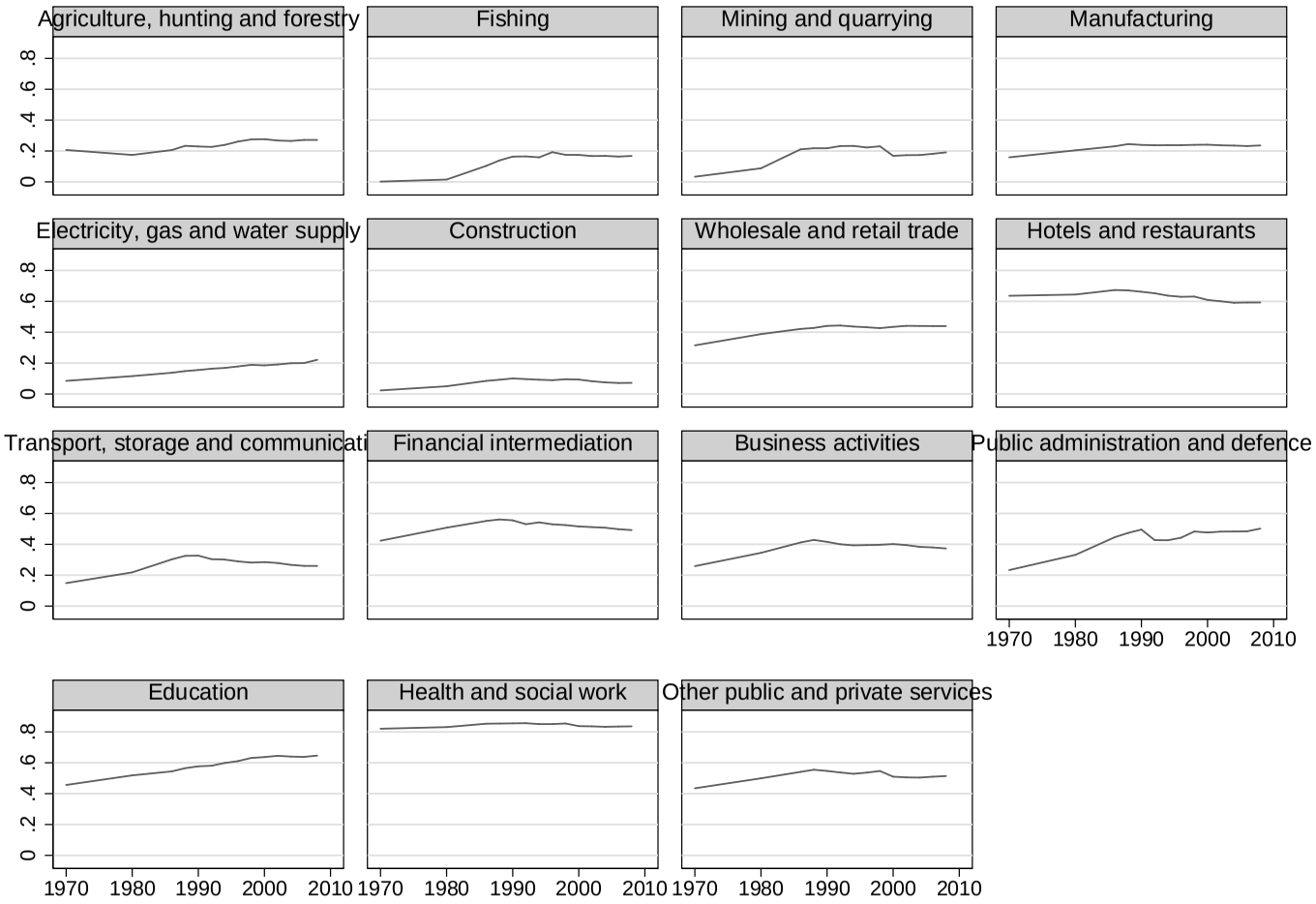
Note: The figure plots the male and female labor force participation rate in Norway in years 1970-2014 for the population aged 15-74, as well as the female share of total employment. Source: Statistics Norway.

Figure 3.3: Male and female employment by industry sector 1970-2009



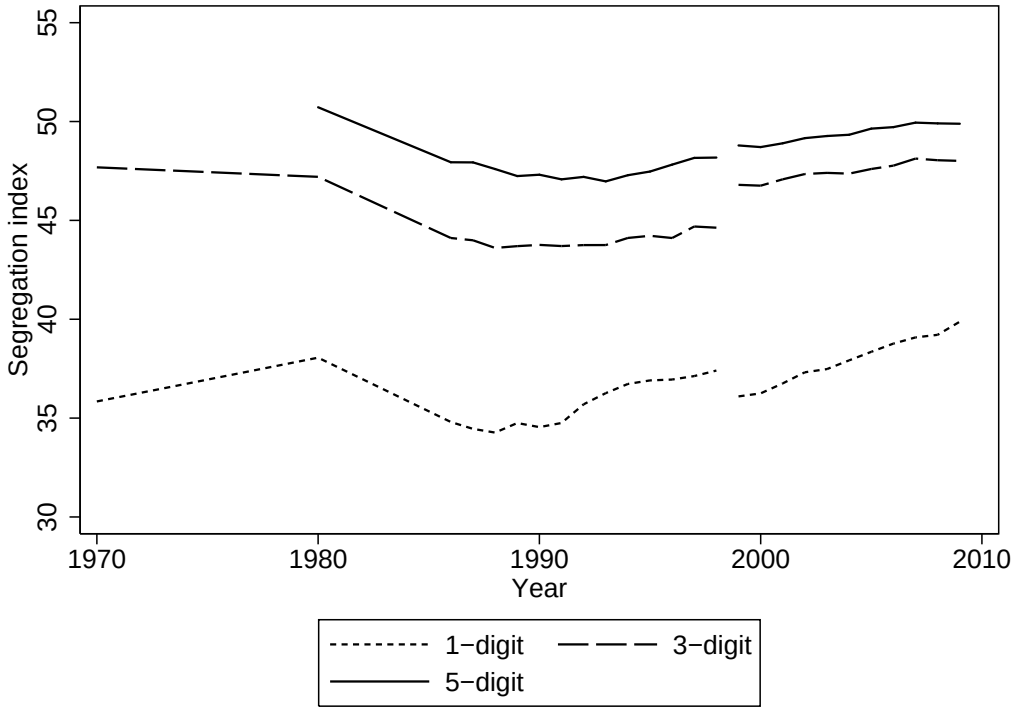
Note: The figure plots the distribution of male and female workers across industry sectors. The sample is restricted to employed individuals aged 18-66. Source: register data as described in Section 3.4.

Figure 3.4: Female share of employment by industry sector 1970-2009



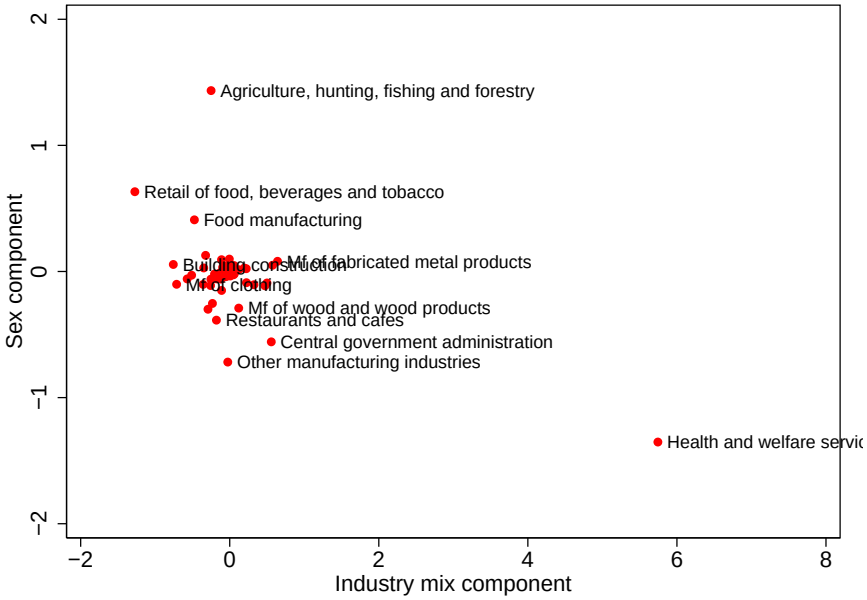
Note: The figure plots the female share of employment by industry sector. The sample is restricted to employed individuals aged 18-66. Source: register data as described in Section 3.4.

Figure 3.5: Trends in industrial segregation as expressed by the dissimilarity index



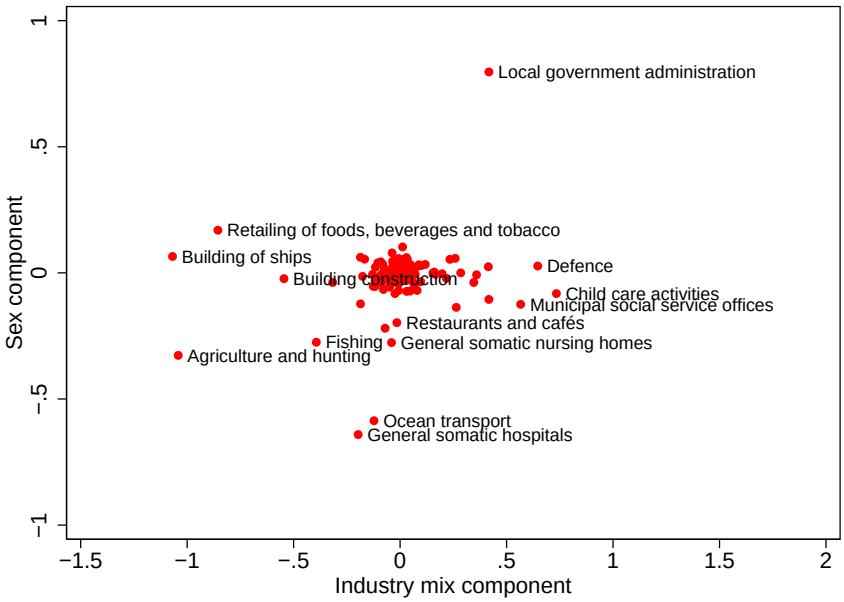
Note: The figure plots the dissimilarity index calculated at the one-, three- and five-digit level.

Figure 3.6: Decomposition of dissimilarity index 1970-1980



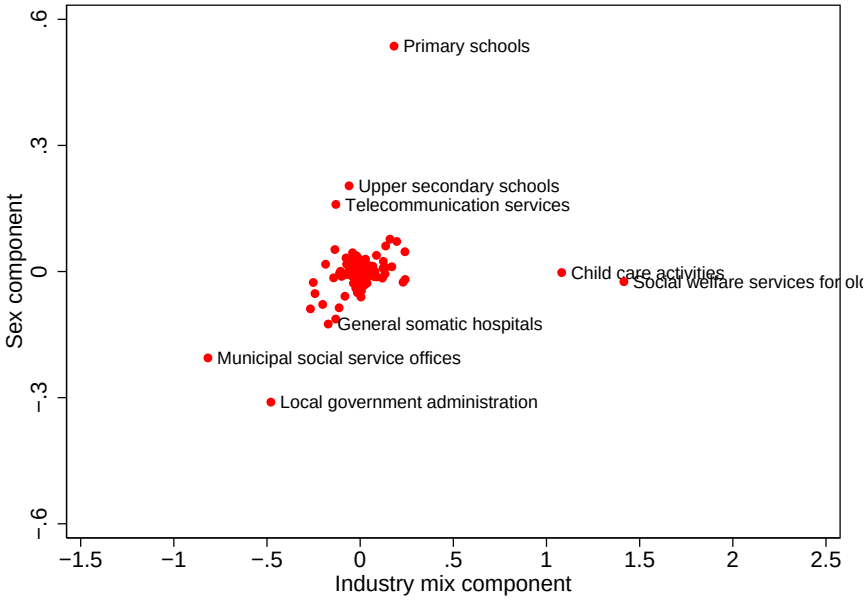
Note: The figure plots the sex and industry mix components that are the result of decomposing the change in the dissimilarity index between two years. The method is described in Section 3.5. The sex component is informative of whether the gender composition of an industry became more or less segregated relative to the overall female employment share in the economy. A positive sex component means that an industry became more gender segregated. The industry mix component is informative of whether an industry contributed to a reduction or increase in the dissimilarity index through changes in its relative size. A positive industry mix component means that in industry contributed to an increase in the level of segregation either because of an increase in the relative size of a segregated industry or decrease in the relative size of an integrated industry.

Figure 3.7: Decomposition of dissimilarity index 1980-1990



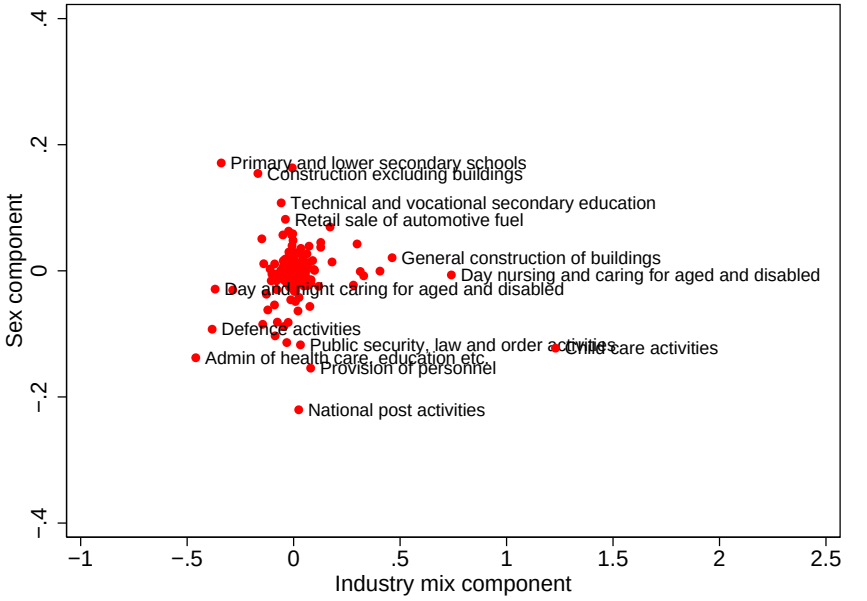
Note: See notes for Figure 3.6.

Figure 3.8: Decomposition of dissimilarity index 1990-1998



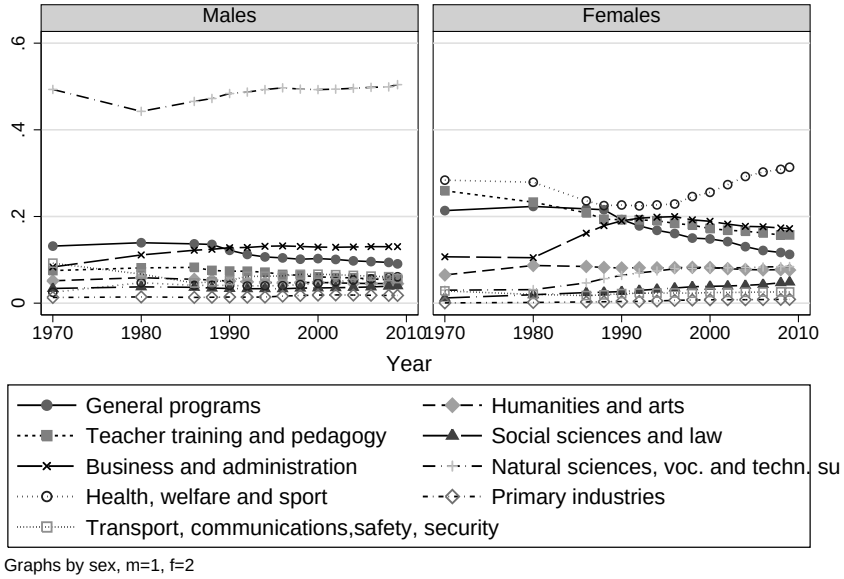
Note: See notes for Figure 3.6.

Figure 3.9: Decomposition of dissimilarity index 1999-2009



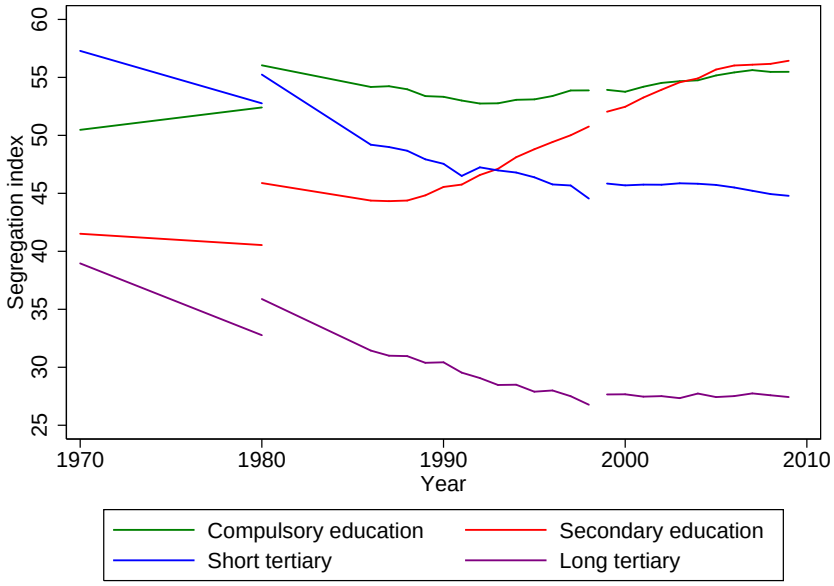
Note: See notes for Figure 3.6.

Figure 3.10: Workers with secondary and tertiary education by field of study and gender



Note: The figure plots the distribution of male and female workers with secondary and tertiary degrees across field of study.

Figure 3.11: Dissimilarity by education groups



Note: The figure plots industrial segregation expressed by the dissimilarity index by educational groups. 3-digit data is used for 1970-1980 and 5-digit data for 1980-1998 and 1999-2009. Education groups defined as follows: compulsory education = compulsory education or two years of secondary education (up to 12 years of education), secondary education = high school graduate (13-14 years of education), short tertiary education = 2-4 years of higher education, and long tertiary education = more than four years of higher education.

Table 3.1: Sample characteristics

	1970	1980	1990	1998	1999	2009
Full sample						
(%) female	24,0	35,0	44,5	45,9	45,7	47,7
N(individuals)	1,080,711	1,285,058	1,445,213	1,601,328	1,702,038	1,902,979
N(men)	821,607	836,167	802,742	866,843	925,040	1,006,641
N(women)	259,104	448,891	642,471	734,485	776,998	896,338
N(3-digit ind.)	67	68	68	68	212	211
N(5-digit ind.)	-	370	370	370	584	587
Compulsory education						
(%) female	24,4	36,8	48,8	50,9	50,8	48,9
N(individuals)	810,271	877,556	778,503	662,627	674,832	511,571
N(men)	612,261	554,993	398,633	325,634	332,123	261,472
N(women)	198,010	322,563	379,870	336,993	342,709	250,099
N(3-digit ind.)	67	68	68	68	211	211
N(5-digit ind.)	-	370	370	370	581	582
Secondary education						
(%) female	15,1	24,3	34,1	35,1	35,5	39,4
N(individuals)	135,550	182,529	314,529	452,105	503,960	626,706
N(men)	115,015	138,096	207,427	293,235	324,902	379,746
N(women)	20,535	44,433	107,102	158,870	179,058	246,960
N(3-digit ind.)	67	68	68	68	211	208
N(5-digit ind.)	-	370	370	369	584	585
Short tertiary education						
(%) female (%)	38,5	44,5	51,3	55,1	54,7	59,2
N(individuals)	94,188	161,880	264,008	369,023	395,604	507,002
N(men)	57,957	89,923	128,516	165,736	179,168	206,931
N(women)	36,231	71,957	135,492	203,287	216,436	300,071
N(3-digit ind.)	67	68	67	67	209	209
N(5-digit ind.)	-	370	365	366	575	580
Long tertiary education						
(%) female	8,5	13,3	20,9	29,4	29,8	40,3
N(individuals)	33,417	52,135	77,020	106,402	116,271	172,734
N(men)	30,575	45,214	60,960	75,107	81,622	103,103
N(women)	2,842	6,911	16,060	31,295	34,649	69,631
N(3-digit ind.)	66	66	67	67	201	196
N(5-digit ind.)	-	342	355	357	524	515

Note: The table shows the sample size, number of males and females, female employment share and the number of observed industries in each of the five samples used in the analysis: the full sample that covers the whole work force, as well as the four education samples.

Table 3.2: Employment by industry sector 1970-2009

	1970	1980	1990	1998	1999	2009
Agriculture, hunting and forestry	7,0	4,3	0,9	0,7	0,8	0,6
Fishing	1,6	1,0	0,3	0,3	0,3	0,3
Mining and quarrying	0,7	0,6	1,4	1,4	1,7	2,2
Manufacturing	28,4	22,7	18,7	16,9	15,4	12,1
Electricity, gas and water supply	1,2	1,1	1,4	1,0	1,0	0,8
Construction	10,4	8,9	7,4	6,7	6,6	7,6
Wholesale and retail trade	14,6	15,2	15,2	14,6	14,3	13,2
Hotels and restaurants	1,9	1,9	2,5	2,4	2,6	2,3
Transport, storage and communication	12,2	10,1	8,0	7,8	8,3	6,7
Financial intermediation	1,6	2,8	3,7	2,7	2,6	2,4
Real estate, renting and business activities	2,0	4,0	6,4	8,1	9,4	11,6
Public administration and defence	6,2	7,4	9,6	8,4	8,3	7,1
Education	5,4	7,0	7,9	9,5	8,7	8,9
Health and social work	4,7	10,9	14,0	16,7	16,9	20,8
Other community, social and personal services	1,9	2,2	2,5	2,9	3,2	3,6
Private households with employed persons	0,3	0,0	0,1	0,0	0,0	0,0
Extra-territorial organisations and bodies	0,0	0,0	0,0	0,0	0,0	0,0
Total	100,0	100,0	100,0	100,0	100,0	100,0

Table 3.3: Decomposition of changes in industrial segregation

	Start year	End year	Change	Sex component	Industry mix component	No. of industries
A. 3-digit						
1970-1980	47.68	47.20	-0.48	-2.07	1.59	67
1980-1990	47.20	43.76	-3.44	-2.91	-0.53	68
1990-1998	43.76	44.63	0.87	-0.19	1.07	68
1999-2009	46.80	48.01	1.21	-0.50	1.71	211
B. 5-digit						
1970-1980	-	-	-	-	-	-
1980-1990	50.72	47.31	-3.41	-2.29	-1.12	370
1990-1998	47.31	48.18	0.87	-0.16	1.03	370
1999-2009	48.79	49.89	1.10	-0.61	1.70	584

Notes: The table shows the decomposition of the changes in the dissimilarity index over time that were plotted in Figure 3.5.

Table 3.4: Sex and industry mix components by industry sector

A. Sex component	1970-80	1980-90	1990-98	1999-2009
Total	-2,07	-2,29	-0,16	-0,61
Agriculture, hunting and forestry	1,41	-0,35	-0,07	-0,02
Fishing	0,03	-0,27	-0,01	0,01
Mining and quarrying	-0,05	-0,03	-0,01	-0,08
Manufacturing	-0,93	0,36	-0,27	-0,22
Electricity, gas and water supply	0,00	0,00	-0,11	-0,09
Construction	-0,06	-0,16	-0,04	0,22
Wholesale and retail trade	0,80	0,34	0,39	0,27
Hotels and restaurants	-0,44	-0,23	-0,14	-0,14
Transport, storage and communication	-0,45	-0,79	0,17	-0,16
Financial intermediation	0,03	0,01	-0,13	0,00
Real estate, renting and business act	-0,04	0,15	-0,05	-0,14
Public administration and defence	-0,59	0,75	-0,39	-0,45
Education	-0,25	-0,24	0,78	0,44
Health and social work	-1,35	-1,82	-0,32	-0,17
Other public and private service act	-0,07	-0,01	0,03	-0,04
Private households w/employed persons	-0,10	-0,01	-0,01	-0,01
Extra-territorial organisations	0,00	0,00	0,00	0,00
B. Industry component	1970-80	1980-90	1990-98	1999-2009
Total	1,59	-1,12	1,03	1,70
Agriculture, hunting and forestry	-0,45	-1,15	-0,05	-0,06
Fishing	-0,35	-0,39	0,00	-0,02
Mining and quarrying	0,00	0,27	0,01	0,36
Manufacturing	-0,86	-2,18	-0,44	-1,03
Electricity, gas and water supply	0,02	0,15	-0,12	-0,06
Construction	-0,28	-0,03	-0,22	1,03
Wholesale and retail trade	-0,68	-1,11	-0,38	-0,58
Hotels and restaurants	-0,35	-0,04	-0,13	-0,15
Transport, storage and communication	-1,03	0,41	0,43	-0,19
Financial intermediation	0,20	-0,03	-0,33	-0,07
Real estate, renting and business act	0,02	0,66	0,50	0,89
Public administration and defence	0,50	0,92	-0,62	-0,61
Education	-0,23	-0,16	0,33	-0,51
Health and social work	5,74	1,62	2,00	2,31
Other public and private service act	-0,26	-0,07	0,10	0,16
Private households w/employed persons	-0,36	0,01	-0,05	0,00
Extra-territorial organisations	0,00	-0,01	0,00	0,00

Notes: The table shows the decomposition of the changes in the dissimilarity index over time that were plotted in Figure 3.5. The industry specific sex and industry mix components are calculated and then aggregated up to the 1-digit level, as described in Section 3.5.

Table 3.5: Distribution of workers and industries by sex composition in industry

	Male	Integrated	Female	Total
A. Distribution of workers				
(year used to define gender composition in parenthesis)				
i. 1970-1980				
1970 (1970)				
Men	51,9	32,4	15,7	100
Women	8,9	28,7	62,4	100
Total	41,6	31,5	26,9	100
1980 (1970)				
Men	43,6	36,8	19,5	100
Women	8,2	25,7	66,1	100
Total	31,4	33,0	35,6	100
ii. 1980-1990				
1980 (1980)				
Men	64,7	19,6	15,6	100
Women	16,1	19,9	64,0	100
Total	47,7	19,7	32,5	100
1990 (1980)				
Men	57,5	24,4	18,1	100
Women	14,4	24,2	61,4	100
Total	38,3	24,3	37,3	100
iii. 1990-1998				
1990 (1990)				
Men	62,6	18,8	18,66	100
Women	17,0	18,6	64,4	100
Total	42,3	18,7	38,9	100
1998 (1990)				
Men	61,2	20,2	18,7	100
Women	15,7	19,0	65,3	100
Total	40,3	19,6	40,1	100
iv. 1999-2009				
1999 (1999)				
Men	63,3	21,2	15,5	100
Women	16,3	22,1	61,5	100
Total	41,9	21,6	36,5	100
2009 (1999)				
Men	62,9	19,9	17,2	100
Women	15,5	21,2	63,3	100
Total	40,6	20,5	38,9	100

Table 3.5: (continued)

	Male	Integrated	Female	Total
B. Distribution of workers				
(current year used to define gender composition)				
1980				
Men	58,8	23,5	17,7	100
Women	14,4	21,7	63,9	100
Total	43,5	22,9	33,6	100
1998				
Men	63,6	19,5	16,9	100
Women	16,8	20,4	62,7	100
Total	42,1	19,9	37,9	100
1999				
Men	63,3	21,2	15,5	100
Women	16,3	22,1	61,6	100
Total	41,8	21,6	36,5	100
2009				
Men	63,8	19,8	16,4	100
Women	15,6	20,3	64,1	100
Total	41,2	20,0	38,8	100
C. Distribution of industries				
1970	34,3	34,3	31,3	100
1980	40,3	28,4	31,3	100
1980	44,5	25,9	29,7	100
1990	50,0	23,8	26,2	100
1998	53,2	21,9	24,9	100
1999	54,9	24,0	21,1	100
2009	56,4	23,7	19,9	100

Notes: Year used to define by gender composition in parenthesis. An industry is defined as male (female) dominated if its female share of employment is more than 0.1 lower (higher) the overall female share of employment. The female share of employment was 24, 35, 44.5, 45.7 and 47.7 % in 1970, 1980, 1990, 1999 and 2009, respectively.

Table 3.6: Educational attainment by gender

	1970	1980	1990	1998	1999	2009
A. All						
Compulsory education	75,5	68,9	54,3	41,7	39,9	28,1
Secondary education	12,6	14,3	21,9	28,4	29,8	34,5
Short tertiary education	8,8	12,7	18,4	23,2	23,4	27,9
Long tertiary education	3,1	4,1	5,4	6,7	6,9	9,5
Total	100,0	100,0	100,0	100,0	100,0	100,0
B. Men						
Compulsory education	75,0	67,0	50,1	37,9	36,2	27,5
Secondary education	14,1	16,7	26,1	34,1	35,4	39,9
Short tertiary education	7,1	10,9	16,2	19,3	19,5	21,8
Long tertiary education	3,7	5,5	7,7	8,7	8,9	10,8
Total	100,0	100,0	100,0	100,0	100,0	100,0
C. Women						
Compulsory education	76,9	72,3	59,5	46,1	44,3	28,9
Secondary education	8,0	10,0	16,8	21,7	23,2	28,5
Short tertiary education	14,1	16,1	21,2	27,8	28,0	34,6
Long tertiary education	1,1	1,6	2,5	4,3	4,5	8,0
Total	100,0	100,0	100,0	100,0	100,0	100,0

Notes: The table shows how workers are distributed across different educational levels.

Table 3.7: Decomposition by education groups

	Start year	End year	Change	Sex component	Industry mix component
A. Compulsory education					
1970-1980	50.48	52.41	1.93	-0.13	2.06
1980-1990	56.04	53.33	-2.72	-2.46	-0.26
1990-1998	53.33	53.88	0.55	-1.07	1.62
1999-2009	53.93	55.48	1.56	-0.40	1.95
B. Secondary education					
1970-1980	41.52	40.54	-0.98	-2.98	1.92
1980-1990	46.71	45.54	-0.17	-2.25	2.07
1990-1998	45.54	50.76	5.23	1.56	3.66
1999-2009	52.04	56.44	4.39	0.36	4.07
C. Short tertiary education					
1970-1980	57.29	52.76	-4.52	-7.08	2.56
1980-1990	55.23	47.54	-7.69	-6.89	-0.80
1990-1998	47.54	44.53	-3.01	-2.85	-0.16
1999-2009	45.84	44.79	-1.06	-2.35	1.29
D. Long tertiary education					
1970-1980	38.96	32.77	-6.19	-4.70	-1.49
1980-1990	35.85	30.43	-5.42	-6.78	1.36
1990-1998	30.43	26.77	-3.65	-3.65	0.04
1999-2009	27.66	27.43	-0.22	0.60	-0.82

Notes: The table shows the decomposition of the changes in the dissimilarity index over time withing education groups, also in Figure 3.11.

Appendix

3.A The IP index

The IP-index that was introduced by Karmel and MacLachlan (1988) is very similar to the dissimilarity index, and is defined as:

$$IP_t = \frac{1}{N_t} \sum_i \left| \left(1 - \frac{M_t}{N_t}\right) M_{it} - \frac{M_t}{N_t} F_{it} \right| \tag{3.5}$$

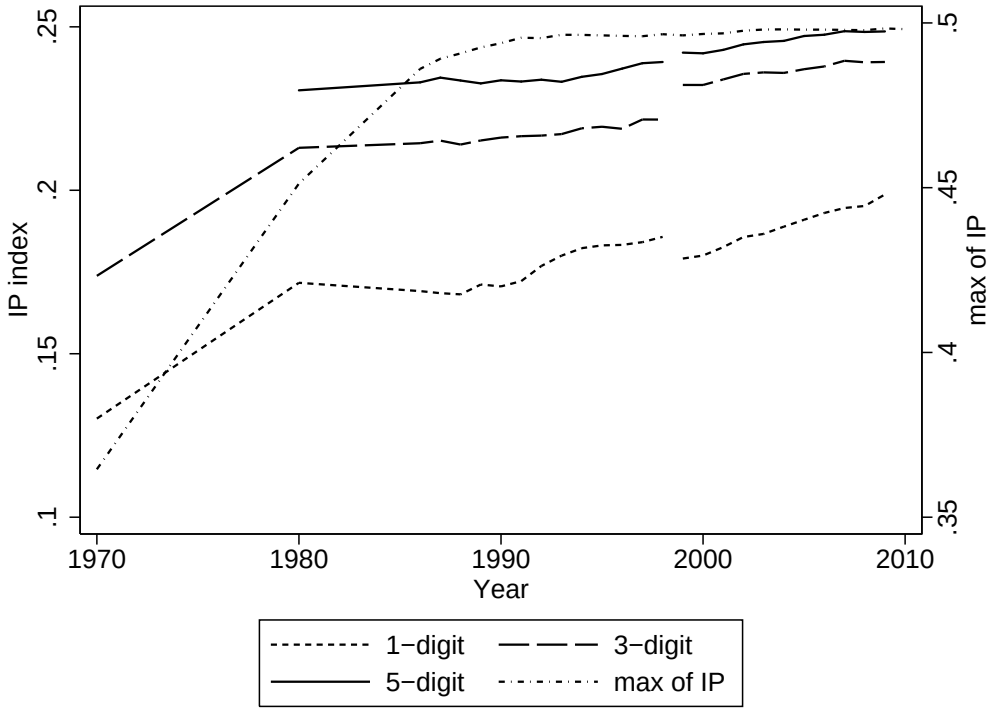
where N_t, M_t and F_t are total, male and female employment in year t respectively. The subscript i denotes the ith industry. The IP-index has the same interpretation as the dissimilarity index but in ranges from 0 in the case of full integration to $(2 * M_t/N_t * F_t/N_t)$ in the case of total segregation. Thus, the maximum value of the index varies with the female share of employment and it reaches its absolute maximum of 50 percent when females constitute half of the total employment. An advantage of the IP-index over the dissimilarity index is that it takes the female share of employment into account directly. But this turns into a disadvantage when studying time trends in segregation, because the level of the IP-index may change simply because of a change in the female participation rate. Thereby, the trends in the index become more difficult to interpret. In periods when the female share of employment has been stable on the other hand, the interpretation is straightforward.

As discussed in Section 3.5, the dissimilarity index has some weaknesses. Therefore, it is useful to use an alternative index to check the robustness of the results. The trends segregation as expressed by the IP-index, along with the maximum of the IP index, is depicted in Figure 3.A.1. The IP-index increased rapidly from 1970 to 1980, but when the increase in the maximum of the index is adjusted for, segregation actually decreased slightly.[25] From 1986 onwards, the trends in segregation were basically identical to those in Figure 3.5. If anything,

[25]If the value of the index (left axis) is divided by the maximum of the index (right axis) the trend is fairly constant over time.

the IP index might indicate a slightly steeper increase in segregation than the dissimilarity index. All in all, the similarity of the trends in segregation as expressed by both the dissimilarity index and the IP-index suggest that the trends in segregation are pretty robust to different measures of segregation.

Figure 3.A.1: Trends in Industrial Segregation using the IP-index



Note: The figure plots industrial segregation using the IP-index (Karmel and MacLachlan, 1988). The index is calculated using 1-, 3-, and 5-digit data on industry. The maximum value of the IP-index is a function of the female employment share in the economy, and the index may therefore increase in periods when female employment is increasing, solely because of this. The maximum value of the IP-index is plotted on the right x-axis. At first sight, the level of segregation seems to be increasing, but when the fact that the range of the index is increasing as well is taken into account, the trend is fairly stable over time.

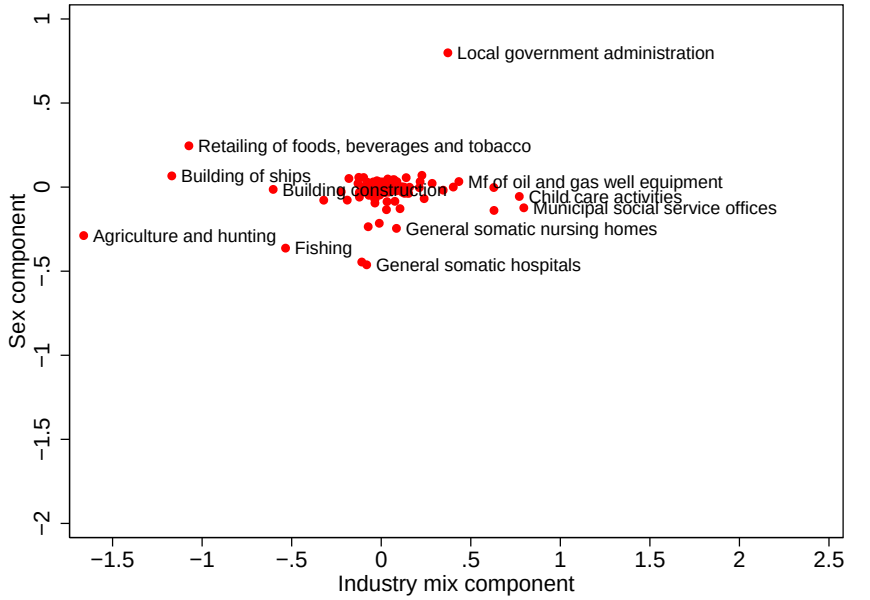
3.B Decomposition of dissimilarity index by education

This appendix contains the Figures that are referred to in Section 3.8.

Figure 3.B.1: Decomposition of dissimilarity index in compulsory education 1970-1980

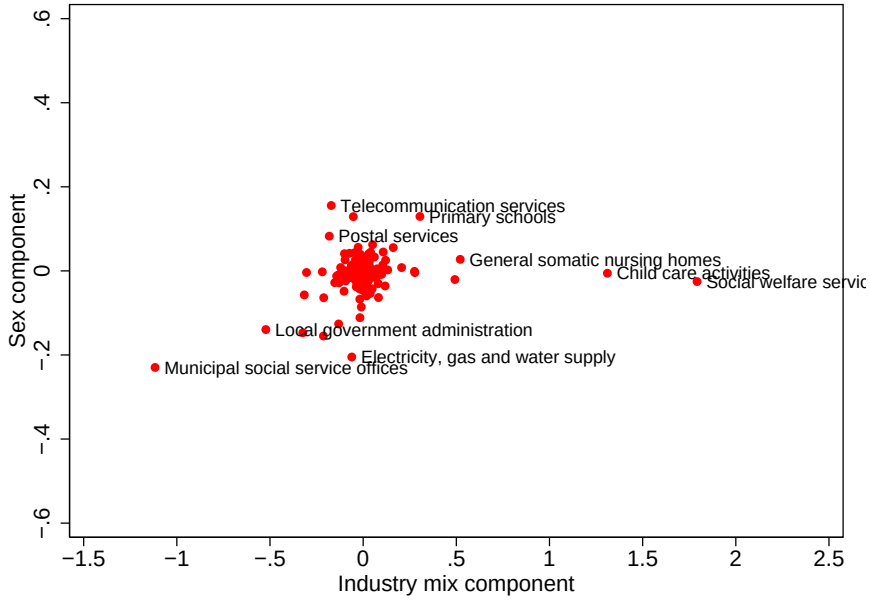
Note: See notes for Figure 3.6.

Figure 3.B.2: Decomposition of dissimilarity index in compulsory education 1980-1990



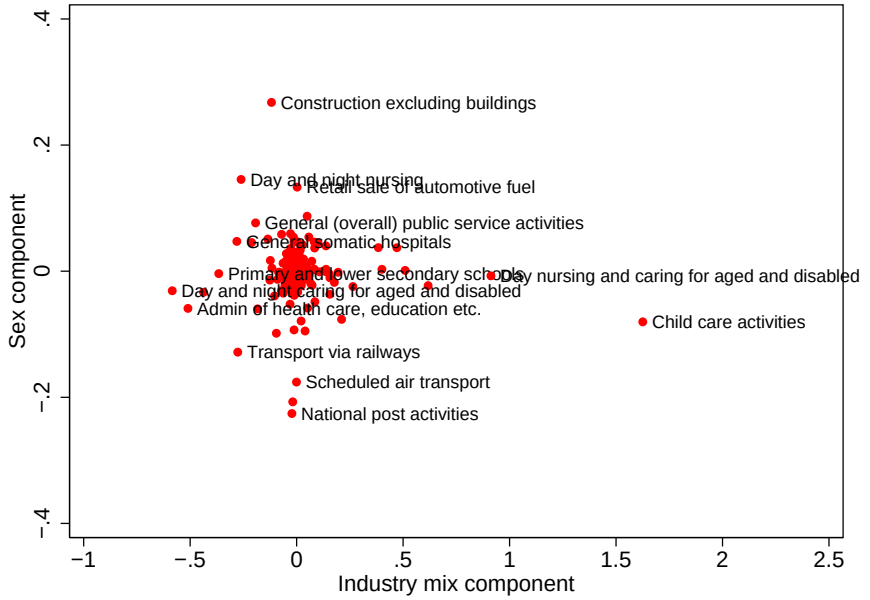
Note: See notes for Figure 3.6.

Figure 3.B.3: Decomposition of dissimilarity index in compulsory education 1990-1998



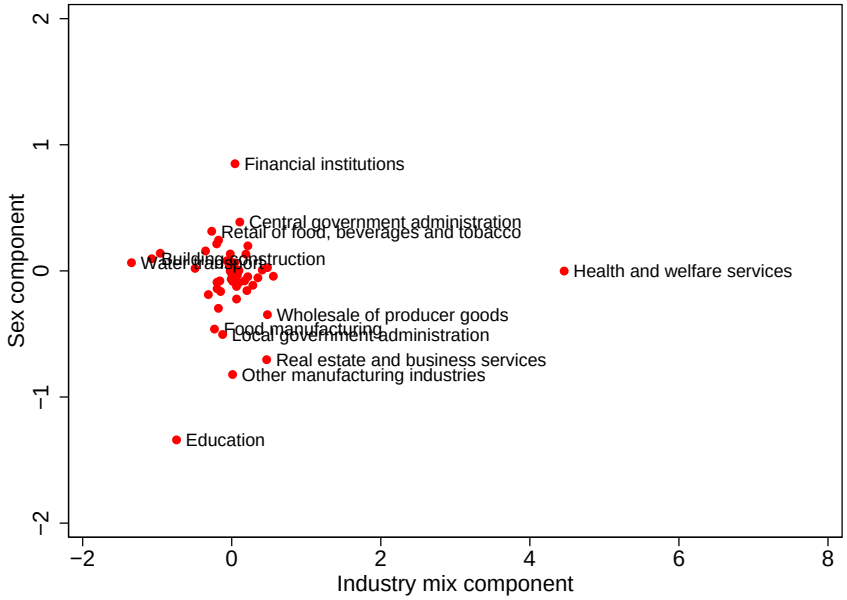
Note: See notes for Figure 3.6.

Figure 3.B.4: Decomposition of dissimilarity index in compulsory education 1999-2009



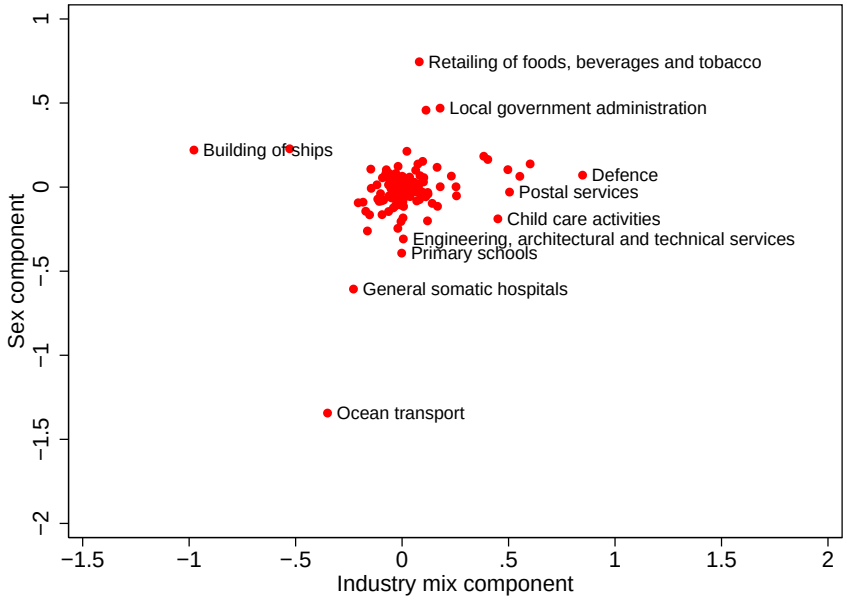
Note: See notes for Figure 3.6.

Figure 3.B.5: Decomposition of dissimilarity index in secondary education 1970-1980



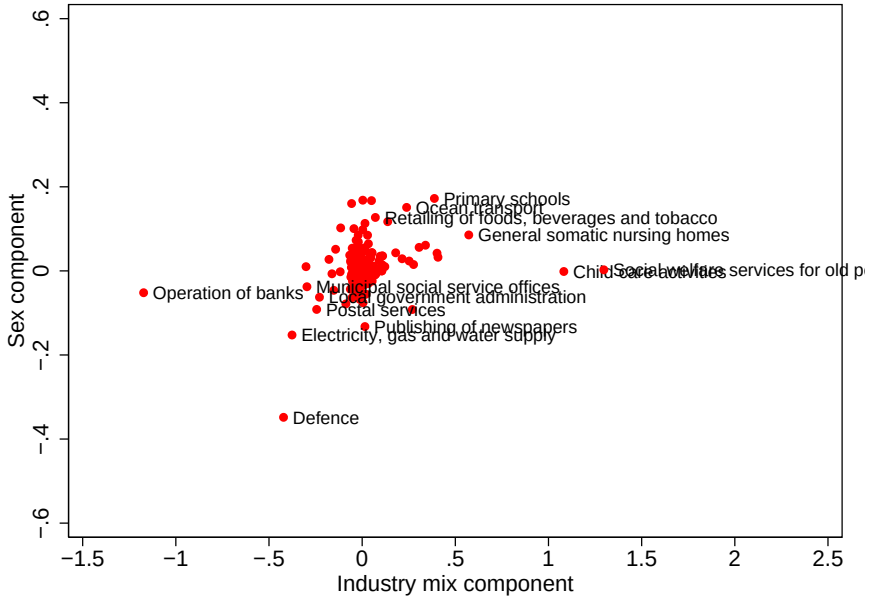
Note: See notes for Figure 3.6.

Figure 3.B.6: Decomposition of dissimilarity index in secondary education 1980-1990



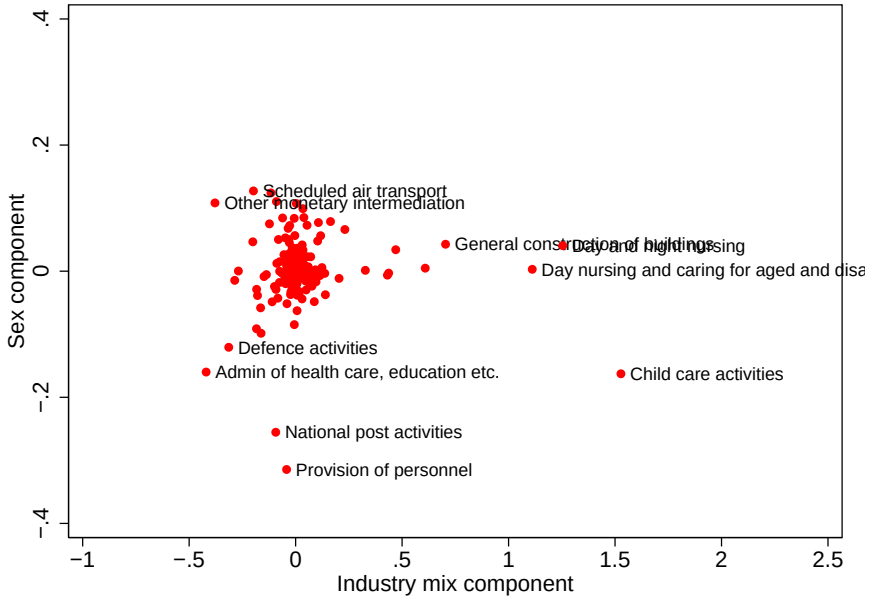
Note: See notes for Figure 3.6.

Figure 3.B.7: Decomposition of dissimilarity index in secondary education 1990-1998



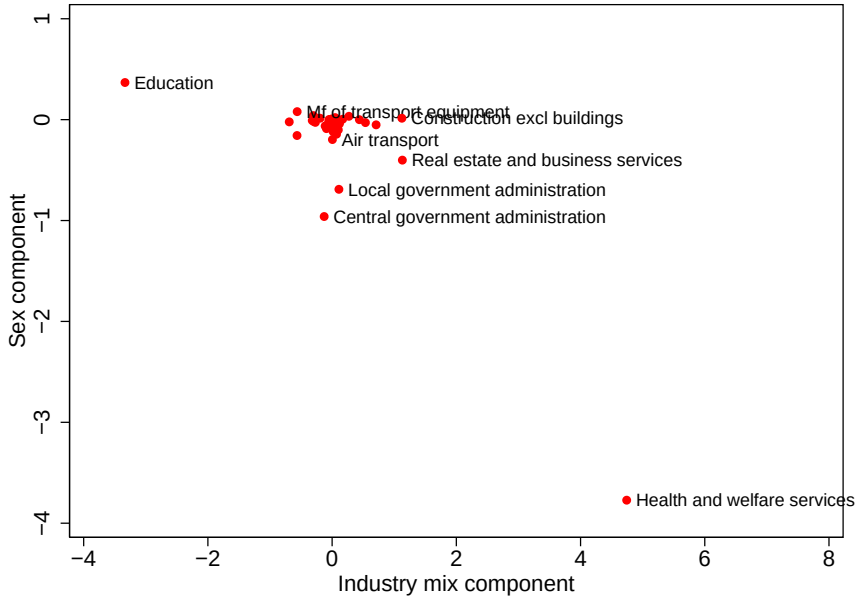
Note: See notes for Figure 3.6.

Figure 3.B.8: Decomposition of dissimilarity index in secondary education 1999-2009



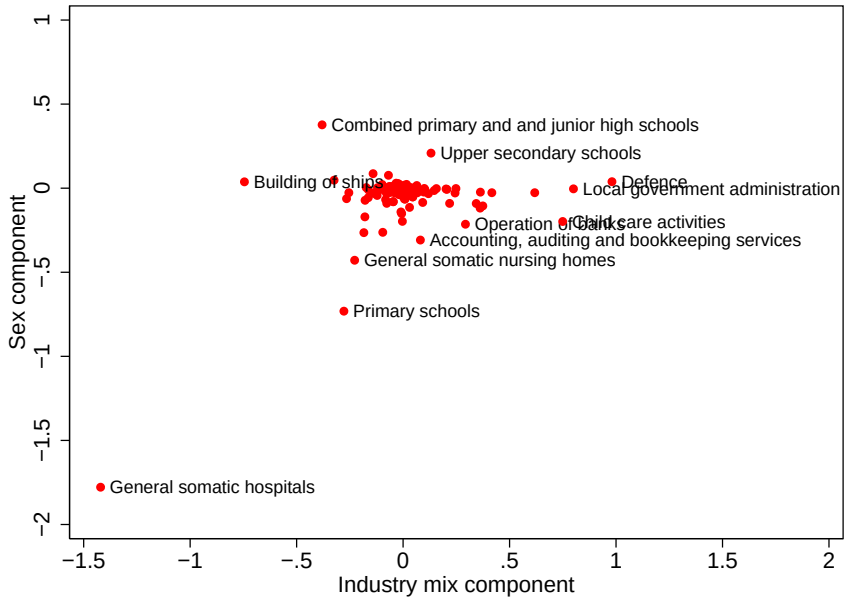
Note: See notes for Figure 3.6.

Figure 3.B.9: Decomposition of dissimilarity index in short tertiary education 1970-1980



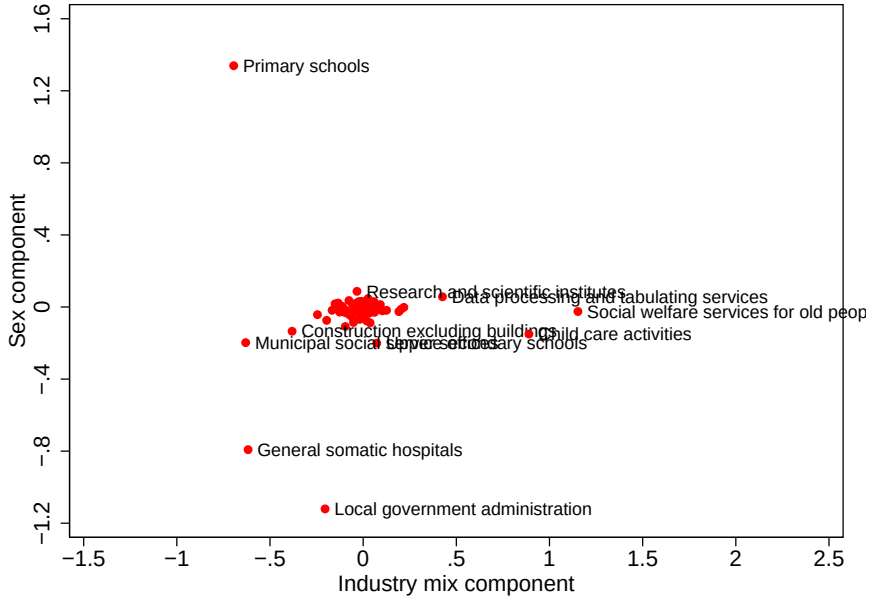
Note: See notes for Figure 3.6.

Figure 3.B.10: Decomposition of dissimilarity index in short tertiary education 1980-1990



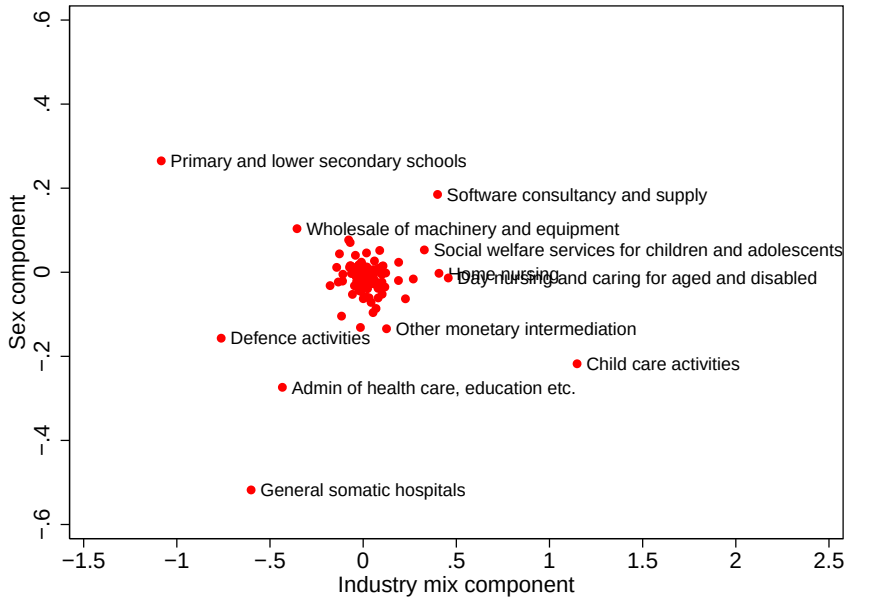
Note: See notes for Figure 3.6.

Figure 3.B.11: Decomposition of dissimilarity index in short tertiary education 1990-1998



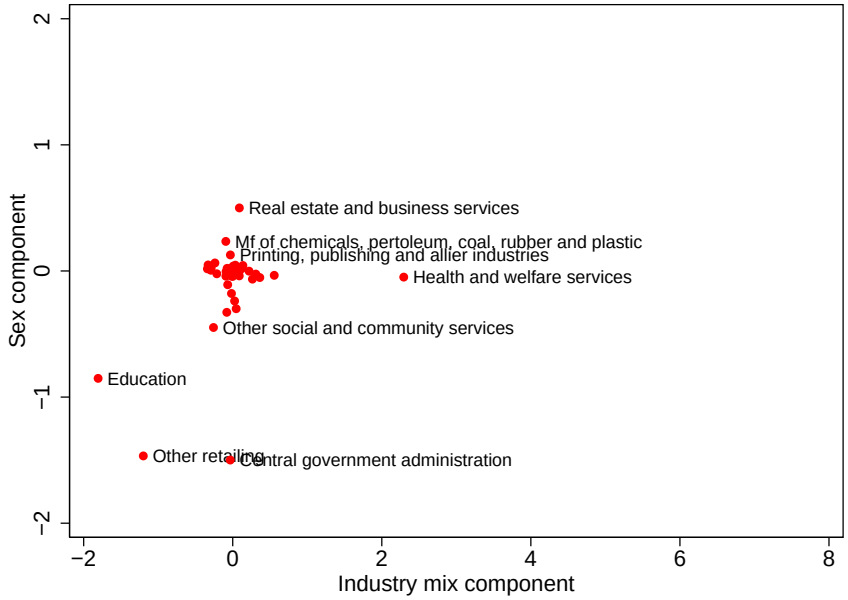
Note: See notes for Figure 3.6.

Figure 3.B.12: Decomposition of dissimilarity index in short tertiary education 1999-2009



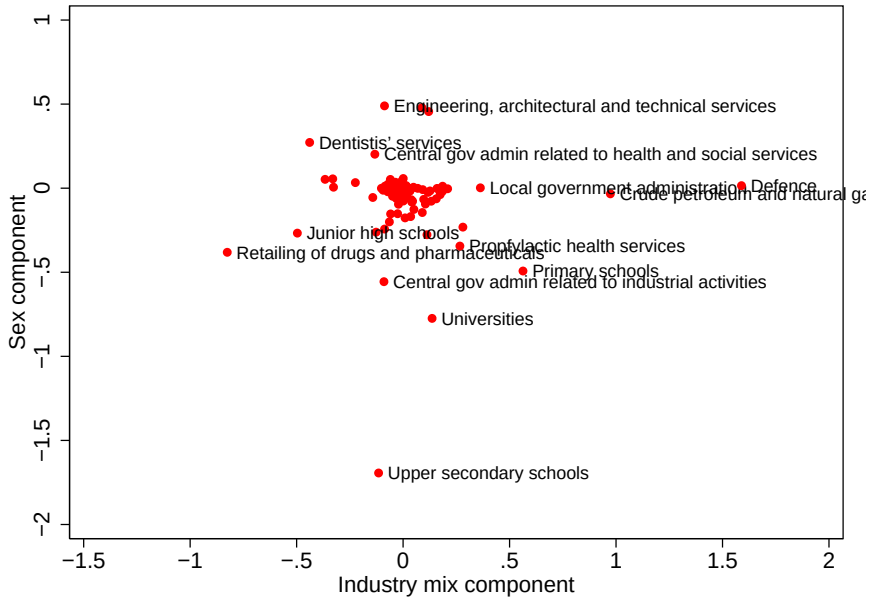
Note: See notes for Figure 3.6.

Figure 3.B.13: Decomposition of dissimilarity index in long tertiary education 1970-1980



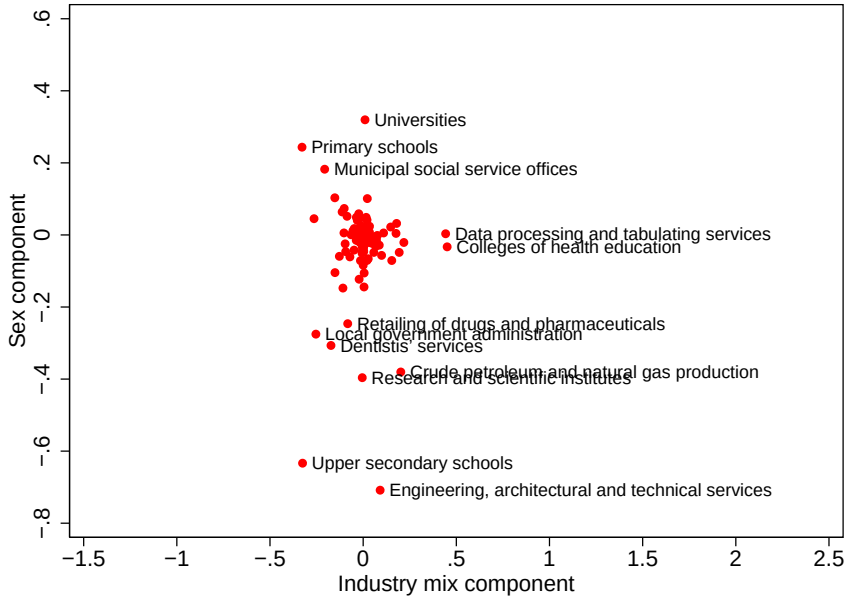
Note: See notes for Figure 3.6.

Figure 3.B.14: Decomposition of dissimilarity index in long tertiary education 1980-1990



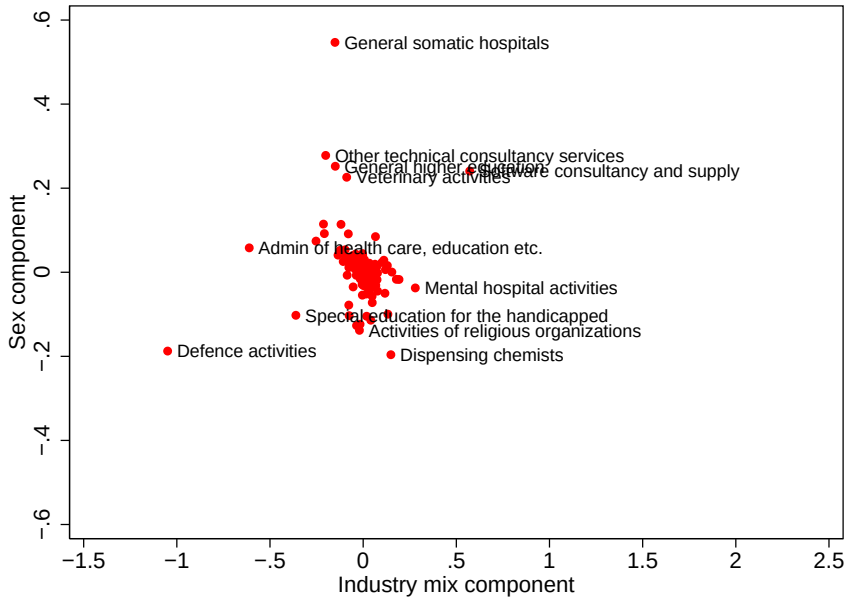
Note: See notes for Figure 3.6.

Figure 3.B.15: Decomposition of dissimilarity index in long tertiary education 1990-1998



Note: See notes for Figure 3.6.

Figure 3.B.16: Decomposition of dissimilarity index in long tertiary education 1999-2009



Note: See notes for Figure 3.6.

Bibliography

Aigner, D. J. and G. G. Cain (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review 30*(2), 175–187.

Anker, R. (1998). *Gender and Jobs: Sex Segregation of Occupations in the World*. Geneva: ILO.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter and A. Rees (Eds.), *Discrimination in Labor Markets*, Chapter The Theory of Discrimination. Princeton, NJ: Princeton University Press.

Bayard, K., J. Hellerstein, D. Neumark, and K. Troske (2003, October). New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data. *Journal of Labor Economics 21*(4), 887–922.

Becker, G. (1985). The allocation of effort, specific human capital, and differences between men and women in earnings and occupations. Technical report, Chicago - Population Research Center.

Becker, G. S. (1957). *The Economics of Discrimination*. The University of Chicago Press.

Bertaux, N. E. (1991). The roots of today's"women's jobs" and "men's jobs": Using the index of dissimilarity to measure occupational segregation by gender. *Explorations in Economic History 28*(4), 433–459.

Bertrand, M. (2011). *New Perspectives on Gender*, Volume 4 of *Handbook of Labor Economics*, Chapter 17, pp. 1543–1590. Elsevier.

Bettio, F. and A. Verashchagina (2009). *Gender segregation in the labour market: root causes, implications and policy responses in the EU*. Directorate-General for Employment, Social Affairs, and Equal Opportunities. European Commission.

Birkelund, G. E. and T. Sandnes (2003). Paradoxes of welfare states and equal opportunities: Gender and managerial power in Norway and the USA. In *Comparative Studies of Culture and Power*, Chapter 10, pp. 203–242.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2008). Staying in the Classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births. *The Economic Journal 118*(530), 1025–1054.

Blackburn, R. M., J. Browne, B. Brooks, and J. Jarman (2002). Explaining gender segregation. *The British Journal of Sociology 53*(4), 513–536.

Blau, F., P. Brummund, and A. Liu (2013). Trends in occupational segregation by gender 1970-2009: Adjusting for the impact of changes in the occupational coding system. *Demography 50*(2), 471–492.

Blau, F., M. A. Ferber, and A. E. Winkler (2013). *The Economics of Women, Men and Work* (7th ed.). Prentice Hall.

Blau, F. D., P. Simpson, and D. Anderson (1998). Continuing Progress? Trends in Occupational Segregation in the United States over the 1970s and 1980s. Working Paper 6716, National Bureau of Economic Research.

Boraas, Stephanie; Rodgers, W. M. (2003). How does gender play a role in the earnings gap? An update. *Monthly Labor Review 126*(3), 9–15.

Bridges, W. P. (2003). Rethinking gender segregation and gender inequality: Measures and meanings. *Demography 40*(3), pp. 543–568.

Carneiro, P., K. V. Lken, and K. G. Salvanes (2015). A Flying Start? Maternity Leave Benefits and Long-Run Outcomes of Children. *Journal of Political Economy 123*(2), pp. 365–412.

Carrington, W. J. and K. R. Troske (1997). On measuring segregation in samples with small units. *Journal of Business & Economic Statistics 15*(4), pp. 402–409.

Charles, M. (1992). Cross-national variation in occupational sex segregation. *American Sociological Review 57*(4), pp. 483–502.

Daatland, S. O. (1997). De siste tiårene. Eldreomsorgen i Skandinavia 1960-95. NOVA-rapport 22/1997, Norsk institutt for forskning om oppvekst, velferd og aldring.

Duncan, O. and B. Duncan (1955). A methodological analysis of segregation indexes. *American Sociological Review 20*(2), 210–217.

Ellingsæter, A. L. (2013). Scandinavian welfare states and gender (de) segregation: Recent trends and processes. *Economic and Industrial Democracy 34*(3), 501–518.

Ellingsæter, A. L. and L. Gulbrandsen (2007). Closing the Childcare Gap: The Interaction of Childcare Provision and Mothers' Agency in Norway. *Journal of Social Policy 36*, 649–669.

England, P. (1992). *Comparable Worth: Theories and Evidence*. Aldine De Gruyter.

Fuchs, V. R. (1975). A Note on Sex Segregation in Professional Occupations. In *Explorations in Economic Research, Volume 2, number 1*, NBER Chapters, pp. 105–111. National Bureau of Economic Research, Inc.

Grusky, D. and M. Charles (1998). The past, present, and future of sex segregation methodoloqy. *Demography 35*(4), 497–504.

Hakim, C. (1993). Segregated and integrated occupations: A new approach to analysing social change. *European Sociological Review 9*(3), 289–314.

Håland, I. and G. Daugstad (2003). Den kjønnsdelte arbeidsmarknaden. *Samfunnsspeilet* (6).

Hallden, K. (2014). *Yrke, karriär och lön: kvinnors och mäns olika villkor på den svenska arbetsmarknaden*, Chapter Könssegregering efter yrke på men svenska arbetsmarknaden år 2000-2010, pp. 49–75. Number 2014:81 in Statens offentliga utredningar. Delegationen för jämställdhet i arbetslivet.

Halvorsen, K. and S. Stjernø (2008). *Work, oil and welfare*. Universitetsforlaget.

Hansen, M. N. (1995). Sex segregation and the welfare state. Report 95:2, Institutt for samfunnsforskning.

Havnes, T. and M. Mogstad (2011a). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics 95*(11), 1455–1465.

Havnes, T. and M. Mogstad (2011b). No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy 3*(2), 97–129.

Hodne, F. and O. H. Grytten (2002). *Norsk økonomi i den 20. århundrede*. Fagbokforlaget.

Høst, H. and M. Evensen (2009). Ny struktur - tradisjonelle mønstre? Rapport 28/2009, NIFU STEP.

International Labour Organization (2015). Gaining Momentum - Global report. ILO, Bureau for Employers' Activities (ACT/EMP).

Jacobs, J. A. (1989). Long-term trends in occupational segregation by sex. *American Journal of Sociology 95*(1), pp. 160–173.

Jensberg, H., R. Mandal, and E. Solheim (2012). Det kjønnsdelte arbeidsmarkedet 1990-2010. Kontinuitet eller endring? Rapport, SINTEF.

Jrgensen, T. (1997). Utdanning i Norge. Statistiske analyser 19, Statistics Norway.

Karmel, T. and M. MacLachlan (1988). Occupational sex segregation - increasing or decreasing? *Economic Record 64*(3), 187–195.

Leira, A. (1992). *Welfare states and working mothers*. Cambridge University Press.

Levanon, A., P. England, and P. Allison (2009). Occupational Feminization and Pay: Assessing Causal Dynamics Using 1950-2000 U.S. Census Data. *Social Forces 88*(2), 865–891.

Mandel, H. and M. Semyonov (2006). A welfare state paradox: State interventions and women's employment opportunities in 22 countries. *American Journal of Sociology 111*(6), pp. 1910–1949.

NOU (1997). Pasienten først! NOU 1997:2, Sosial- og helsedepartementet.

NOU (2012). Politikk for likestilling. Norges offentlige utredninger 2012:15, Barne-, likestillings- og inkluderingsdepartementet.

OECD (2013). *Technical Report of the Survey of Adult Skills (PIAAC)*. OECD.

OECD (2014). Education at a Glance 2014: OECD Indicators.

Olsen, T. E. (2004). Kjønnssegregering i industrisektoren. *Søkelys på arbeidsmarkedet 21*(1).

Otterstad, H. K. (2013). Historisk blikk på eldrebølgen. *Sykepleien 101*(2), 58–60.

Phelps, E. (1972). The statistical theory of racism and sexism. *The American Economic Review 62*(4), 659–61.

Polachek, S. W. (1978). Sex differences in college major. *Industrial and Labor Relations Review 31*(4), pp. 498–508.

Polachek, S. W. (1981). Occupational self-selection: A human capital approach to sex differences in occupational structure. *The Review of Economics and Statistics 63*(1), pp. 60–69.

Reisel, L. and I. Brekke (2013). Kjønnssegregering i utdanning og arbeidsliv: status og årsaker. Rapport 2013:04, Institutt for samfunnsforskning.

Reskin, B. F. and D. D. Bielby (2005). A sociological perspective on gender and career outcomes. *Journal of Economic Perspectives 19*(1), 71–86.

Rubery, J., M. Smith, and C. Fagan (1999). *Women's Employment in Europe : Trends and Prospects*. Routledge.

Solheim, E. (2012). Det kjønnsdelte arbeidsmarkedet: horisontal og vertikal kjønnssegregering, en sammenligning mellon Norge og EU-land 1998-2010. Arbeidsnotat til likestillingsutvalget.

Statistics Norway (1978). Standard industrial classification. Statistisk sentralbyrås håndbøker 2, Statistisk sentralbyrå.

Statistics Norway (1983). Standard industrial classification. Standarder for norsk statistikk 9, Statistisk sentralbyrå.

Statistics Norway (1994). Standard industrial classification 1994.

Statistics Norway (1995). Historical statistics 1994. NOS C 188, Statistics Norway.

Statistics Norway (2002). Standard industrial classification 2002.

Statistics Norway (2007). Standard industrial classification 2007.

Statistics Norway (2015a). Annual national accounts. Table 09174: Wages and salaries, employment and productivity, by industry.

Statistics Norway (2015b). Employment, register based. Table 07984: Employed persons per 4th quarter, by place of residence, place of work, sex, age and industry division.

Statistics Norway (2015c). Kindergartens. Table 09344: Employed persons in kindergartens, by ownership, employment position and sex.

Statistics Norway (2015d). Sysselsatte personer etter næring 1900-1969.

Teigen, M. (2006). Det kjønnssegregeringnnsdelte arbeidslivet: En kunnskapsoversikt. Rapport 2006:2, Institutt for samfunnsforskning.

Try, S. (2000). Veksten i høyere utdanning: Et vellykket arbeidsmarkedstiltak? Rapport 2/2000, Norsk institutt for studier av forskning og utdanning.

UNDP (2010). *Human Development Report 2010 - The Real Wealth of Nations: Pathways to Human Development*. Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

UNDP (2011). *Human Development Report 2011 - Sustainability and Equity: A Better Future for All*. Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

UNDP (2013). *Human Development Report 2013 - The Rise of the South: Human Progress in a Diverse World*. Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

UNDP (2014). *Human Development Report 2014 - Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience*. Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Vassenden, K. (1987). Folke- og boligtellingene 1960, 1970, 1980. Dokumentasjon av de sammenlignbare filene. Rapporter fra Statistisk sentralbyrå 87/2, Statistisk sentralbyrå.

Watts, M. (1998). Occupational gender segregation: Index measurement and econometric modeling. *Demography 35*(4), pp. 489–496.