CBS
COPENHAGEN
BUSINESS SCHOOL
HANDELSHØJSKOLEN

Gender Differences in Returns to Skills

MSc in Business Administration and Philosophy
Master's Thesis

Mathias Fjællegaard Jensen

Student number:	113047
Supervisor:	Fane Naja Groes
Number of characters:	131,373
Number of pages:	64
Date of Submission:	15/11/2018

Abstract

Women, men, and their differences in the labour market are widely studied, e.g. in terms of segregation, participation, pay gaps, and discrimination. In this literature, workers are normally categorised into mutually exclusive occupations, so that job characteristics and their variation across occupations can be imported from databases such as O*NET and DOT. However, recently available data from online job vacancies enable analyses that move beyond across-occupation variation to also include within-occupation variation in terms of skills: in which occupations, but also, in which firms do workers employ certain task-specific skills? Such a question can be answered by the use of job vacancy data alone. More interestingly, I also test how the employment of skills and their returns depend on the gender of the worker by exploiting a novel combination of Danish job vacancy data and Danish individual-level register data. I am the first to operationalise this combination of data.

I use the novel combination of vacancy and register data to show that, in aggregate, similar skills are required by women and men. Some gender differences are observed, but they are small relative to the high degree of gender segregation in the labour market. Furthermore, variation in skills cannot be fully accounted for by occupations and other sets of covariates. Even with a very extensive set of control variables, regression analyses indicate that women face lower returns to character, writing/language, customer service, management, financial, and computer skills when compared to men. This result is driven by workers in the private sector. Whereas these results are merely correlational, they serve to warrant future casual studies of the impact of task-specific skills on wages, and thus, om the gender pay gap.

An empirical analysis of gender differences in the deployment of and returns to skills does not only provide further insights into pay differentials between women and men. Equally important, my analysis of skills opens up a discussion of the difference between the female worker and female work, as well as between the male worker and male work. This represents a move away from the essentialist approach taken by many economists where no such distinction is made. The distinction becomes vital as the skills framework breaks down mutually exclusive occupational categories, which are inherently gendered. Occupational gender segregation is costly, not only by causing gender pay inequalities, but also because of the tokenisation of workers whose occupational choice and gender identity do not align. However, reconceptualising work through the skills framework facilitates a transgression of the boundaries between occupations, and thus, between female and male work. Female work becomes multi-dimensional and so does its male counterpart; it becomes clear that the perceived reality of the gendered worker is only superficially constructed. This realisation allows a move beyond segregation; a move beyond an occupational division of workers to a multi-dimensional approach to work.

Conceptualising work as skills, rather than simply occupational categories, therefore both enrich quantitative studies of pay inequalities, and it potentially allows workers to forsake identities of work based on inherently gendered occupational categories.

Table of Contents

List of Figures

List of Tables

Introduction

Conceptualising work as the utilisation of skills, rather than as mutually exclusive occupational categories, has gendered implications. These gendered implications form an overarching theme of this thesis. An empirical reconceptualisation of work as skills is enabled by the combination of Danish job vacancy data and Danish register data. By operationalising and utilising this novel combination of vacancy and register data, it is possible to examine if different skills are required by women and men. Furthermore, I can evaluate if differences in returns to skills contribute towards gender pay differentials both *across* and *within* occupations.

Occupational categorisation plays a central role in the development of workers' social identities. Predominantly, work is conceptualised as a limited set of mutually exclusive occupational categories, which are inherently gendered. Conceptualising work as the utilisation of skills with an indefinite number of potential values and combinations is therefore not without implications. Through an application of the skills framework, workers may realise that they are similar across occupations and different within occupations. Occupational gender segregation is costly, not only because of gender pay inequalities, but also because of the tokenisation of workers whose occupational choice and gender identity do not align. The skills framework may enable transgressions of the gendered boundaries between occupational categories.

Accordingly, I aim to answer the following three questions by reconceptualising work as the utilisation of skills:

1. Do women and men face different returns to the same task-specific skills? E.g. social skills, cognitive skills, management skills.
2. Does ignoring the gender dimension in returns to skills lead to biased estimates of returns to task-specific skills?
3. How may a reconceptualisation of work as the utilisation of skills affect the gendered occupational identities of workers and with which implications?

In order to answer these questions, the thesis at hand is divided into two major parts. In Part I, I undertake an empirical analysis of skills and answer questions 1 and 2. The empirical analysis opens up a theoretical discussion of gendered work and workers, answering question 3, forming Part II of the thesis. Together, the two parts exemplify a combination empirical and theoretical analyses, for which the study programme MSc in Business Administration and Philosophy allows. In other words, my thesis *"...is organized as an encounter between philosophical theory and a specific theme that has relevance to business practice or general economics."*[1]

[1] https://www.cbs.dk/uddannelse/kandidat/msc-in-business-administration-and-philosophy [accessed 3 September 2018]

Methodologies

At first, a definition of "skill" is necessitated. In the economics literature, "skill" has been used both to denote education levels and task-specific skills, such as social and cognitive skills. I follow the second tradition and define skills as follows:

> *Occupational skill requirements are based on the activities that employees have to perform at the workplace. We pool these activities into [...] task categories... (Black & Spitz-Oener 2010)*

Note how skill requirements and tasks categories are equivalent according to Black and Spitz-Oener. To simplify my terminology, I simply refer to skills in what follows below. In Part I, the extraction of skills from job posts is described in detail.

In order to answer research questions 1 and 2, I rely on quantitative methods, which are traditionally applied within the discipline of economics. In addition to simple statistical tools, also linear regressions, quantile regressions and Oaxaca-Blinder decompositions are applied. These methods are further described in Part I. This quantitative approach makes it possible to pinpoint gender differences in returns to skills in aggregate. A shortfall of the choice of methods, however, is that the presented results cannot be interpreted as causal evidence of gender differences in returns to skills.

Question 3 is approached in a very different way. First, I revisit theories of work and identity from economic sociology and social psychology in order to establish the interdependence of work, occupations, and the identities of workers. Next, I apply a deconstructionist philosophical approach to show how gender and occupations are "done" or performatively produced (Butler 2006; West & Zimmerman 1987). Part II is therefore a theoretical enquiry, but it remains motivated by the above-mentioned quantitative analysis throughout.

I do not adhere to either the quantitative or non-quantitative "paradigm", but rely on methods from both (for more details on the "paradgim wars," see Jensen 2017). By bringing different methodological approaches together, I reconceptualise work as the utilisation of skills both quantitatively and theoretically with the aim of furthering the understanding of gendered labour market outcomes.

PART I: Gendered Returns to Skills

Introduction

By analysing data from two sources 1) Danish job vacancy data, and 2) Danish register data, I aim to elucidate gender differences in the deployment of skills and particularly in the returns to skills in the Danish labour market. Existing literature has already documented the increasing importance of social and cognitive skills and their interaction in the modern labour market (Deming 2017; Deming & Kahn 2018). Furthermore, it is well-known that a significant and sizeable gender gap in earnings remain, especially in the top of the income distribution (Blau & Kahn 2017). Thus, an obvious question arises: can the gender pay gap, at least in part, be explained by different returns to the same skills? Beaudry and Lewis (2014) find that changing returns to education do indeed seem to account the for narrowing of the gender pay gap over the period 1980-2000. However, the literature on gender differences in task-specific skills, such as social and cognitive skills, is sparse.

With the internet's omnipresence in the Global North, online posting of job vacancies is now an integrated part of firms' recruitment of new employees. As firms' recruiting processes were digitalised, a new data source also became readily available. The text of each job post is highly informative when studying modern labour markets: Typically, job posts state expected skills, education, and experience of potential applicants, as well as certain characteristics of the job itself, e.g. its occupation, industry, and region. Crucially, the text of digital job posts can easily be scraped from various sources on the web. I have access to a recently available dataset that contains all Danish online job posts, including job descriptions, covering the period 2010-2016. From the job posts, data is obtained on demanded skills, occupations, as well as firm identifiers by applying computer-based text analysis and machine learning methods. Most importantly, these data can be matched with the Danish register data at the firm*occupation*month-level. This combination of data has not been operationalised before, and thus, the operationalisation of the data constitutes one of the key contributions of this thesis.

Utilising the novel combination of firm-level vacancy and individual-level register data, it is possible to evaluate gender differences in returns to skills both *across* and *within* occupations. Therefore, in Part I, I aim to answer the following two questions:

1. Do women and men face different returns to the same task-specific skills? E.g. social skills, cognitive skills, management skills.
2. Does ignoring the gender dimension in returns to skills lead to biased estimates of returns to task-specific skills?

Throughout, keep in mind that the estimated returns to skills are correlations, and they should not be interpreted as causal effects. However, the analyses that follow constitute the first operationalisation of the combination of Danish job vacancy data and register data, and thus, it provides the first step in an exhaustive analysis of returns to skills in the Danish labour market. In the following section, the relevant existing literature is outlined, and its findings are discussed. After that, I move on to present the Danish job vacancy data, the Danish register data, and my analyses of these data.

Background

At least three related literatures are relevant to consider when studying gender differences in returns to skills using a combination job vacancy data and register data. First, the literature on the gender pay gap should be considered. The study at hand adds another perspective to the discussion of gender pay differentials. Rather than merely considering gender pay differentials at the mean, adding a skills dimension has the potential to further decompose pay differentials.

Secondly, the literature on technological change, skills, and tasks in the labour market is relevant to consider. Through the skilled-biased technological change hypothesis, economists have argued that technological change favoured skilled (here: highly educated) workers by augmenting their productive potential in the labour market, and thus, increasing their wages. Since Autor, Levy, and Murnane's (2003) famous study, however, the effects of technological change have mainly been studied in the context of disaggregate task-specific skills. After introducing the general literature on technological change and employment polarisation, I move on to studies which consider gender differences in skills and in employment polarisation.

Thirdly, I consider the increasing number of studies in which job vacancy data are analysed to further the understanding of economic outcomes. Increasing computer power has – together with new methods for collecting and codifying text data – provided economists with this new source of information about labour market outcomes. Most of the studies that utilise job vacancy data also explore technological change in the labour market, but they do not tend to point at gender differences in outcomes.

Gender in the labour market

The empirical analysis that follows in the chapters below draw on two well-known conclusions from the literature on gender differences in labour market outcome. Firstly, women still receive substantially lower hourly wages when compared to men. Nevertheless, some convergence in women's and men's labour market outcomes has been observed internationally over the last few decades, both in terms of hours worked, earnings and educational attainment (Blau & Kahn 2017; Goldin 2014; Lindley & Machin 2012; Olivetti & Petrongolo 2016). The "unexplained" gender pay gap remains even after controlling for large sets of covariates, including human capital. Due to the convergence of educational attainment, human capital differences explain little of the gender pay gap today (Blau & Kahn 2017). Despite today's similar levels of human capital, wage differentials between women and men are still observed. Even *within* occupations, large and significant gender pay gaps remain (Blau & Kahn 2017; Goldin 2014; Lindley 2016). Goldin (2014) points out that some of the wage differentials within occupations can be explained by non-linear returns to hours. Working longer hours pays off as the hourly returns increase for each additional hour worked. Thus, "flexibility" in hours is needed to further narrow the gender pay gap, as women still undertake most of the work in the home and caregiving in the family. Utilising Danish register data, Kleven, Landais, & Søgaard (2018) further emphasise parenthood's gendered effects on earnings. Women's earnings and hours decrease substantially when becoming mothers whereas fathers' earnings are unaffected. Furthermore, they also document how women's occupational choices are affected by childbirth. Hence, when becoming mothers, women are less likely to enter high paying and managerial occupations. This provides another argument for looking at gender differences in earnings both across and within occupations.

Secondly, women and men are, to a large extent, segregated in the labour market as they tend to work in different occupations and industries (Levanon & Grusky 2016; Olivetti & Petrongolo 2014). Counter-intuitively, occupational segregation is particularly pronounced in Scandinavian countries, including Denmark, despite the fact that we also observe some of the smallest gender pay gaps in these countries (Jarman, Blackburn, & Racko 2012). Gender segregation in the labour market has been pointed out as being both a cause and a cure of the gender pay gap. On the one hand, women are concentrated in occupations and industries with lower wages, e.g. because of occupational downgrading when working part-time (Manning & Petrongolo 2008), the "sticky floor" and because of "the glass ceiling" (Arulampalam, Booth, & Bryan 2007). On the other hand, the same segregation has also been emphasised as one of key drivers of the observed convergence in hours and earnings of women and men (Ngai & Petrongolo 2017; Olivetti & Petrongolo 2014, 2016). These studies find that women are concentrated in the service sector and that they have comparative advantage in producing services. Due to recent structural transformation, namely the expanding service sector, women's hours and wages have increased relatively to their male counterparts. Thus, in order to better understand the development of the gender pay gap, the changing structure of the labour market must be considered. The literature on technological change and its consequences for employment are considered in the next section.

Technological change, skills, and tasks

Numerous economic models and theories incorporate the impact of technology on the labour market. An extensive literature has developed on the topic of the skill-biased technological change hypothesis (e.g. Autor, Katz, & Krueger 1998; Machin 2001, 2011; Machin & Van Reenen 1998). This literature defines skill in terms of education levels and explores the complementarity between highly educated workers and new technologies. In order not to confuse education levels with my task-specific skill measures, I will refer to the skill-biased technological change hypothesis as the education-biased technological change hypothesis (EBTC). The studies on EBTC emphasise how returns to education have increased despite an increasing supply of highly educated workers. Increasing returns to education further wage inequalities between the highly educated and the less educated workers. Autor, Levy, and Murnane (2003), however, introduced an alternative hypothesis, the task- or routine-biased technical change (RBTC) hypothesis, arguing that technological change primarily impacts the labour market by substituting workers undertaking routine cognitive and routine manual tasks, and by complementing workers undertaking non-routine tasks. RBTC can explain some empirically observed patterns in the labour market, which cannot be explained by EBTC, e.g. job polarisation (Autor & Dorn 2013; Goos, Manning, & Salomons 2014). Not only have studies documented the impact of technological change on the labour market in the past, but technological change is also predicted to continue to impact workers, and the process of technological replacement of workers may even speed up. Frey and Osborne (2013) predict that 47 per cent of US jobs are at high risk of being computerised within the next two decades due to technological change. Looking at 702 detailed occupations, their research shows that the risk of replacement varies substantially across occupations. The level of risk depends on multiple factors, but in particular, they find strong negative relationships between both wages and educational levels, and an occupation's probability of being computerised. Hence, technology and technological change remains crucial to understand when considering the development of pay inequalities in modern labour markets.

Both the EBTC and RBTC hypotheses predict that workers in different industries and occupations will be differently impacted by technological change, depending on the education and task compositions within their industries and occupations respectively (Autor et al. 2003; Machin 2011). Overcrowding (cf. Bergmann 1974) in less-educated or more routine task-dominated occupations or industries is a likely consequence of technological replacement of workers. As Machin (2011) points out, it follows that the standard labour supply-demand model would predict rising wage inequalities between workers in the highly educated and the less educated occupations/industries, and between workers in routine and non-routine task-dominated occupations/industries. Goos and Manning (2007) show that job polarisation, predicted and explained by the RBTC hypothesis, accounts for a substantial share of the UK rise in wage inequality from 1976 to 1995. Importantly, Goos, Manning, and Salomons (2009; 2014) show that employment polarisation is not just a US and UK phenomenon, but that polarisation is also observed across Europe, including in Denmark, and that these patterns of polarisation cannot be explained by offshoring.

The link between task-specific skills, job polarisation and rising inequality have spurred research into the demand for certain skills and research on their prices or returns. As pointed out above, cognitive skills have traditionally been highlighted as complementing new technologies, and thus, the demand and returns to cognitive skills should increase with technological change. However, recent evidence suggests that since year 2000, the demand for cognitive skills reversed (Beaudry, Green, & Sand 2016). Instead, the interaction between social skills and cognitive skills seems to be of particular importance, rather than cognitive skills alone (Deming 2017; Weinberger 2014). Although it seems obvious that women and men could be differently endowed with cognitive and social skills, most of the above-mentioned studies on changing technology and skill demands rely on skills data at the occupation-level. Hence, gender differences in skills can only be inferred from the fact women and men tend to work in different occupations. A few studies explore this fact to evaluate gendered consequences of technological change.

Technological change and gender

Keeping in mind that both the demand of skills and the returns to skills have changed over time, an obvious question to ask is whether or not the labour market outcomes of women and men are differentially impacted by technological change. In order to answer this question, Black and Spitz-Oener (2010) match the German Qualification and Career Survey and the IAB employment sample at the occupation-level and show that women have experienced substantial growth in non-routine tasks from 1979 to 1999. Furthermore, they show that women have experienced a large decline in routine tasks, whereas men have not. Thus, they find that the pattern of polarisation noted by Autor, Levy, and

Murnane (2003) and predicted by RBTC is more pronounced for women than for men in Germany. In line with this, Cerina et al. (2017) confirm that gender differences in employment polarisation are pronounced also in the US, where patterns of polarisation are similarly driven by changes in women's employment.

Concurrently, Bacolod and Blum (2010) use US occupation-level skills data from DOT to show that the returns to people skills and cognitive skills increased from 1968-1990. Since women are particularly well-endowed with people skills and cognitive skills, they find that increasing returns to these two skills can explain up to 20 % of the decline in the gender pay gap. Beaudry and Lewis (2014) find a similar result, when they analyse how skills prices change with the adoption of IT. Because women are well-endowed with cognitive skills, which complement IT, Beaudry and Lewis find that gender pay gap narrows with the adoption of IT. Thus, IT driven changes in skill prices can explain more than 50 % of the narrowing of the gender pay gap between 1980 and 2000. Yamaguchi (2018) also look at DOT data and find little gender differences in the endowment of cognitive skills from 1980-2000, but large differences in motor or "brawn" skills, with which men are typically more endowed. As returns to motor skills declined significantly from 1980-2010, the gender gap narrowed. Rendall (2010) confirms these findings, also using DOT data. Rendall's model use gender differences in cognitive and "brawn" skills not only to predict the narrowing of the gender pay gap, but also to predict the timing of this narrowing of the gap.

All of the above-mentioned studies on the interaction between technological change and gender use skills and task data at the occupation-level, i.e. they do not observe within-occupation variation in skills. A few studies, however, have utilised survey data on individual-level variation in skills, which precisely enabled them to look at within-occupation gender variation in skills. Using UK Skills Survey data from 1997 and 2006, Lindley (2012) shows that women tend to be less endowed with skills that complement tasks related to technological change, and thus, concludes that overall women lost out on technological change, despite the increasing number of women obtaining a university degree. Bizopoulou (2017) use data on 9 countries from the OECD Survey of Adult Skills (PIAAC) to show that within 4-digit occupations task segregation tend to favour men, who typically undertake more high paying tasks. Thus, within-occupation gender variation in tasks explains some of the within-occupation gender pay gap. Fedorets (2014) apply a new method to impute data on tasks from the German Qualification and Career Survey onto the IAB employment sample (the same data used by Black & Spitz-Oener 2010). The new imputation method preserves some variation in skills within occupations, particularly between women and men. Using these data, Fedorets (2014) finds that gender-specific returns to tasks is one of the drivers of the gender pay gap. Lastly, Stinebrickner, Stinebrickner, and Sullivan (2018) use the Berea Panel Study, which follows cohorts of graduates from Berea College, to show that individual-level variation in tasks is an important predictor of wages as well as the gender wage gap. Because of the narrow selection of their sample, however, they warn not to interpret their level estimates of the pay gap in a broader context. Still, they emphasise that individual variation in tasks should be considered in the context of the gender wage gap. In sum, it is clear that the data on within-occupation variation, and thus, gender variation, in skills and tasks are sparse.

Current studies rely on small surveys or on matching data from different sources at a coarse level, but even so, they find that also in this context gender seems to play an important role.

Vacancy data

Before the availability of job vacancy data, researchers typically relied on skills and tasks data from relatively small surveys or from the O*NET- and DOT-databases, which were infrequently updated and provided job characteristics that only varied at the occupational level. Job vacancy data provide novel insights into variation in skill demands *within* occupations. With the internet's omnipresence in the Global North, online posting of job vacancies is now an integrated part of firms' recruitment of new employees. As firms' recruiting processes were digitalised, a new data source also became readily available. The text of job posts themselves is highly informative when studying modern labour markets: Typically, job posts state expected skills, education, and experience of potential applicants, as well as certain characteristics of the job itself, e.g. its occupation, industry, and region. Crucially, the text of digital job posts can easily be scraped from various sources on the web. The richness of the scraped job vacancy data becomes evident as the literature relying on these data expands. New analyses of the effects of business cycles (Hershbein & Kahn 2018; Modestino, Shoag, & Ballance 2016a), changing skill requirements (Atalay et al. 2018b, 2018a; Berkes, Mohnen, & Taska 2018; Deming & Kahn 2018; Grinis 2017; Modestino, Shoag, & Ballance 2016b), the geography of job search (Azar et al. 2018; Marinescu & Rathelot 2018), and the impact of unemployment insurance programmes (Marinescu 2017) all rely on job vacancy data from the US and the UK.

US

Before the availability of job vacancy data, researchers interested in job openings in the US relied on data from the Job Openings and Labor Turnover Survey (JOLTS). JOLTS provides information on job openings, but also job closings in terms of layoffs and quits.[2] However, firms participate in the survey on a voluntary basis, and the coverage is far from universal. When studying changing skill and task demands, the most utilised data sources are O*NET and the Dictionary of Occupational Titles (DOT). Both databases provide detailed information about job characteristics, but only at the occupational-level. Thus, more detailed data on both job vacancies and on skills are in high demand.

<u>Online job vacancy data</u>

First, job vacancy data from a single source, CareerBuilder.com, was utilised and applied in studies of economic outcomes by Marinescu and co-authors. Marinescu and Wolthoff (2016) exploit the text-based data, and find that job titles are particularly important when employers and employees match. As most job posts do not include a wage, job titles serve as an important signal of wages and explain more than 90% of the variance in posted wages (Marinescu & Wolthoff 2016:27). Marinescu (2017) use the same data to analyse the partial and general equilibrium effects of an extension of unemployment benefits in the US during the Great Recession. Lastly, Marinescu and Rathelot (2018) show that workers prefer jobs closer to their home, but that increasing mobility of workers would only

[2] https://www.bls.gov/jlt/jltwhat.htm

marginally decrease unemployment, since workers generally are able to find vacancies near their home. The three studies by Marinescu and co-authors demonstrate that job vacancy data has a different potential than traditional labour market data.

However, Hershbein and Kahn (2018) point out that relying on one source of job vacancy data, i.e. CareerBuilder.com, can limit representativeness of the included vacancies. Thus, numerous recent studies utilise another source of job vacancy data collected by Burning Glass Technologies (BG). BG scrape data from around 40,000 online job boards, and thus they believe to include the near universe of US vacancies posted online (Hershbein & Kahn 2018). The richness and representativeness of the data has enabled a variety of economic questions to be answered. For example, Azar et al. (2018) use the data to examine monopsony power or labour market concentration across the US. A significant number of studies exploit that job posts include skills requirements. Hershbein and Kahn (2018) show that during the Great Recession, skill requirements in job posts increase more in areas that were hit harder by the recession. Also utilising the BG data, Modestino, Shoag, and Ballance (2016b, 2016a) find a similar relationship between skill requirements and the availability of workers, i.e. that skill requirements increased during the recession and decreased again through the recovery. Furthermore, Berkes, Mohnen, and Taska (2018) find that skill-mismatch for college graduates have both immediate and long-term (6 years) negative effects on wages.

Of the papers utilising online job vacancy data, Deming and Kahn's (2018) is the one closest related to my analysis. First, they use BG data from 2010 to 2015 to extract 10 general skill measures at the firm*occupation level. Next, they match these skill measures to data on individual firms and to wage data from metropolitan statistical areas (MSA). Thus, they can estimate the relationships between skills and wages as well as between skills and firm performance. They find their skills measures generally correlate positively with both wages and firm performance. High paying and high performing firms require higher levels of social and cognitive skills. When a job requires both social and cognitive skills, they find that the level of wages is particularly higher than otherwise. Although Deming and Kahn (2018) explore variation in returns to skills both across and within occupations, they cannot say whether or not their results hold at the individual level. This follows from the fact that they cannot match their skills and firm data with employees, but only on wage data at the MSA-level. Furthermore, their data only covers the period from 2010 to 2015, so they can neither analyse time trends in skill demands nor trends in returns to skills.

Newspaper vacancy data

Online job vacancy data all come with the same restriction: the earliest data is available only from 2007 and onwards. At the same time, technological change is a long term phenomenon. In order to study long term trends, Atalay et al. (2018a, 2018b) extract and utilise new measures of tasks from job ads in a range of US newspapers from 1960 to 2000. With these data, they show that within-occupation task changes account for at least as much of the decline in routine tasks as across-occupation changes in employment over this long period of time. Furthermore, they show that the introduction of ICTs correlates with decreasing routine tasks and increasing nonroutine analytic tasks. Thus, the introduction of ICTs increases the college/skill premium.

Deming and Noray (2018) utilise BG data together with the newspaper data from Atalay et al. (2018a, 2018b) when studying the relationship between STEM jobs and technological change. This combination of data enables them to consider a longer period of time. They find that STEM workers face large returns to their STEM degree when entering the labour market, but that the returns decline over time. They explain this finding by showing that STEM skills are especially susceptible technological change, and thus, STEM skills become outdated relatively fast.

Cortes, Jaimovich, and Siu (2018) also analyse the newspaper data from Atalay et al. (2018a, 2018b) together with DOT data. Using both data sources, they evaluate changes in skill demands within occupations, over time. Significantly, they find that when social skills become more important within an occupation, the occupation's female share of employment also increases. After merging their skills measure to a sample of US census data, they also indicate that returns to social skills have increased over time. This consistent with Deming's (2017) findings.

Rest of the world
Job vacancy data from countries other than the US have only been studied in few papers. Grinis (2017) analyse UK online job vacancy data from 2012 to 2016, these data are also supplied by Burning Glass Technologies. Grinis finds that a large share of STEM jobs, i.e. jobs that require STEM skills, are in non-STEM occupations. If variation in STEM skills within occupations is ignored, the demand for STEM skills is underestimated. Kuhn and Shen (2013) analyse job vacancy data from the third largest job board in China, Zhaopin.com. They collect the universe of vacancies, but for some subperiods between 2008 and 2010. Gender discrimination in hiring is legal in China, and Kuhn and Shen utilise the Chinese job data when analysing how job posts differentially target women and men. Gender targeted ads are common in the sample. Interestingly, employers' gender preferences vary even within firms and within firm*occupation cells. Again, vacancy data's added dimension of within-occupation variation yields further insights into labour market outcomes.

Individual/employee match
Although job vacancy data enable analyses of labour market outcomes, which would be impossible to undertake with traditional data sources, e.g. of within-occupation variation in skills, all job vacancy datasets are constrained by the fact that information on the hired worker is hidden. Vacancy data is often matched with firm-level data, for example by using firm names in Deming and Kahn (2018), but matching at the individual-level is impossible in settings where only datasets with samples of workers are available, e.g. in the US. Austria is an exception. Their social security database provides information on the universe of workers, and thus, Kettemann, Mueller, and Zweimüller (2018) obtain matched vacancy-employer-employee data. As their data and their data matching procedure is the one closest related to mine, I describe it in detail. Specifically, they utilise job vacancy data from a single online job board, namely the state-run Austrian "Arbeitsmarktservice". The dataset is unique as it does not only contain a firm identifier for each posting firm, but also a personal identifier for each hired worker (but only if the person is hired through the "Arbeitsmarktservice"). Using these two sets of identifiers, they can match the vacancies

with the Austrian social security database which provides further information about workers and firms. However, the match on both firm and personal identifiers are often ambiguous. The dataset totals 5,354,139 vacancies from the period 1997-2014, and out of these 2,183,199 are matched on the firm-level. Next, 439,341 vacancies (or 8.2 % of the total) are matched on both the firm- and individual-level (see Kettemann, Mueller, & Zweimüller 2018:16-19 for more details on their match and sample restriction criteria). Finally, they use their matched vacancy-employer-employee to show that growing firms fill their vacancies faster, and that wages and vacancy duration are negatively correlated when adjusting for worker heterogeneity.

In sum, vacancies are becoming an increasing popular source of data in the applied labour economics, but a link to individual-level data on wages and other control variables continues to be hard to establish.

Why my analysis is needed

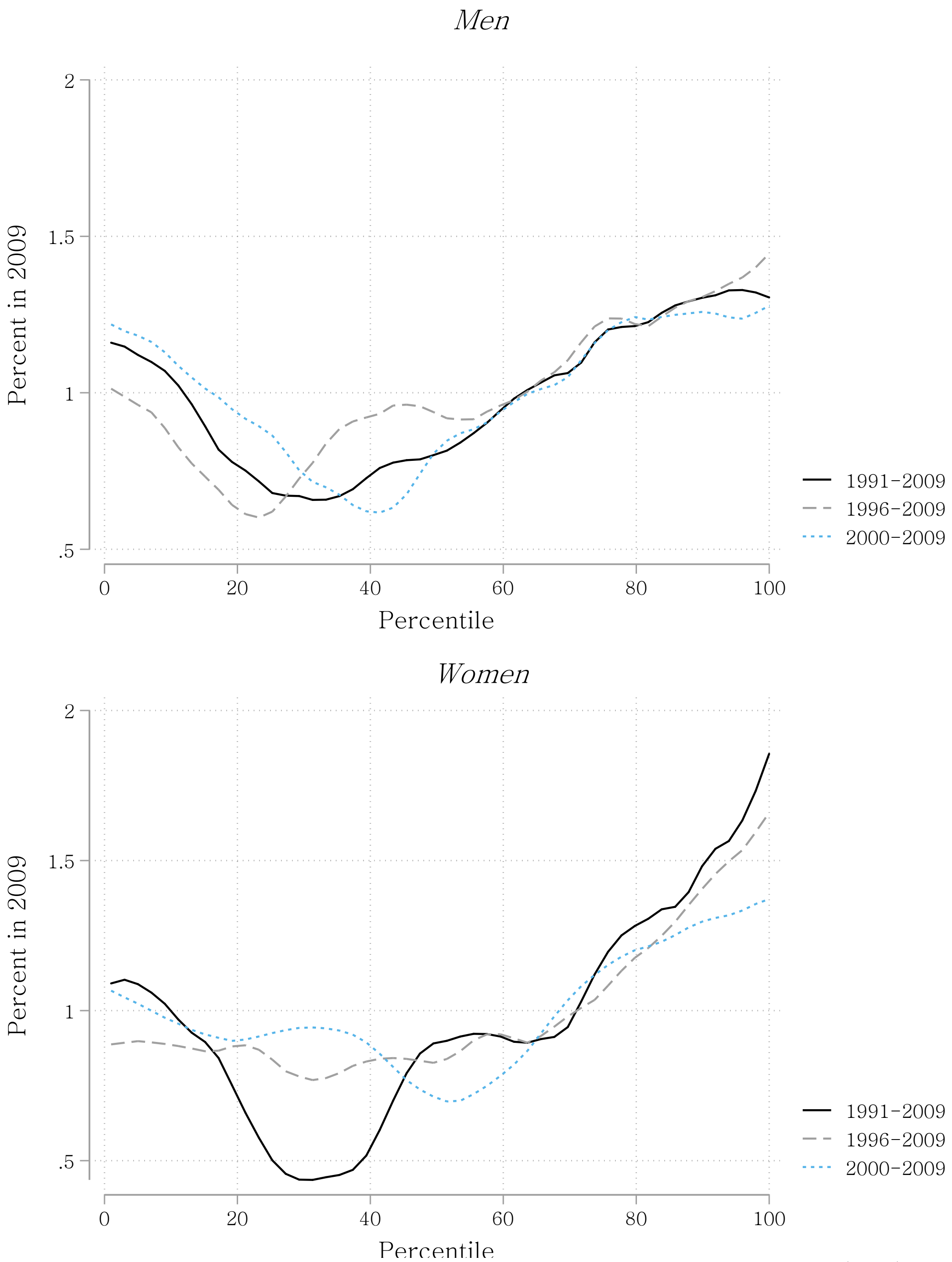
Going through the existing literature, a few cross-national trends stand out:

1) Women tend to be paid lower hourly wages than men, both within and across occupations. The pay gap has narrowed over the last few decades.
2) Women and men remain segregated in the labour market, i.e. high levels of occupational segregation are observed.
3) Technological change and disappearing routine jobs have led to employment polarisation. More so for women than for men.
4) Changes in skill demands and skills prices over the last few decades seem to favour women (this is supported by a few studies, but Lindley 2012 concludes the opposite).
5) Returns to cognitive skills alone have stalled, whereas the returns to the interaction of cognitive and social skills appear to have increased.
6) Vacancy data provide additional insight into variation in skill demands.
7) Vacancy data is difficult to match with individual-level data, even in settings where population data is available. Thus, vacancy data is also difficult to match with information on gender of hired workers.

I aim to connect the dots between these seven trends. Although it may seem like a big task, the novel combination of Danish vacancy and register data makes it possible to estimate returns to skills at the individual-level. Thus, it is also possible to examine how gender differences in returns to skills affect the gender pay gap *across* occupations and *within* occupations. In many ways, my analyses are similar to that of Deming and Kahn (2018), but as they only utilise regional-level wage information, they cannot explore additional heterogeneity in returns to skills.

To further motivate my study of gender differences in returns to skills in Denmark, a first step is to determine whether or not patterns in employment polarisation differ between women and men. To do so, I follow the approach by Acemogly and Autor (2011: Figure 10). Occupations are mapped into percentiles according to their mean hourly wage in 1991. Next, I examine if occupations in the 1st percentile of the wage distribution in 1991 account for more or less than 1% of workers in 2009, and then repeat for each of the remaining percentiles. This is done separately for women and men and repeated for the periods 1996-2009 and 2000-2009. Figure 1 is a smoothed plot of the results from this exercise.

Figure 1: Employment polarisation in Denmark



Source: AKM, BEF, IDAN 1991-2009. Note: Method adopted from Figure 10 in Acemoglu and Autor (2011). *Each occupation is mapped into percentiles according to their mean hourly wage in the base year. This is the x-axis. The kernel-smoothed share of employment, which the occupations from each base year-percentile account for in 2009, is mapped on the y-axis.*

It is clear from Figure 1 that polarisation patterns indeed are more pronounced for women than for men when considering the full timespan 1991-2009. This corresponds with the findings from the international literature outlined in the above. The literature on RBTC also emphasises the link between task-specific skills and employment polarisation. Thus, Denmark does provide an interesting setting to explore gender differences in skill demand and in skill prices. Not only does Danish labour market exhibit high levels of occupational segregation, women's and men's employment also seem to be differentially impacted by RBTC. Exploring gender differences in returns to skills provides a first step towards a better understanding of these gendered employment patterns. In order to examine the gendered aspects of skills, I utilise both job vacancy data and data from the rich population-level Danish registers. I describe these datasets in detail in the following section.

Data sources

As alluded to above, my analyses rely on two sources of data. Firstly, Statistics Denmark provide register data on employment, education, demographics, firm characteristics etc. Crucially, these registers include the entire population of both workers and firms in Denmark. Furthermore, it is possible to match the different registers at the firm- and individual-levels. Monthly employment data are available, and they include a firm identifier and an occupational code for each employment relation. Secondly, Danish online job vacancy data from 2010-2016 are supplied by the Danish consultancy firm, Højbjerre Brauer Schultz (HBS). These data also include a firm identifier and an occupational code for each job post as well as a posting date. Thus, it is possible to match data from the two sources using firm identifiers and occupational codes, and by exploiting the data's time dimension. In the following sections, I describe the register and job vacancy data in more detail separately, before I move on to describe the match between the data sources.

Register data

In this section, I briefly outline which data I extract from Statistics Denmark's registers. [3] Detailed monthly employment data for the entire Danish population is available from the BFL-register until 2016. Particularly, monthly wages, start and end dates, monthly hours, a firm identifier (CVR-number), and an occupational code are provided for each monthly observation. A person will appear in the register multiple times if they have more than one job in a given month, i.e. jobs are not aggregated at the individual-level, but are included as separate observations. In what follows, I define a job spell as the period over which a worker remains within the same firm*occupation cell. Thus, a new job spell starts when a worker enters a new role in the same firm (new occupation code), or when a worker gets a job in another firm (new firm identifier). From this definition, I construct two datasets:

[3] For data documentation, see: https://www.dst.dk/da/TilSalg/Forskningsservice/Data/Register_Variabeloversigter

1) **Stock data:**
   I include all jobs in BFL from 2010-2016 and aggregate monthly observations for each job to get averages of hourly wages etc. for each calendar year. I include the total number of hours per year which I normalise to a measure of full-time equivalents.[4] This measure of full-time equivalents will be used as weights in the analyses that follows.

2) **Flow data:**
   I identify new jobs in BFL, i.e. jobs where workers are registered with either a new occupational code or a new firm identifier in a given month.[5] I construct a sample of those new jobs with the first 12 months of observations in BFL (or fewer, if the job spell ends before). Next, I aggregate to get the 12-months-averages of hourly wages, full-time equivalents, and other relevant variables. Thus, this dataset contains all new jobs and information on the first 12 months of employment. Since I need 12 months of observations, the latest job spells included start in January 2016.

The stock data provide information about the full population of workers, whereas the flow data yield information only on workers during their first year of employment in a certain job. I make two versions of the stock and flow data: one using *3-digit* occupation codes (DISCO-codes), and one using *4-digit* DISCO-codes.

For both samples impose a number of restrictions and exclude observations of workers:

a) With a missing DISCO-code or firm identifier
b) With a total number of hours for a given year below the equivalent of a full-time month (1923.96/12 as defined by DST) or above 3,500 hours. Part-time workers remain included.
c) Aged under 15, or over 70 on 1 January in the given year
d) With an hourly wage below 30 DKK or above 5,000 DKK (in 2016-levels)
e) With total wages exceeding 10,000,000 DKK (in 2016-levels)

Criterion a) and b) are the most restrictive. Criterion a) is necessary to construct job spells at the firm*occupation level, missing DISCO-codes are mostly observed in the private sector. I describe the DISCO-coding in detail below. Criterion b) is imposed to avoid observations where hours of work may be misreported, e.g. freelance work. In addition, I believe that jobs spells with fewer hours than the equivalent of a full-time month are less likely to appear in the job vacancy data due to fixed costs of hiring. To complement the employment data from BFL, I extract data on demographics from BEF, years of education from UDDA, employment experience from IDAN, and firm data from FIRM. This completes my stock and flow datasets. In the next section, I briefly outline the

[4] A full-time job is defined as 1923.96 hours per year Statistics Denmark, so full-time equivalents = total number of hours per year / 1923.96.
https://www.dst.dk/da/Statistik/dokumentation/Times/moduldata-for-arbejdsmarked/fuldtid

[5] "New" in the sense that the worker was not observed in same firm*occupation cell in the month before. Furthermore, I detect gaps between spells of work in the same firm*occupation cell. If the gap between two spells is less than 6 months, I do not code reoccurring work in a firm*occupation cell as a new job, rather I code them as the same job.

Danish job vacancy data before moving on the match between the vacancy and register data.

Job vacancy data

HBS collects online job vacancy data from numerous Danish online jobs boards, and thus, they believe that their data contains the near universe of publicly accessible Danish online job posts.[6] They remove duplicates and clean the data before machine reading the job posts. By applying data driven methods, HBS extracts the date on which a given job vacancy was posted online, the identification number (CVR-number) of the posting firm, and a 6-digit DISCO-code. If the firm identifier is not listed directly in the job post, HBS imputes it from publicly accessible registers using the firm name listed in the job post. Importantly, HBS also extract keywords from the raw text in the job post. In many ways, the resulting data is similar to the US job vacancy data supplied by BG.

I restrict my dataset to include job posts with non-missing firm identifiers and occupational codes only. Furthermore, I exclude job posts with no extracted keywords. These observations are likely to represent job posts that just state the firm name and job title of the open position without any accompanying text.

Table 1: Skills categories and their corresponding keywords

Cognitive	problem solving, research, analytical, critical thinking, math, statistics, systematic
Social	communication, teamwork, collaboration, negotiation, presentation, social, extrovert, network, relations
Character	organised, detail-orientated, multi-tasking, time management, meeting deadlines, energetic, busy, engaged, overview, motivate, independent, professional, goal-oriented, ambitious, efficient, love of order, stabile, result-oriented, adaptable
Writing/language	writing, language, English, German, Swedish, Norwegian
Customer Service	customer, sales, client, patient
Management	management, supervisory, leadership, mentoring, staff, control, planning, implementing
Financial	budgeting, accounting, finance, cost, tender/bids
Computer (general)	computer, spreadsheets, common software, systems (e.g. Microsoft Excel, PowerPoint)

Note: Categories and corresponding keywords adapted from Deming and Kahn (2018).

[6] For more details, see: http://www.hbseconomics.dk/wp-content/uploads/2017/09/Eftersp%C3%B8rgslen-efter-sproglige-kompetencer.pdf

In order to extract skill requirements from the job vacancy data, I follow the method of Deming and Kahn (2018). They map a selection of keywords into skills categories. For example, the keyword "teamwork" is indicative of a job requiring social skills. My eight skill categories and their mapping to keywords can be found in Table 1. I construct dummy variables for the 8 skill categories. If a job post contains one or more keywords that is indicative of a certain skill, the dummy takes a value of 1 and 0 otherwise. Thus, all job posts are marked as either requiring or not requiring each of the 8 skills. The next step is to match the job vacancy data and the associated skill requirements with the register data outlined above.

Data match

As unique firm identifiers and occupational codes are included in both the register data and job vacancy data, the data can be match along those two dimensions. Furthermore, I exploit the data's time dimension. For the match on DISCO-codes to be reliable, the codes must be consistently coded across the register data and job vacancy data. Thus, I briefly outline how DISCO-codes are coded in the two data sources.

Although variables on wages and hours in BFL are automatically imputed from the Danish tax authorities' data, the *6-digit* DISCO-codes are not. As they require some "manual" coding, i.e. placing a worker in a category, they do not appear in the Danish tax authorities' data. Hence, DST collect the DISCO-codes in a separate procedure. For public employees, DST impute DISCO-codes directly from the public wage data where every employee's job title/position is recorded. In the private sector, DST collect data on employees from firms with 100 or more employees every year.[7] Smaller firms are sampled to report DISCO-codes on their employees from year to year. Private employers are supplied with a correspondence table between job titles/positions and DISCO-codes in order to secure consistent reporting.[8] If a private firm is not sampled, DST impute an individual's DISCO-code from the previous year given that changes no in the individual's employment are observed. Otherwise, they estimate a DISCO-code from register data on each individual's education, the industry of the individual's employer, and the individual's membership of an unemployment insurance fund (these funds are often occupation-specific).[9]

In the case of the job vacancy data, HBS first extract a job title from each job post. Using a correspondence table between job titles/positions and DISCO-codes similar to that supplied by DST to DISCO-reporting firms, HBS can then identify the *6-digit* DISCO-code which corresponds to the extracted job title.[10] Thus, both the register data's and the job vacancy data's *6-digit* DISCO-codes are imputed from detailed job titles/positions.

[7] For more details, see: https://www.dst.dk/ext/loen/vejl_lon_2017--pdf
[8] For more details, see: https://www.dst.dk/da/Indberet/oplysningssider/loenstatistik/stillingsbetegnelser-disco-08-i-loenstatistikken
[9] For more details, see: https://www.dst.dk/Site/Dst/SingleFiles/hojkvalbilag.aspx?varid=107187&bilagid=183191
[10] For more details, see: http://www.hbseconomics.dk/wp-content/uploads/2017/09/Eftersp%C3%B8rgslen-efter-sproglige-kompetencer.pdf

Although DISCO-codes are generally imputed in a similar manner in both the register data and the job vacancy data, some inconsistencies are to be expected at the very detailed *6-digit* level. For example, there are three subdivisions of nurses at the *6-digit* level and only one at the *4-digit* level.[11] A job title may not capture the detailed *6-digit* division. In order to avoid coding inconsistencies, I perform the following matching procedure at two levels: one using *3-digit* DISCO-codes (174 unique values), and one using *4-digit* DISCO-codes (575 unique values). The *4-digit* version is mainly included as a robustness check, so statistics and results on this version are reported in the appendix throughout. The register and job vacancy data are matched as follows:

1) Stock data
   In order to merge skill levels from the job vacancy data onto the individual-level BFL stock data, I take the following approach: First, I calculate the average skill levels for each firm*occupation cell. I do this by averaging required skills across all job posts from 2010-2016 in that particular firm*occupation cell. Next, I merge these skills levels onto the BFL stock data at the firm*occupation level. Hence, every person in the same firm*occupation cell will have the same skill levels. This goes for both new hires and existing employees. I ignore any time variation in skill requirements, but I do observe within-occupation variation skills, namely across firms.

   The average skill requirements can fall anywhere in the interval 0 to 1, and thus, the average skill requirements are not dummy variables. I take this particular approach to obtain individual-level data with skill measures comparable with those of Deming and Kahn (2018). Match rates and number of jobs and vacancies are reported for the dataset using *3-digit* DISCO-codes in Table 2 (see appendix, Table 13, for results on *4-digit* data). Match rates are consistently high, and it is evident that the number of job posts was lower in the post-recession years but have recently increased.

2) Flow data
   I construct flow data on skills and wages to further exploit time and individual variation in skill requirements and returns. In order to do so, I assume that most vacancies are posted in same month as the vacancy is filled or maximum four months prior. For example, if a job spell starts in May, the corresponding vacancy would be posted anytime between the beginning of January to the end of May in the same year. With this assumption, I use the job vacancy data to construct a rolling average of skill levels in each firm*occupation cell. If a new job spell appears in the BFL flow data, I match it with skill levels averaged over the relevant 5 months. Because 4 months of job vacancy data before job start is needed, my matched flow data is limited jobs commencing in the period May 2010 to January 2016. The flow-matching strategy gives a pseudo-individual level match with much less aggregation compared to the stock data. Again, the average skill requirements can fall anywhere in the interval 0 to 1; they are not dummy variables. Table 3 shows

[11]For more details, see: https://www.dst.dk/da/Statistik/dokumentation/nomenklaturer/disco-08

match rates aggregated to the yearly level for the dataset using *3-digit* DISCO-codes (see appendix Table 14, for results on *4-digit* data). Match rates are – expectedly – lower than those for the stock data both in terms of BFL-jobs and job posts. My definition of a new job is not very restrictive, so it is not surprising that only 27% of new BFL-jobs can be matched with a job post. Many of the new jobs are likely to be internal hires, informal hires (the job is not publicly posted), or DISCO-recodes. However, 44 % of job posts are matched to new BFL-jobs. This is a very high match rate when compared to Kettemann, Mueller, and Zweimüller (2018). At the *4-digit-level,* the number of jobs is naturally slightly higher, and the match rates slightly lower (see Table 13 and Table 14 in the appendix).

Table 2: 3-digit stock data: Match rates

Year	Number of jobs	Matched jobs	% of jobs matched	Number of job posts	Matched posts	% of posts matched
2010	2,776,911	1,470,265	0.53	91,184	58,056	0.64
2011	2,800,660	1,503,113	0.54	96,367	60,440	0.63
2012	2,738,233	1,492,915	0.55	108,950	71,235	0.65
2013	2,733,638	1,520,014	0.56	116,788	79,690	0.68
2014	2,797,925	1,564,600	0.56	124,802	84,505	0.68
2015	2,846,993	1,580,575	0.56	136,320	93,109	0.68
2016	2,889,036	1,577,849	0.55	140,080	92,314	0.66
Total	19,583,396	10,709,331	0.55	814,491	539,349	0.66

Source: BFL, HBS-Jobindex 2010-2016, excluding observation with missing CVR- or DISCO-codes.

Table 3: 3-digit flow data; Match rates

Year	Number of new jobs	Matched new jobs	% of jobs matched	Number of job posts	Matched posts	% of posts matched
2011	839,275	206,695	0.25	96,367	38,108	0.40
2012	769,250	205,562	0.27	108,950	47,173	0.43
2013	779,259	224,086	0.29	116,788	53,705	0.46
2014	885,229	236,647	0.27	124,802	56,673	0.45
2015	843,607	224,930	0.27	136,320	62,817	0.46
Total	4,116,620	1,097,920	0.27	583,227	258,476	0.44

Source: BFL, HBS-Jobindex 2010-2016, excluding observation with missing CVR- or DISCO-codes.

The main advantage of the stock data is that skill levels are imputed for *all* employees: both new hires and existing employees. This is similar to the approach taken by Deming and Kahn (2018), although they only have wage data at the MSA-level and not at the individual-level. However, the imputation relies on a rather strict assumption: the skills requirements of already employed workers equal those of new hires in the same firm*occupation cell. In other words, it is assumed that skill requirements in a job post reflect the skills of workers already employed in the same firm*occupation cell. This assumption also implies that skills cannot be learned at the job, but rather stay constant.

The flow data allow me to relax the assumption of time-invariant skill levels at the firm*occupation level. Instead, it is only assumed that new employees' skills levels are reflected in the job posts in their firm*occupation cell just around the start of their job spell. Furthermore, focussing on the first 12 months of wages in a job spell should limit bias from additional skills learned in the firm*occupation cell. Since only few workers tend to start in the same firm*occupation cell in a given month, the level of aggregation is also much lower: for the *3-digit* data, the median number of workers in matched firm*occupation*start-month cells in the flow data is 2, whereas the equivalent median number of observations in firm*occupation cells in the stock data is 11.[12]

Aggregating the job vacancy/skills data at the firm*occupation and firm*occupation*start-month levels is a potential drawback of my data: I cannot separate women and men in the job vacancy data, and thus, I assume that everyone has the same skills at the firm*occupation and firm*occupation*start-month levels respectively. In other words, I assign women and men in the same cells with the same skills; I do not observe any gender variation in skills at the firm*occupation/firm*occupation*start-month levels. If women and men tend to work in the same cells, this would restrict my analysis. However, as pointed out above, women and men tend to work in different occupations in the Danish labour market, i.e. high levels of occupational segregation are observed (Jarman et al. 2012). Due to the smaller cell sizes, gender segregation is likely to be even more pronounced at the firm*occupation and firm*occupation*start-month levels. To explore gender segregation at these levels, I first calculate the female share of hours in each firm*occupation/ firm*occupation*start-month cell. Next, I graph the kernel density estimate of hours worked by women and men respectively on the cell's female share of hours. To make my point even clearer, I also graph the cumulative distribution of hours worked for women and men respectively on the cell's female share of hours.

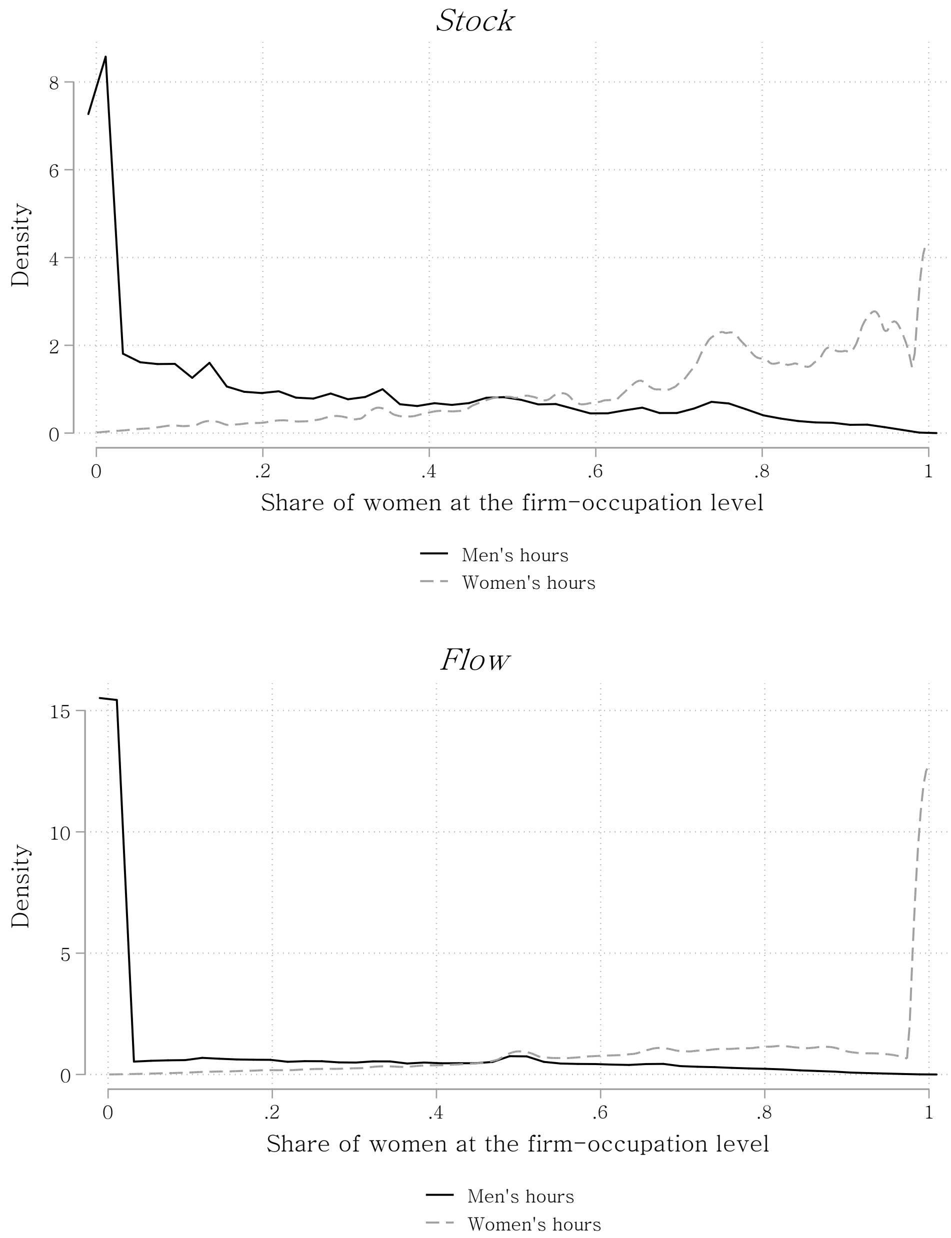
Figure 2 and Figure 3 show that women and men do not tend to mix in the same firm*occupation, and even more seldom, women and men get employed at the same time in the same firm*occupation*start-month cell. So, despite the fact that I cannot observe any gender variation in skills within firm*occupation/firm*occupation*start-month cells, I still observe plenty of gender variation in skills across these cells. Furthermore, I do observe gender differences in wages and in all other characteristics within a cell; these variables vary at the individual level.

An average match rate of 55 % of BFL-jobs in the 3-digit stock data, and 27 % of new BFL-jobs in the 3-digit flow data can be problematic if the matched jobs are not representative of the population of the stock job or new jobs respectively. To check whether or not all occupations and industries are represented in the matched data, I compare the occupational and industrial distribution in the complete BFL data and in the matched subsample. List of occupations and industries and their titles are included in the appendix, Table 15 and Table 16.

---

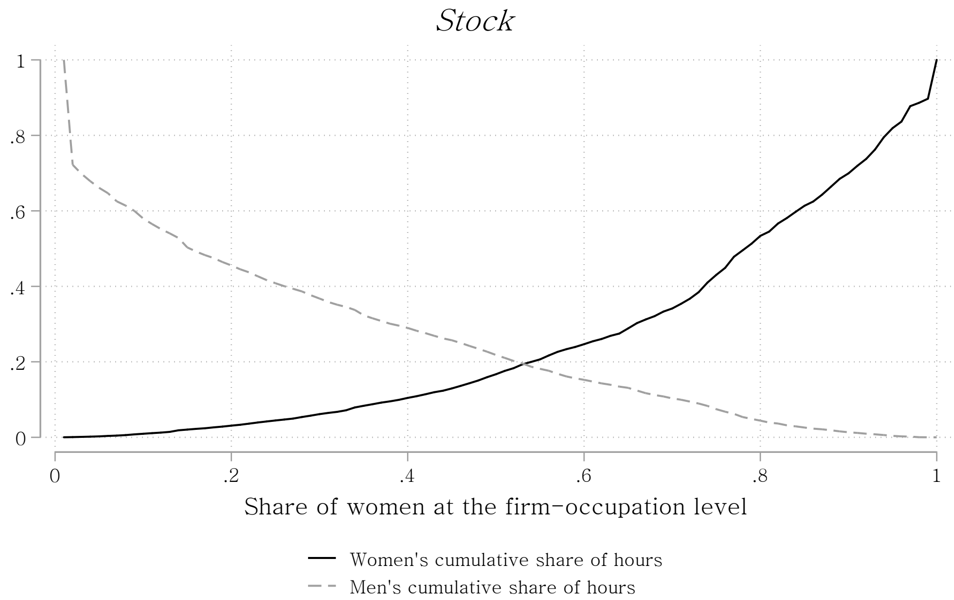
[12] Median number of workers in 4-digit flow firm*occupation*start-month cells is also 2, and the median number of observations in 4-digit stock firm*occupation cells is 10.

Figure 2: 3-digit data: Density of hours worked



*Source: BFL 2010-2016, excluding observation with missing CVR- or DISCO-codes. Kernel density estimates of hours worked on the share of women in the workers' firm*occupation cell.*

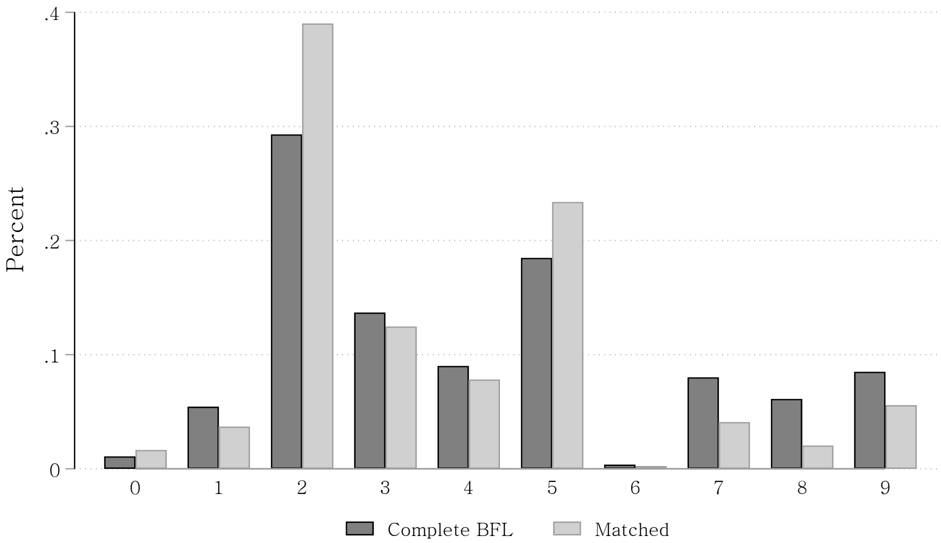
Figure 3: 3-digit data: Cumulative distribution of hours worked



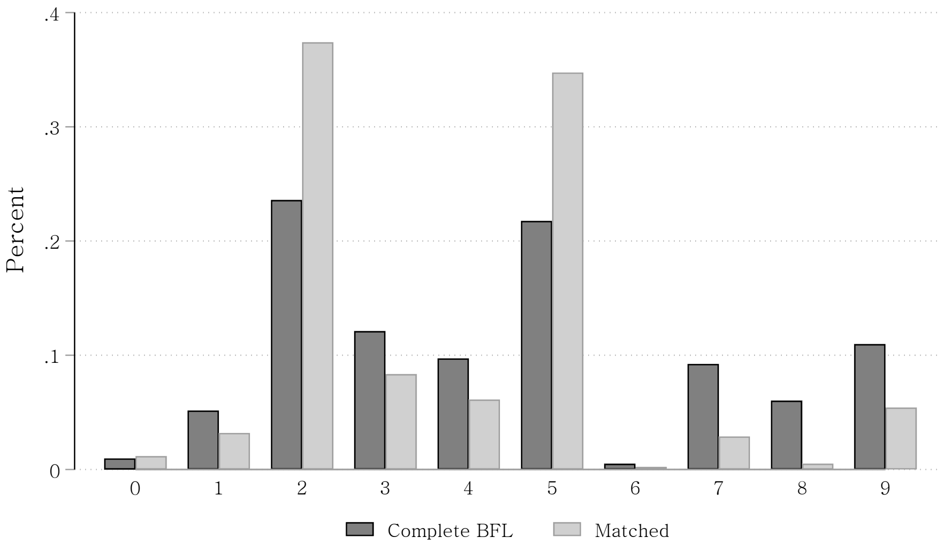
*Source: BFL 2010-2016, excluding observation with missing CVR- or DISCO-codes. Cumulative distribution of hours worked on the share of women in the workers' firm*occupation cell.*

Figure 4: 3-digit data: Distributions of occupations in BFL and matched data

Stock

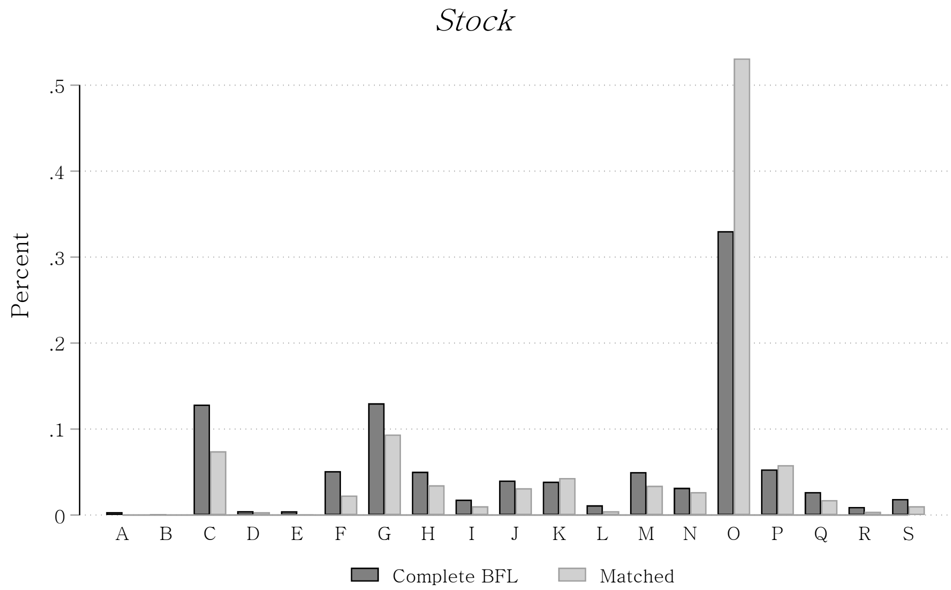


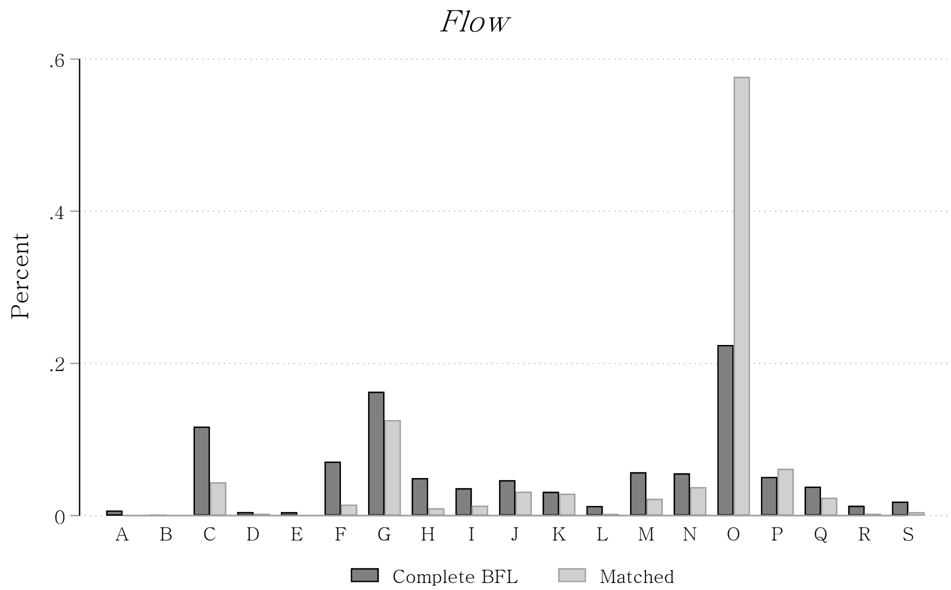
Flow



Source: BFL, HBS-Jobindex 2010-2016.
Note: Observations weighted by full-time equivalents.

Figure 5: 3-digit data: Distributions of industries in BFL and matched data





Source: BFL, HBS-Jobindex 2010-2016.
Note: Observations weighted by full-time equivalents.

Notice the striking overrepresentation of the occupations "2 Professionals" and "5 Services and Sales Workers" as well as of the industries "O Public administration and defence; compulsory social security" and "P Education". These occupations are dominated by large groups of public employees, namely teachers, nurses and care assistants (the Danish public employers are in generally included in industry "O", even if they undertake work in the health care sector). The significant overrepresentation of public employees in the matched sample follows from the fact that all permanent public sector jobs by law must be publicly advertised. Thus, public sector job vacancies are also overrepresented in the vacancy data. Notice, however, how all occupations and industries are represented in the matched data, despite the overrepresentation of public sector jobs. I include a variable in the matched data to indicate whether a job is in the public or private sector, which is utilised in as a control in the data analyses that follow below.

To conclude, match rates between Danish register data and job vacancy data are relatively high. This holds even after relaxing the strict assumption that skills requirements of already employed workers equal those of new hires in the same firm*occupation cell, i.e. when focussing on new jobs in the flow data. Furthermore, jobs in all 1-digit occupations and 1-letter industries are well-represented, although public sector jobs are overrepresented. Keeping this in mind, the next section provides some summary statistics on the skill levels of workers in the Danish labour market.

Descriptive statistics

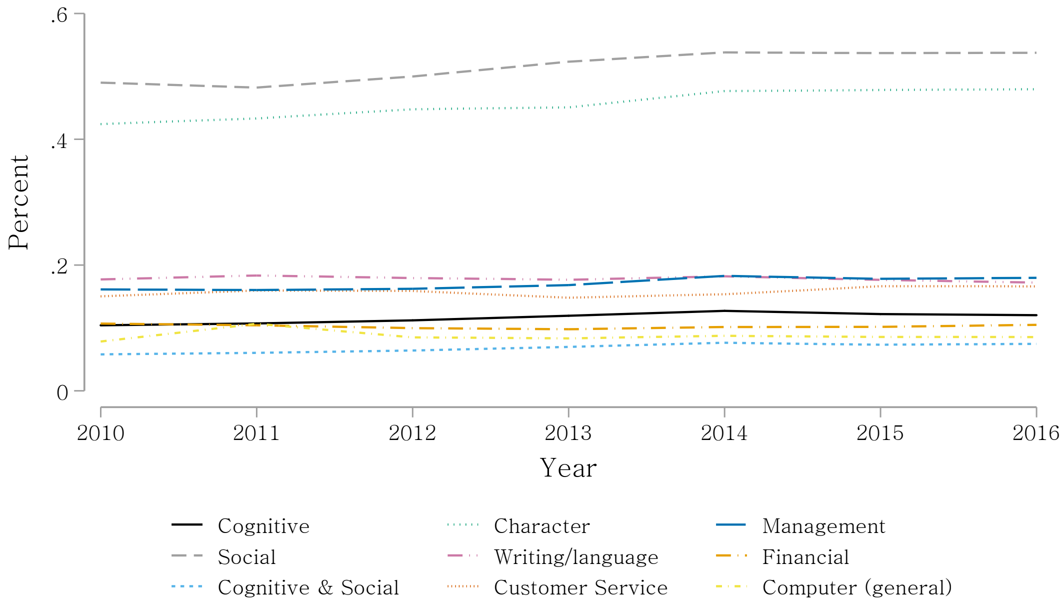
Before moving on to regression analyses of gender differences in returns to skills in the next chapter, some introductory descriptive statistics are provided here. First, I show time trends in skills levels, and I show how these trends are relatively well accounted for by changes in the occupational distribution. Throughout, the demand for both cognitive and social skills is also included, as Deming and Kahn (2018) find that this skill combination is of particular importance. Second, I exploit the match between the job vacancy data and register data when reporting gender differences in skills levels. Third, it is highlighted how the skill measures correlate with each other and with other variables such as wage and years of education. Lastly, I show that not all variation in skill measures can be accounted for by individual-level variation in other variables. Thus, I demonstrate that the skill measures yield explanatory power beyond that of standard labour market data (cf. Deming & Kahn 2018).

Skill levels and trends

Levels of skill demands are plotted from 2010-2016 in Figure 6. Notice that social skills are by far the most demanded type of skill; around half of all job posts state a requirement for social skills. Furthermore, the demand for social skills increases over time. The second most demand skill is “character” which is required in more than 40 % of job posts. All other skills are demanded in between 10 and 20 % of all job posts. Only the requirement of both social and cognitive skills is less common.

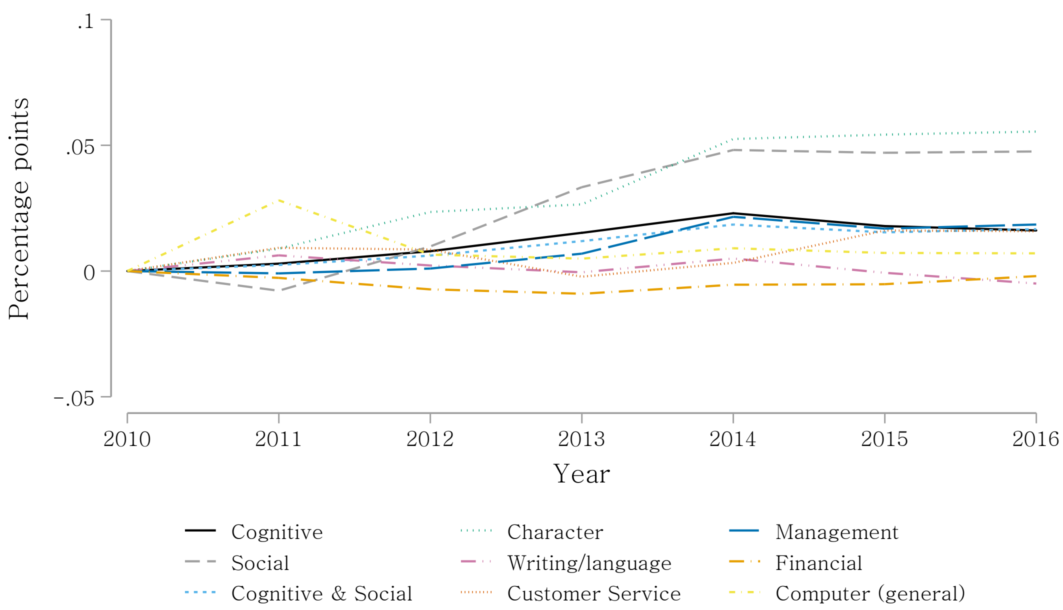
In order to further explore time trends in skills, I regress the demand for each skill on a set of year dummies. 2010 is the reference group. The coefficients on the year dummies are plotted in Figure 7. I plot 2010 as zero to indicate that this year is the reference group. Thus, growth in skill demands are presented in percentage points, not percent. Throughout the demand for all skills tend to be slightly increasing, except for demands for “Writing/language” and “Financial” skills which are marginally decreasing.

Figure 6: Skill demand 2010-2016



Source: HBS-Jobindex 2010-2016.

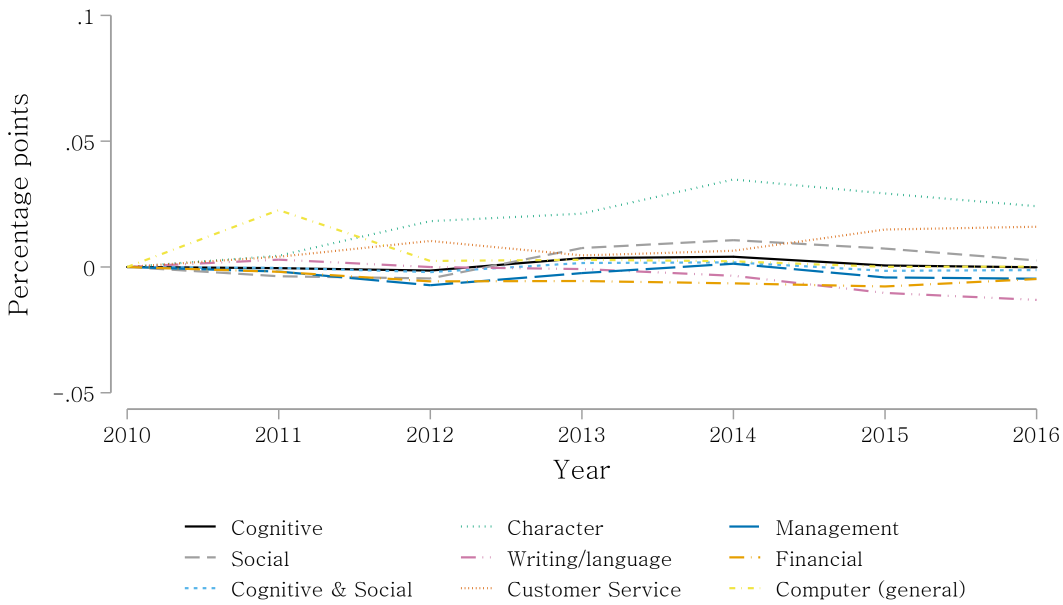
Figure 7: Unadjusted growth in skill demand 2010-2016



Source: HBS-Jobindex 2010-2016.
Note: Skills are regressed on year dummies. 2010 is the reference group.

Time trends in skills constitute one of Deming and Kahn's (2018) arguments for ignoring the time dimension in their skill measures. Instead, they aggregate over their sample period 2010-2015: "This aggregation has the advantage of smoothing out time variation in skill requirements driven by factors outside the scope of this analysis, such as labor supply shocks or labor market conditions." (Deming & Kahn 2018:344). I follow their method when constructing my stock data, but the flow dataset keeps the time dimension. Thus, it is necessary to show that time trends in skills will not bias my results from the flow data. Therefore, instead of regressing skill demands on just year dummies, I now also include occupation fixed-effects and number of keywords observed for each job post. The corresponding coefficients on year dummies are plotted in Figure 8.

Figure 8: Adjusted growth in skill demand 2010-2016



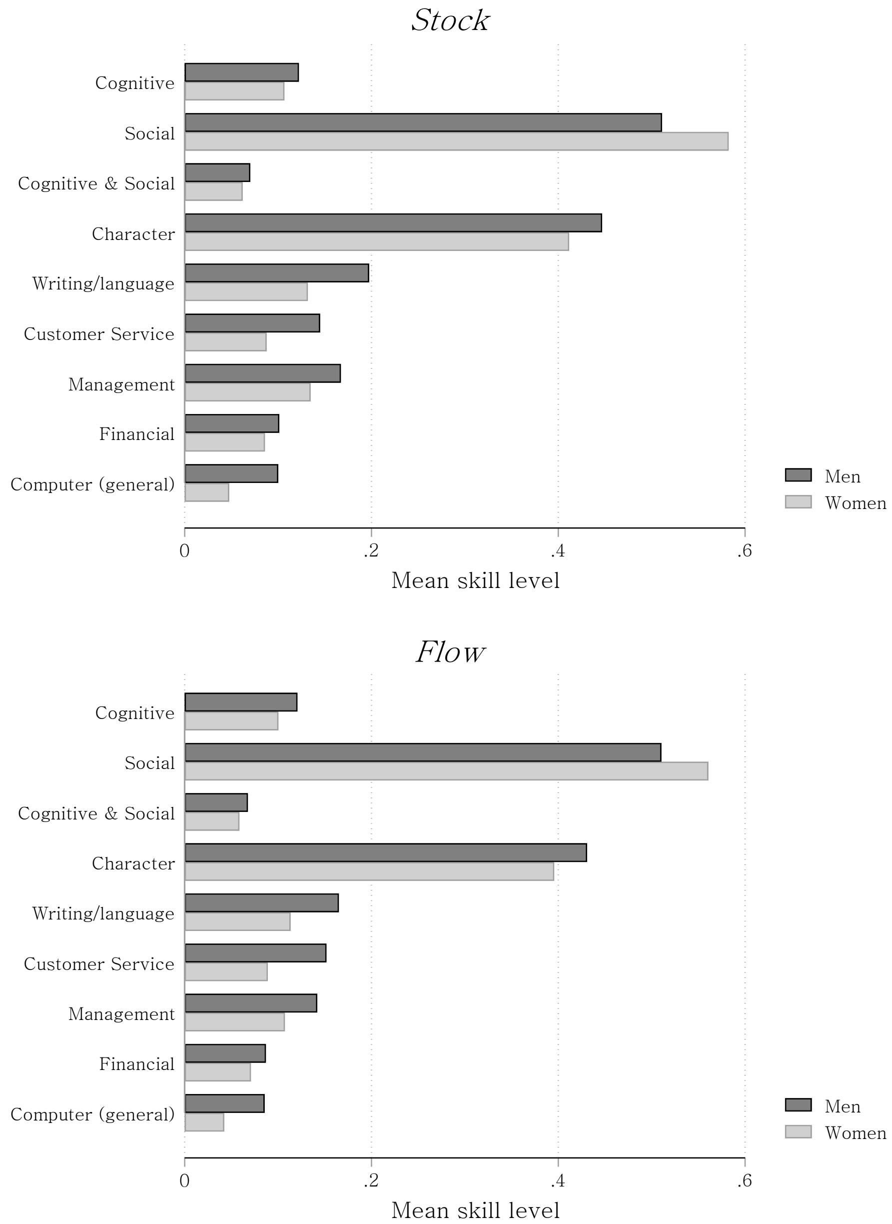
Source: HBS-Jobindex 2010-2016. Note: Skills are regressed on year dummies, number of keywords, and 4-digit occupations dummies. 2010 is the reference group.

Most of the growth in skill demands can be accounted for by a changing occupational distribution and by an increasing number of keywords. The demand for "Character" skills is the most increasing from 2010-2016 in terms of percentage points. However, this is also one of the most frequently observed skills, so in percent, the increase is small. Including controls for occupations and number of keywords should therefore limit bias from time trends in skills when analysing the flow data.

The reported time trends and levels in skills are determined solely by analysing job vacancy data. In what follows, I finally exploit the match between the Danish job vacancy data and register data.

Gender differences in skill levels
The vacancy-register data match enables analyses of skills together with the rich sets of variables provided by the Danish registers. In the context of this thesis, an essential piece of information to exploit is – of course – the gender of workers. Figure 9 maps average skill levels for women and men respectively.

Figure 9: 3-digit data: Mean skill levels by gender



Source: BEF, HBS-Jobindex 2010-2016.
Note: Observations weighted by full-time equivalents

Figure 9 shows that women are overrepresented in jobs that require social skills when compared to men. The opposite is the case for the remaining seven skills. Distributions of skills are – as expected – similar in the stock and flow samples. Generally, skills levels are similar for women and men. The largest relative gender difference observed is in general computer skills. Men are about twice as likely to be employed in a firm*occupation/firm*occupation*start-month cell where computer skills are required. To further explore the relationship between skills, gender and wages, the next section reports correlations between all of these variables.

Correlations

Before moving on to regression analysis, simple correlation coefficients between the skill measures, wages, and gender are important to consider for at least two reasons. Firstly, the skill measures should not be too highly correlated, as that could result in multicollinearity issues in regressions. Second, the correlations themselves may give us some idea of whether or not the skill measures make sense to include in wage regressions. For example, one would expect that skilled workers tend to work in cells with higher wages, i.e. that skills measures and wages are positively correlated.

Table 4 includes correlations between log hourly wages, a female dummy variable (=1 for women), and finally, all eight skill measures as well as the requirement of both cognitive and social skills. I also include the dual requirement of both cognitive and social skills. All skill measures are positively correlated with wages. As apparent in Figure 9, all skills measures except social skills are negatively correlated with being a woman. Most skills are positively correlated with each other, although there are a couple of exceptions: "Character" and "Customer Service" are negatively correlated with a few skills, including social and cognitive skills. All signs of the correlation coefficients are the same for stock and flow datasets with the exception of the correlation between "Cognitive & Social" and "Character" skills. Importantly, no skill measures are correlated to a degree that should cause problems of multicollinearity in regression models, although the dual requirement of "Cognitive & Social" by definition is highly correlated with cognitive skills. Note that this is not a standard interaction term. Instead, it represents the share of job posts within the worker's firm*occupation/firm*occupation*start-month cell, which requires both cognitive and social skills.

Table 4: Correlation tables of skills, wages and gender

3-digit stock data

	ln(wage)	Female	Cognitive	Social	Cognitive & Social	Character	Writing/language	Customer Service	Manage-ment	Financial	Computer (general)
ln(wage)	1										
Female	−0.218	1									
Cognitive	0.250	−0.0400	1								
Social	0.0885	0.119	0.0544	1							
Cognitive & Social	0.217	−0.0300	0.793	0.238	1						
Character	0.0538	−0.0596	−0.0223	−0.0605	0.0118	1					
Writing/language	0.186	−0.131	0.171	0.0423	0.169	0.0785	1				
Customer Service	0.0349	−0.120	−0.0297	−0.0740	−0.00843	0.0880	0.0625	1			
Management	0.214	−0.0726	0.120	0.191	0.144	0.0559	0.0692	−0.0463	1		
Financial	0.159	−0.0384	0.109	0.0247	0.102	0.0456	0.105	0.0648	0.128	1	
Computer (general)	0.126	−0.156	0.152	−0.0541	0.111	0.0313	0.191	0.0381	0.0701	0.0540	1

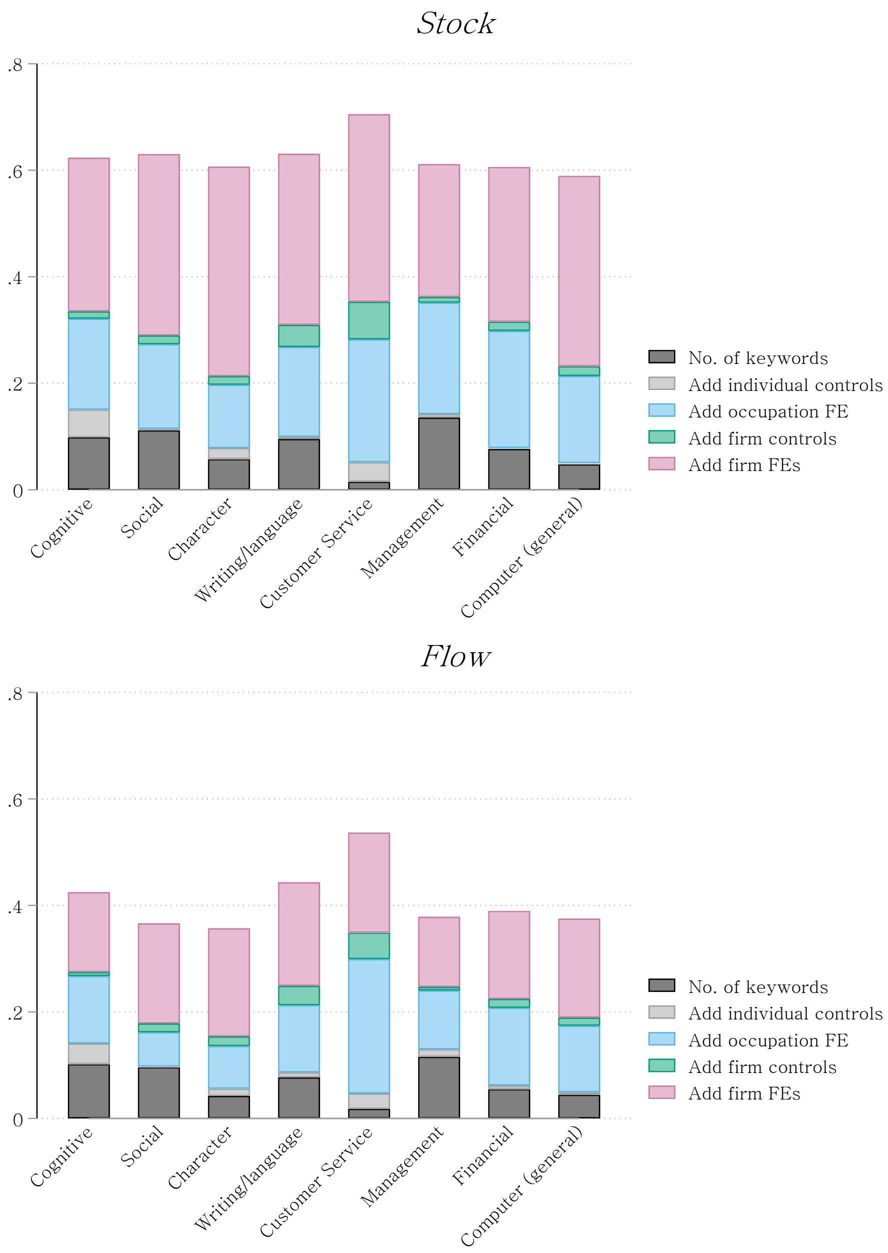
3-digit flow data

	ln(wage)	Female	Cognitive	Social	Cognitive & Social	Character	Writing/language	Customer Service	Manage-ment	Financial	Computer (general)
ln(wage)	1										
Female	−0.158	1									
Cognitive	0.218	−0.0433	1								
Social	0.0451	0.0686	0.0559	1							
Cognitive & Social	0.174	−0.0257	0.773	0.254	1						
Character	0.0307	−0.0479	−0.0493	−0.0439	−0.0149	1					
Writing/language	0.144	−0.0938	0.174	0.0371	0.150	0.0358	1				
Customer Service	0.00600	−0.115	−0.0474	−0.00359	−0.0251	0.0865	0.000604	1			
Management	0.195	−0.0705	0.146	0.0714	0.140	0.00787	0.0869	−0.0201	1		
Financial	0.156	−0.0379	0.134	0.0236	0.127	0.00425	0.0672	0.0426	0.124	1	
Computer (general)	0.126	−0.111	0.168	−0.00239	0.118	0.0257	0.176	0.0355	0.0706	0.0697	1

Note: Observations weighted by full-time equivalents.

Although the correlation coefficients indicate that my skill measures are not correlated to a degree that would cause multicollinearity issues in regression models, the variance of the skill measures should also be explored. Before moving on to regression analyses it must be established that skill requirements cannot be entirely predicted by potential covariates. If so, the skill measures would not add any explanatory to a regression model. Thus, I regress the eight skill measures on various sets of control variables, and plot the adjusted R^2 from each regression:

Figure 10: 3-digit data: Adjusted R^2 from regressions of skills level on various controls



Source: Various registers, HBS-Jobindex 2010-2016.
Note: Skills are regressed on various sets of controls.
The sets of controls are described in the section on regression models.

Figure 10 shows that around 60 % of the variance in skills in the stock data can be explained by the most extensive set of covariates. For the flow sample, 40 % of the variance in skills can be explained. Notice that occupation and firm fixed-effects explain particularly large fractions of the variance in skill requirements. Still, a significant share of the variance in skill demands cannot be explained by even the most extensive set of covariates. Thus, the skill measures appear as suitable regressors in regressions in which similar sets of covariates are included. The next section introduces the regression analyses which aim to shed light on potential gender differences in returns to skills.

Regression analyses

Undertaking regression analyses is the next step to further understand gender differences in returns to skills. First, I outline the regression models, and next results are reported. Finally, a few robustness checks are included.

Although the regression models outlined below do control for some confounding factors, the estimated returns to skills should interpreted as correlational rather than causal evidence. In the section on limitations, I propose future strategies that may identify potential causal effects of skills.

Models

Regressing hourly wages on skills and their gender interactions will indicate whether women and men with same skill requirements also receive the same wage. The econometric methods I rely on are simple but suitable for the question at hand. To give a full picture the interactions between gender and skills, I estimate four models:

A. A simple gender pay gap estimation:
   I regress log hourly wages on a female dummy and controls
B. An estimation of returns to skills without a female dummy:
   I regress log hourly wages on skills and controls *without a female dummy*
C. An estimation of returns to skills with a female dummy:
   I regress log hourly wages on skills and controls *with a female dummy*
D. An estimation of returns to skills with a female dummy and female*skills interactions:
   I regress log hourly wages on skills and controls *with a female dummy and female*skills interactions.*

Each of the four models are estimated with four different sets of controls and fixed effects. Before writing out the relevant regression models, I outline the four sets of control variables and fixed effects that are applied. The order reveals the sequence of which they will be introduced in the regressions that follow:

1. Base controls (individual controls and number of keywords): *age, years of experience, migrant dummy, marriage dummy, part-time dummy, year FEs, (only for flow data: start-month FEs), number of keywords*
2. Occupation fixed effects: *see Table 5.*

3. Firm controls: *1-letter industry dummies (19 in total), firm location, number of employees, a private dummy*
4. Firm fixed effects: *see Table 5.*

Table 5: Fixed effects

	3-digit		4-digit	
	Stock	Flow	Stock	Flow
Occupation FEs	136	124	389	325
Firm FEs	23,428	13,990	21,544	12,532

The advantages and disadvantages of including the different sets of controls are discussed in the results section. Controls are included additively in the regression models, so it suffices to write out the full model including all controls and fixed effects. I do so for both the stock and flow dataset. For the stock datasets, the following linear model is estimated:

$$w_{iofy} = \beta_0 + I_{iy}\beta_1 + F_{fy}\beta_2 + S_{of}\gamma_1 + S_{of} \times g_i \times \gamma_2 + \lambda_o + \phi_f + \theta_y + \varepsilon_{iofy}$$

Where i indicates variation at the individual-level, f at the firm-level, o at the occupation-level, and y at the year-level. w_{iofy} is the log hourly wage. I_{iy} is a matrix of individual year-varying characteristics and includes a female dummy, F_{fy} a matrix of firm year-varying characteristics, S_{of} is a matrix of the eight skill measures that vary at the firm*occupation-level. g_i is a female dummy variable, which equals 1 for women only. λ_o are the occupation FEs, ϕ_f the firm FEs. Finally, θ_y are year FEs. The vector γ_1 gives the coefficients on the eight skill measures for men and $\gamma_1 + \gamma_2$ for women, i.e. γ_2 is the gender differences in the skill measures' coefficients.

For the flow data, where the skill measures also have a time dimension, the linear regression model can be written as follows:

$$w_{iofym} = \beta_0 + I_{iy}\beta_1 + F_{fy}\beta_2 + S_{ofym}\gamma_1 + S_{ofym} \times g_i \times \gamma_2 + \lambda_o + \phi_f + \theta_y + \delta_m + \varepsilon_{iofym}$$

The notation is analogue to that of the previous model but note the additional subscript m, which indicates further variation at the *start*-month-level. Wages are still aggregated to 12-months averages, but a job spell can start in any given month, as indicated by the added time dimension. Furthermore, start-month FEs are added, δ_m. Note they are *not* interacted with year FEs, and thus, they are not year*start-month FEs.

As the eight skill measures only vary at the firm*occupation/firm*occupation*start-month levels, the error terms ε_{iofy} and ε_{iofym} may be correlated within these cells. Thus, I follow the approach taken by Hersch (1998) and cluster my standard errors at these levels. Such an approach is also recommended by Cameron and Miller (2015), who note that cluster-robust standard errors increase with:

(1) the within-cluster correlation of the regressor
(2) the within-cluster correlation of the error
(3) the number of observations in each cluster.
(Cameron & Miller 2015:322)

The eight skill measures are perfectly correlated within clusters (they do not vary within), and the cells/clusters are relatively large (particularly for the stock data), and thus, applying cluster-robust standard errors significantly inflate the estimated errors. Using White-Huber (=robust) standard errors, all estimated coefficients on the skill measures and their gender interaction are highly statistically significant. This is not the case when applying cluster-robust standard errors, which stresses the importance of doing so. In the next section, the estimated coefficients and their cluster-robust standard errors are presented.

Results

The estimated coefficients are reported in Table 6 and Table 7. Note that military personnel with the 1-digit DISCO-code "0" are excluded from all regressions. Table 6 include the results from the 3-digit stock data, and Table 7 the results from the 3-digit flow data. Results from the 4-digit equivalents are reported in the appendix, in Table 17 and in Table 18 respectively.

First, consider models 1A-1D in both Table 6 and Table 7. Model 1A gives an estimate of the gender pay gap adjusted only for individual characteristics. Note that the gender pay gap is larger in the stock than in the flow data. One explanation of this difference could be that the gender pay gap increases with firm*occupation-tenure. Model 1B shows that all coefficients on all skills are highly significant and positive, except for social skills for which the coefficient is negative. From model 1C, it becomes clear that model 1B's negative coefficient on social skills is partly driven by a gender effect, as introducing a female dummy in the regression reduces both the level and significance of the negative coefficient on social skills, but only in the stock data.

Model 1D introduces female*skill interactions. For both the stock and flow data, note that coefficients on all skills, but social skills, remain positive and significant. The coefficients' magnitudes increase, except for cognitive and social skills. Next, the coefficients on the female*skill interactions are of particular interest. Both in the stock and flow data, the coefficient on the female*cognitive interaction is positive, but it is only significant in the flow data. The remaining coefficients on female*skill interactions are either insignificant, or significant and negative. In the stock data, only the coefficients on female*character and female*(customer service) are negative and significant, but this also goes for coefficients on female*management, female*financial, and female*computer in the flow data. This is the first indication that women and men with the same skills may face different returns to them. Of course, many other factors influence this result; the next models try to control for a number of these. Finally, note that the unexplained gender pay gap, i.e. the coefficient on the female dummy, deflates already when introducing skills in the regression (model 1C). The pay gap further decreases after including female*skills interactions.

In model 2A-2D, occupation FEs are added to the regression. First, the differences in the estimated gender pay gap between model 1A and 2A – occupational segregation explains a relatively large share of the gender pay gap. Pay differences due to occupational segregation may be an outcome of interest, and thus, occupation FEs can be seen as "bad controls" in an estimation of the gender pay gap (Angrist & Pischke 2008). However, this is not the outcome of interest here, rather model 2A-2D takes out occupation FEs to give *within*-occupation estimates of returns to skills. Note that the coefficients' signs generally stay the same, but their levels decrease. This makes sense: some occupations may require higher skill levels than others and pay their workers accordingly more. This cross-occupation effect is controlled for here. The coefficients on cognitive skills and on their female-interaction are small and insignificant across models 2B-2D. I return to this later, connecting it with "The Great Reversal in the Demand for Skill and Cognitive Tasks." (Beaudry et al. 2016). Also note how the negative coefficient on social skills appear to be driven in part by a gender effect, as the magnitude and significance level falls when introducing the female*social interaction. The coefficient on management skills is now insignificant in the stock data, but it remains significant in the flow data. The coefficients on the female*computer interaction is now positive, but also insignificant in model 2D in the flow data. As with model 1D, note the smaller magnitude of the gender pay gap after introducing the female*skills interactions in model 2D

Firms may play a significant role when estimating returns to skills. Some firms may reward certain skills highly, whereas other firms will not. To control for such effects, model 3 includes a number of firm specific controls, such as firm size, 1-letter industry dummies and a private/public dummy. This is particularly important as public sector jobs are overrepresented in the matched vacancy-register data. First, going from model 2A to model 3A further decreases the unexplained gender pay gap (the coefficient on the female dummy). Even after controlling for occupation FEs, firms still play a role in gender pay differences. Across models 3B-3D and for both stock and flow data, the coefficients on the eight skill measures fall in magnitudes, and several become insignificant. However, this is not the case for the coefficients on the female*skills interactions. Except for the female interactions with cognitive and social skills (and female*management in the stock data, and female*computer in flow data), all female*skills interactions are now negative and significant in model 3D. Furthermore, in model 3D, the inclusion of female*skills interactions reduces the unexplained gender pay gap.

Model 4 fully controls for firm specific effects by additionally including firm FEs in the regressions. Note that including both occupation and firm FEs is not analogue to including firm*occupation FEs, and thus, variation specific to the firm*occupation interactions remains. The estimated coefficients can be interpreted as *within*-firm*occupation returns to skills and *within*-firm*occupation gender differences in returns. Firm FEs does not change the estimation of the gender pay gap much – there is little difference between the coefficient on the female dummy in models 3A and 4A. Generally, the results for model 4 are similar to those of model 3. The key result is that the female interaction with character, writing/language, customer service, management, financial, and computer skills are consistently negative. Furthermore, all these coefficients are statistically significant, except the coefficient on female*financial in the stock data. Lastly, note that including female∗skills interactions cause a particularly large drop in the unexplained gender pay gap in the stock data.

Table 6: 3-digit stock data: Regressions results

	(1) Model 1A		(2) Model 1B		(3) Model 1C		(4) Model 1D	
Female=1	−0.158***	[0.005]			−0.134***	[0.005]	−0.101***	[0.010]
Cognitive			0.118***	[0.018]	0.116***	[0.019]	0.101***	[0.022]
Social			−0.042***	[0.012]	−0.015	[0.012]	−0.012	[0.015]
Character			0.044***	[0.012]	0.041***	[0.012]	0.061***	[0.014]
Writing/language			0.102***	[0.016]	0.079***	[0.015]	0.080***	[0.019]
Customer Service			0.112***	[0.017]	0.083***	[0.016]	0.103***	[0.019]
Management			0.128***	[0.018]	0.110***	[0.018]	0.123***	[0.021]
Financial			0.089***	[0.023]	0.088***	[0.023]	0.111***	[0.031]
Computer (general)			0.132***	[0.017]	0.093***	[0.016]	0.100***	[0.018]
Female=1 # Cognitive							0.030	[0.017]
Female=1 # Social							−0.007	[0.011]
Female=1 # Character							−0.043***	[0.011]
Female=1 # Writing/language							0.000	[0.016]
Female=1 # Customer Service							−0.046**	[0.016]
Female=1 # Management							−0.026	[0.016]
Female=1 # Financial							−0.041	[0.026]
Female=1 # Computer (general)							−0.015	[0.017]
Base controls	X		X		X		X	
Occupation FEs								
Firm controls								
Firm FEs								
Observations	10313978		10313978		10313978		10313978	
R^2	0.370		0.386		0.417		0.418	

Table 6 continued

	(5) Model 2A		(6) Model 2B		(7) Model 2C		(8) Model 2D	
Female=1	−0.091***	[0.003]			−0.090***	[0.003]	−0.051***	[0.006]
Cognitive			0.007	[0.009]	0.010	[0.008]	0.012	[0.010]
Social			−0.022***	[0.006]	−0.020***	[0.005]	−0.015*	[0.007]
Character			0.017**	[0.006]	0.017**	[0.006]	0.031***	[0.007]
Writing/language			0.057***	[0.008]	0.054***	[0.008]	0.063***	[0.010]
Customer Service			0.040***	[0.009]	0.034***	[0.008]	0.057***	[0.011]
Management			−0.005	[0.008]	−0.004	[0.008]	0.000	[0.009]
Financial			0.028**	[0.009]	0.027**	[0.009]	0.051***	[0.011]
Computer (general)			0.029**	[0.010]	0.027**	[0.010]	0.034**	[0.011]
Female=1 # Cognitive							−0.005	[0.009]
Female=1 # Social							−0.013	[0.007]
Female=1 # Character							−0.029***	[0.007]
Female=1 # Writing/language							−0.022	[0.012]
Female=1 # Customer Service							−0.055***	[0.012]
Female=1 # Management							−0.009	[0.008]
Female=1 # Financial							−0.051***	[0.011]
Female=1 # Computer (general)							−0.017	[0.012]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls								
Firm FEs								
Observations	10313978		10313978		10313978		10313978	10313978
R^2	0.370		0.386		0.417		0.418	0.370

Table 6 continued

	(9) Model 3A		(10) Model 3B		(11) Model 3C		(12) Model 3D	
Female=1	−0.083***	[0.003]			−0.083***	[0.003]	−0.043***	[0.006]
Cognitive			0.000	[0.008]	0.003	[0.008]	0.007	[0.009]
Social			−0.011*	[0.005]	−0.010*	[0.005]	−0.010	[0.006]
Character			0.002	[0.005]	0.003	[0.005]	0.019**	[0.006]
Writing/language			0.026***	[0.006]	0.025***	[0.006]	0.041***	[0.008]
Customer Service			−0.002	[0.007]	−0.006	[0.007]	0.023*	[0.009]
Management			0.004	[0.007]	0.005	[0.007]	0.011	[0.008]
Financial			0.020*	[0.009]	0.017*	[0.008]	0.038***	[0.010]
Computer (general)			0.012	[0.008]	0.011	[0.008]	0.019*	[0.009]
Female=1 # Cognitive							−0.011	[0.008]
Female=1 # Social							−0.002	[0.006]
Female=1 # Character							−0.033***	[0.006]
Female=1 # Writing/language							−0.037***	[0.010]
Female=1 # Customer Service							−0.070***	[0.011]
Female=1 # Management							−0.013	[0.007]
Female=1 # Financial							−0.042***	[0.010]
Female=1 # Computer (general)							−0.023*	[0.010]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs								
Observations	10309784		10309784		10309784		10309784	
R^2	0.558		0.550		0.559		0.560	

Table 6 continued

	(13) Model 4A		(14) Model 4B		(15) Model 4C		(16) Model 4D	
Female=1	−0.079***	[0.003]			−0.079***	[0.003]	−0.036***	[0.006]
Cognitive			−0.005	[0.008]	−0.003	[0.008]	0.003	[0.009]
Social			−0.010*	[0.005]	−0.009*	[0.005]	−0.009	[0.006]
Character			0.016**	[0.006]	0.018**	[0.005]	0.040***	[0.007]
Writing/language			0.006	[0.006]	0.006	[0.006]	0.026**	[0.008]
Customer Service			−0.008	[0.008]	−0.009	[0.008]	0.019	[0.011]
Management			0.005	[0.008]	0.006	[0.007]	0.012	[0.008]
Financial			0.022**	[0.007]	0.019**	[0.007]	0.030**	[0.009]
Computer (general)			0.002	[0.007]	0.004	[0.007]	0.020**	[0.008]
Female=1 # Cognitive							−0.015	[0.008]
Female=1 # Social							0.000	[0.006]
Female=1 # Character							−0.044***	[0.006]
Female=1 # Writing/language							−0.042***	[0.011]
Female=1 # Customer Service							−0.065***	[0.012]
Female=1 # Management							−0.013*	[0.007]
Female=1 # Financial							−0.020	[0.010]
Female=1 # Computer (general)							−0.038***	[0.009]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs	X		X		X		X	
Observations	10309784		10309784		10309784		10309784	
R^2	0.592		0.584		0.592		0.593	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: 3-digit flow data: Regressions results

	(1) Model 1A		(2) Model 1B		(3) Model 1C		(4) Model 1D	
Female=1	−0.129***	[0.002]			−0.114***	[0.002]	−0.091***	[0.004]
Cognitive			0.085***	[0.008]	0.082***	[0.008]	0.066***	[0.011]
Social			−0.031***	[0.005]	−0.019***	[0.005]	−0.021***	[0.006]
Character			0.025***	[0.005]	0.021***	[0.005]	0.041***	[0.006]
Writing/language			0.063***	[0.006]	0.051***	[0.006]	0.053***	[0.008]
Customer Service			0.047***	[0.011]	0.025*	[0.010]	0.041***	[0.009]
Management			0.132***	[0.007]	0.123***	[0.006]	0.132***	[0.008]
Financial			0.087***	[0.008]	0.085***	[0.008]	0.113***	[0.009]
Computer (general)			0.110***	[0.009]	0.089***	[0.009]	0.099***	[0.009]
Female=1 # Cognitive							0.028**	[0.009]
Female=1 # Social							0.002	[0.005]
Female=1 # Character							−0.035***	[0.004]
Female=1 # Writing/language							−0.002	[0.007]
Female=1 # Customer Service							−0.034***	[0.006]
Female=1 # Management							−0.018**	[0.007]
Female=1 # Financial							−0.051***	[0.007]
Female=1 # Computer (general)							−0.021**	[0.008]
Base controls	X		X		X		X	
Occupation FEs								
Firm controls								
Firm FEs								
Observations	1202198		1202198		1202198		1202198	
R^2	0.436		0.446		0.463		0.464	

Table 7 continued

	(5) Model 2A		(6) Model 2B		(7) Model 2C		(8) Model 2D	
Female=1	−0.072***	[0.001]			−0.072***	[0.001]	−0.053***	[0.003]
Cognitive			0.006	[0.006]	0.008	[0.006]	0.005	[0.007]
Social			−0.018***	[0.004]	−0.017***	[0.004]	−0.014**	[0.004]
Character			0.008*	[0.003]	0.008*	[0.003]	0.015***	[0.004]
Writing/language			0.030***	[0.004]	0.030***	[0.004]	0.035***	[0.005]
Customer Service			0.021***	[0.005]	0.019***	[0.005]	0.031***	[0.006]
Management			0.009*	[0.004]	0.009*	[0.004]	0.016**	[0.005]
Financial			0.025***	[0.005]	0.024***	[0.005]	0.043***	[0.006]
Computer (general)			0.036***	[0.008]	0.034***	[0.008]	0.033***	[0.007]
Female=1 # Cognitive							0.004	[0.005]
Female=1 # Social							−0.007	[0.004]
Female=1 # Character							−0.013***	[0.004]
Female=1 # Writing/language							−0.011*	[0.005]
Female=1 # Customer Service							−0.027***	[0.005]
Female=1 # Management							−0.014**	[0.005]
Female=1 # Financial							−0.036***	[0.006]
Female=1 # Computer (general)							0.004	[0.007]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls								
Firm FEs								
Observations	1202198		1202198		1202198		1202198	
R^2	0.552		0.548		0.553		0.554	

Table 7 continued

	(9)		(10)		(11)		(12)	
	Model 3A		Model 3B		Model 3C		Model 3D	
Female=1	−0.066***	[0.001]			−0.066***	[0.001]	−0.050***	[0.003]
Cognitive			0.001	[0.004]	0.003	[0.004]	0.001	[0.005]
Social			−0.011**	[0.004]	−0.011**	[0.004]	−0.011**	[0.004]
Character			−0.000	[0.003]	0.000	[0.003]	0.009*	[0.004]
Writing/language			0.008*	[0.004]	0.008*	[0.004]	0.016***	[0.005]
Customer Service			−0.008	[0.005]	−0.009	[0.005]	0.003	[0.005]
Management			0.013**	[0.004]	0.012**	[0.004]	0.022***	[0.005]
Financial			0.020***	[0.005]	0.019***	[0.005]	0.032***	[0.006]
Computer (general)			0.016**	[0.006]	0.015**	[0.006]	0.018**	[0.006]
Female=1 # Cognitive							0.004	[0.005]
Female=1 # Social							0.001	[0.004]
Female=1 # Character							−0.015***	[0.003]
Female=1 # Writing/language							−0.016***	[0.004]
Female=1 # Customer Service							−0.025***	[0.005]
Female=1 # Management							−0.020***	[0.005]
Female=1 # Financial							−0.025***	[0.005]
Female=1 # Computer (general)							−0.005	[0.006]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs								
Observations	1201910		1201910		1201910		1201910	
R^2	0.568		0.564		0.569		0.569	

Table 7 continued

	(13)		(14)		(15)		(16)	
	Model 4A		Model 4B		Model 4C		Model 4D	
Female=1	−0.064***	[0.001]			−0.064***	[0.001]	−0.052***	[0.003]
Cognitive			−0.005	[0.004]	−0.004	[0.004]	−0.008	[0.004]
Social			−0.008*	[0.004]	−0.009*	[0.004]	−0.010*	[0.004]
Character			0.004	[0.003]	0.005	[0.003]	0.013***	[0.004]
Writing/language			−0.001	[0.004]	−0.001	[0.004]	0.008*	[0.004]
Customer Service			−0.013*	[0.006]	−0.012*	[0.006]	−0.004	[0.006]
Management			−0.003	[0.004]	−0.003	[0.004]	0.006	[0.004]
Financial			0.009	[0.005]	0.008	[0.005]	0.018**	[0.006]
Computer (general)			0.010	[0.006]	0.010	[0.006]	0.017***	[0.005]
Female=1 # Cognitive							0.007	[0.004]
Female=1 # Social							0.003	[0.004]
Female=1 # Character							−0.015***	[0.003]
Female=1 # Writing/language							−0.015***	[0.004]
Female=1 # Customer Service							−0.018***	[0.005]
Female=1 # Management							−0.016***	[0.004]
Female=1 # Financial							−0.016***	[0.005]
Female=1 # Computer (general)							−0.014*	[0.006]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs	X		X		X		X	
Observations	1201910		1201910		1201910		1201910	
R^2	0.593		0.589		0.593		0.594	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets
** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

The results from the 4-digit stock and flow data are similar and are included in the appendix in Table 17 and in Table 18 respectively. Thus, the results outlined in the above are robust to including more detailed occupation FEs. Despite the fact that occupational segregation can explain some of the variation in returns to skills, even *within* detailed occupations, significant gender differences in returns to most skills are observed. This finding holds after controlling for both individual characteristics and firm specific effects. Another consistent finding is that requirements of cognitive and social skill generally only correlate weakly with wages after introducing controls. Thus, my results do not align with those of Deming and Kahn (2018) who find that high levels of cognitive, social and particularly of both skills correlate with higher wages. However, they only consider a subsample of workers, namely professionals, and they include the dual requirement of both social and cognitive skills in their regressions.

To be able to compare results directly with Deming and Kahn (2018), subsamples professionals from both the 3-digit stock and flow datasets are also analysed, and the requirement of both cognitive and social skills in combination is added in regressions.[13] Consider particularly model (4) in their Table 3 (see copy in the appendix). This model is similar to my model 3C in terms of controls, but Deming and Kahn (2018) only have MSA-level wage information. Results are reported in Table 8.[14] For the stock data, the coefficient on cognitive skills is positive and marginally significant in model 3C, but after introducing the female*skills interactions in model 3D, the coefficient is no longer significantly different from zero. A similar story holds for the flow data, but the coefficient remains marginally significant in model 3D. The coefficient on social skills is negative and significant for both model 3C and 3D when estimated on the stock data, and insignificant when estimated on the flow data. Finally, the requirement of both cognitive and social skills is insignificant for both specifications and both datasets. Thus, the results here a very different from those of Deming and Kahn (2018). One explanation for Deming and Kahn's (2018) significant coefficients on requirements of cognitive, social and both skills could be that they do not cluster their standard errors. Even so, the magnitudes and signs of the coefficients are also different from those estimated on individual-level data here. The small and insignificant coefficients on cognitive skills may be explained by the "The Great Reversal in the Demand for Skill and Cognitive Tasks" (Beaudry et al. 2016). Although cognitive skills used yield high returns, the increase in educational attainment has created an abundance of skilled workers with high levels of cognitive skills.

Even so, significant, positive and large coefficients are observed on writing/language, customer service, and financial skills across all specifications. Furthermore, the female dummy interactions with character, customer service, management, and financial skills are significant and negative both when estimated on the stock and flow data. The female*computer interaction is positive and significant – this was not the case for the full

[13] Here, professionals are crudely defined as workers with the 1-digit DISCO-codes "1", "2", or the 3-digit code "321". Deming and Kahn (2018) use SOC-codes, but the crosswalk between SOC- and DISCO-codes include numerous many-to-many walks, so it cannot be applied in this context.
[14] In the appendix, tables 19-22, I report results from all models estimated with the additional skill requirement of both social and cognitive skills.

sample of workers. This corresponds somewhat with Lindley (2012) who shows that women lost out from technological change because of lower levels of numeracy skills (perhaps similar to financial skills). However, the results also suggest that even with the same levels of financial skills, female professionals face lower returns them than men.

The next section includes estimates of model 4D on numerous different subsamples, not only professionals. This exercise serves as a robustness check – the estimates reported above may be driven by particular subpopulations.

Table 8: 3-digit stock and flow data: Regressions results, professionals only

	(9)		(10)		(11)		(12)	
	Stock				*Flow*			
	Model 3C		Model 3D		Model 3C		Model 3D	
Female=1	−0.080***	[0.003]	−0.048***	[0.008]	−0.067***	[0.001]	−0.055***	[0.004]
Cognitive	0.036*	[0.015]	0.028	[0.017]	0.040***	[0.011]	0.033*	[0.013]
Social	−0.029**	[0.010]	−0.031**	[0.011]	−0.009*	[0.004]	−0.008	[0.006]
Cognitive & Social	0.016	[0.023]	0.030	[0.026]	−0.004	[0.014]	−0.002	[0.016]
Character	0.026*	[0.010]	0.045***	[0.011]	0.003	[0.004]	0.010*	[0.005]
Writing/language	0.070***	[0.014]	0.065***	[0.013]	0.026***	[0.005]	0.023***	[0.006]
Customer Service	0.048**	[0.015]	0.084***	[0.016]	0.031***	[0.008]	0.059***	[0.009]
Management	−0.025	[0.014]	−0.013	[0.014]	0.005	[0.005]	0.011	[0.006]
Financial	0.040**	[0.013]	0.057***	[0.016]	0.043***	[0.006]	0.056***	[0.008]
Computer (general)	0.023	[0.014]	0.005	[0.013]	0.031**	[0.011]	0.014	[0.011]
Female=1 # Cognitive			0.017	[0.014]			0.011	[0.008]
Female=1 # Social			−0.000	[0.010]			−0.003	[0.005]
Female=1 # Cognitive & Social			−0.036	[0.020]			−0.005	[0.011]
Female=1 # Character			−0.040***	[0.008]			−0.013**	[0.004]
Female=1 # Writing/language			0.008	[0.010]			0.004	[0.005]
Female=1 # Customer Service			−0.101***	[0.011]			−0.073***	[0.008]
Female=1 # Management			−0.029**	[0.011]			−0.012*	[0.006]
Female=1 # Financial			−0.040**	[0.013]			−0.028***	[0.006]
Female=1 # Computer (general)			0.051***	[0.014]			0.042***	[0.007]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs								
Observations	4186116		4186116		433659		433659	
R^2	0.495		0.497		0.525		0.526	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets
** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Robustness

A couple of robustness checks are performed. First, model 4D, which includes firm FEs and all the other controls, are estimated on a number of subpopulations. Next quantile regressions are introduced to allow coefficients on skills requirements to changes along the wage distribution. Finally, the gender difference in mean hourly wages is decomposed using Oaxaca-Blinder decomposition methods.

Subpopulations

The first robustness check focusses on model 4D, which include all occupation and firm FEs as well as numerous other controls. The model is estimated again on the following subpopulations:

a. Professionals (as defined above)
b. Workers in large firms (with 100 or more employees)
c. Workers in small firms (with fewer than 100 employees)
d. Workers in public "firms"
e. Workers in private firms

This exercise may reveal that the results from the full datasets may be "driven" by certain subgroups. Results for each subpopulation for both stock and flow data are reported in Table 9. First, note that the insignificance of the coefficients on cognitive and social skills generally hold across subpopulations (not for the public sector). Furthermore, the coefficients on the female interactions with character, writing/language, customer service, management, and financial skills are consistently negative, and they are also significant across most subpopulations. The most notable exception is for public sector workers where the coefficients on skill measures and their interaction with the female dummy generally are insignificant. As public workers are overrepresented in the matched vacancy-register data, this distinction is important.

Table 9: Model 4D, different subpopulations

3-digit stock data

	(1) Prof.		(2) Large		(3) Small		(4) Public		(5) Private	
Female=1	−0.036***	[0.005]	−0.034***	[0.006]	−0.053***	[0.003]	−0.041***	[0.006]	−0.054***	[0.010]
Cognitive	0.010	[0.009]	0.002	[0.009]	0.000	[0.007]	−0.006	[0.012]	0.016	[0.011]
Social	−0.002	[0.009]	−0.010	[0.007]	0.004	[0.004]	−0.023**	[0.008]	0.010	[0.007]
Character	0.045***	[0.011]	0.043***	[0.007]	0.016***	[0.004]	0.012	[0.008]	0.044***	[0.009]
Writing/lang.	0.024**	[0.009]	0.027**	[0.009]	0.016**	[0.005]	0.004	[0.013]	0.038***	[0.010]
Customer Ser.	0.032*	[0.014]	0.020	[0.012]	0.019**	[0.006]	−0.010	[0.013]	0.011	[0.012]
Management	0.010	[0.013]	0.012	[0.008]	0.010	[0.006]	0.026**	[0.010]	0.005	[0.011]
Financial	0.016	[0.012]	0.029**	[0.010]	0.024***	[0.007]	0.000	[0.011]	0.031*	[0.012]
Computer	0.020	[0.012]	0.021*	[0.009]	0.015*	[0.007]	0.003	[0.013]	0.013	[0.009]
F=1#Cognitive	−0.008	[0.007]	−0.017	[0.009]	−0.006	[0.007]	−0.007	[0.011]	−0.031**	[0.011]
F=1#Social	0.011	[0.007]	0.000	[0.007]	−0.007	[0.004]	0.002	[0.008]	−0.016	[0.008]
F=1#Character	−0.049***	[0.006]	−0.047***	[0.007]	−0.025***	[0.004]	−0.014	[0.010]	−0.041***	[0.010]
F=1#Writing/lan.	−0.005	[0.007]	−0.043***	[0.012]	−0.031***	[0.005]	−0.024	[0.021]	−0.037**	[0.014]
F=1#Cus. Ser	−0.094***	[0.010]	−0.066***	[0.013]	−0.048***	[0.005]	0.010	[0.013]	−0.040**	[0.013]
F=1#Management	−0.019**	[0.007]	−0.013	[0.007]	−0.020**	[0.007]	−0.020*	[0.009]	−0.023	[0.012]
F=1#Financial	−0.029**	[0.011]	−0.019	[0.012]	−0.031***	[0.007]	0.020*	[0.010]	−0.035**	[0.012]
F=1#Computer	−0.010	[0.008]	−0.039***	[0.010]	−0.032***	[0.007]	−0.027	[0.018]	−0.009	[0.009]
Base controls	X		X		X		X		X	
Occupation FEs	X		X		X		X		X	
Firm controls	X		X		X		X		X	
Firm FEs	X		X		X		X		X	
Observations	4185064		9318381		991403		6101901		4207883	
R^2	0.572		0.593		0.601		0.564		0.618	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9 continued

3-digit flow data

	(1) Prof.		(2) Large		(3) Small		(4) Public		(5) Private	
Female=1	−0.058***	[0.003]	−0.052***	[0.003]	−0.049***	[0.006]	−0.054***	[0.004]	−0.049***	[0.004]
Cognitive	0.002	[0.005]	−0.008	[0.005]	0.001	[0.009]	−0.031***	[0.007]	0.010	[0.006]
Social	−0.001	[0.004]	−0.011**	[0.004]	0.003	[0.006]	−0.000	[0.005]	−0.011*	[0.005]
Character	0.001	[0.004]	0.014***	[0.004]	0.002	[0.006]	0.002	[0.005]	0.014**	[0.005]
Writing/lang.	0.002	[0.005]	0.008	[0.004]	0.007	[0.007]	0.010	[0.007]	0.012*	[0.005]
Customer Ser.	0.024**	[0.007]	−0.004	[0.006]	0.013	[0.008]	−0.029**	[0.010]	−0.006	[0.006]
Management	0.002	[0.005]	0.005	[0.005]	0.021*	[0.009]	−0.005	[0.006]	0.015*	[0.006]
Financial	0.017**	[0.006]	0.018**	[0.006]	0.017	[0.013]	−0.002	[0.008]	0.018*	[0.007]
Computer	0.016**	[0.006]	0.016**	[0.006]	0.017*	[0.009]	0.019	[0.011]	0.012*	[0.005]
F=1#Cognitive	0.014***	[0.004]	0.007	[0.004]	−0.005	[0.011]	0.026***	[0.005]	−0.012	[0.006]
F=1#Social	0.006	[0.004]	0.004	[0.004]	−0.002	[0.006]	−0.005	[0.005]	0.000	[0.004]
F=1#Character	−0.008	[0.004]	−0.016***	[0.003]	−0.000	[0.006]	−0.001	[0.004]	−0.018***	[0.004]
F=1#Writing/lan.	0.004	[0.005]	−0.016***	[0.004]	−0.012	[0.008]	−0.013*	[0.006]	−0.018***	[0.005]
F=1#Cus. Ser	−0.068***	[0.007]	−0.017***	[0.005]	−0.027**	[0.009]	0.017	[0.009]	−0.016***	[0.005]
F=1#Management	−0.008	[0.005]	−0.015***	[0.004]	−0.034**	[0.011]	−0.004	[0.005]	−0.029***	[0.007]
F=1#Financial	−0.013*	[0.005]	−0.015**	[0.005]	−0.041**	[0.014]	0.007	[0.006]	−0.027***	[0.006]
F=1#Computer	−0.005	[0.006]	−0.012	[0.007]	−0.040***	[0.011]	−0.005	[0.010]	−0.010	[0.006]
Base controls	X		X		X		X		X	
Occupation FEs	X		X		X		X		X	
Firm controls	X		X		X		X		X	
Firm FEs	X		X		X		X		X	
Observations	433605		1134469		67441		755652		446258	
R^2	0.570		0.591		0.674		0.547		0.678	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets

** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Quantile regressions

Blau and Kahn (2017) point out that the relative gender pay gap increases along the wage income distribution. Therefore, gender differences in returns to skills may also change along the income distribution. To check if this is the case, I run a set of quantile regressions at the 0.25 quantile (the lower quartile), 0.50 quantile (the median), and 0.75 quantile (the upper quartile), but due to computational requirements only on model 1D where base controls alone are included. Recall that the quantile regression estimator, $\hat{\beta}_\tau$, minimises the following objective function (Cameron & Trivedi 2005:87):

$$Q_N(\beta_\tau) = \sum_{i:y_i \geq I_i\,\beta}^{N} \tau|w_i - I_i\,\beta_\tau| + \sum_{i:y_i < I_i\,\beta}^{N} (1-\tau)|w_i - I_i\,\beta_\tau|$$

Where w_i is the log hourly wage, and I_i is a matrix of individual characteristics and skill levels, and τ is the quantile of interest. The objective function is an asymmetrically weighted average of absolute errors. Positive errors are weighted by τ and negative errors by $(1-\tau)$. Finally, note that the expression collapse to the Least Absolute Deviations Estimator at the median (since $\tau = 1 - \tau$ if $\tau = 0.5$). See the documentation for the "qreg" STATA-package for information on the WLS algorithm which is applied to estimate $\hat{\beta}_\tau$.[15] The estimated coefficients are reported in Table 10, but note that the

[15] https://www.stata.com/manuals13/rqreg.pdf

"qreg"-package cannot estimate cluster-robust standard errors, so robust standard errors are applied instead.[16] Thus, disregard the standard errors because they are not clustered.

First, note that the unexplained gender pay gap clearly increases at the upper quartile of the wage distribution both when estimated using the stock and flow data. Coefficients on cognitive skills are also larger at the median and upper quartile. The same goes for character, customer service, management, financial skills in the stock data, but only management and financial skills in the flow data. The coefficients on the remaining skill measures are roughly similar along the wage distribution.

In the stock data, the coefficients on the female*skills interactions are pretty stable. An exception is the coefficient on cognitive skills which is large and positive at the lower quartile. Another exception is the female*writing/language interaction, which is positive at the upper quartile for the stock data. For the flow data, the coefficients on the female*skills interactions tend to be of smaller magnitudes at the lower quartile, but of similar magnitudes at the median and the upper quartile. Again, an exception is the coefficient on cognitive skills which is large and positive at the lower quartile.

In sum, the coefficients on the female interactions with character, customer service, management, and financial skills are consistently negative. The coefficient on female*cognitive skills is particularly large at the lower quartile. The next section further examines gender differences in returns to skills and their contribution towards the mean gender pay gap applying Oaxaca-Blinder decomposition methods.

[16] The STATA-package "qreg2" computes cluster-robust standard errors, but does not allow weights when doing so. Weighting with "qreg" is prioritised to give correct points estimates.

Table 10: Model 1D, Quantile Regression

3-digit stock data

	(1)		(2)		(3)	
	0.25		0.50		0.75	
Female=1	−0.059***	[0.001]	−0.079***	[0.001]	−0.130***	[0.001]
Cognitive	0.063***	[0.001]	0.112***	[0.001]	0.151***	[0.001]
Social	−0.021***	[0.001]	−0.020***	[0.001]	−0.010***	[0.001]
Character	0.036***	[0.001]	0.051***	[0.001]	0.064***	[0.001]
Writing/language	0.071***	[0.001]	0.079***	[0.001]	0.076***	[0.001]
Customer Service	0.042***	[0.001]	0.091***	[0.001]	0.134***	[0.001]
Management	0.085***	[0.001]	0.119***	[0.001]	0.153***	[0.001]
Financial	0.071***	[0.001]	0.085***	[0.001]	0.113***	[0.001]
Computer (general)	0.100***	[0.001]	0.117***	[0.001]	0.107***	[0.001]
Female=1 # Cognitive	0.029***	[0.001]	−0.000	[0.001]	−0.007***	[0.002]
Female=1 # Social	0.018***	[0.001]	0.007***	[0.001]	−0.018***	[0.001]
Female=1 # Character	−0.034***	[0.001]	−0.043***	[0.001]	−0.050***	[0.001]
Female=1 # Writing/language	−0.012***	[0.001]	−0.013***	[0.001]	0.009***	[0.001]
Female=1 # Customer Service	−0.058***	[0.001]	−0.057***	[0.001]	−0.041***	[0.001]
Female=1 # Management	−0.027***	[0.001]	−0.040***	[0.001]	−0.040***	[0.001]
Female=1 # Financial	−0.023***	[0.001]	−0.023***	[0.001]	−0.032***	[0.002]
Female=1 # Computer (general)	−0.056***	[0.001]	−0.055***	[0.001]	0.005**	[0.002]
Base controls	X		X		X	
Occupation FEs						
Firm controls						
Firm FEs						
Observations	10313978		10313978		10313978	

Observations weighted by full-time equivalents. Robust standard errors in brackets

** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

3-digit flow data

	(1)		(2)		(3)	
	0.25		0.50		0.75	
Female=1	−0.071***	[0.001]	−0.068***	[0.002]	−0.096***	[0.002]
Cognitive	0.041***	[0.002]	0.062***	[0.003]	0.091***	[0.003]
Social	−0.022***	[0.001]	−0.018***	[0.002]	−0.023***	[0.002]
Character	0.042***	[0.001]	0.034***	[0.001]	0.037***	[0.002]
Writing/language	0.037***	[0.002]	0.045***	[0.002]	0.053***	[0.003]
Customer Service	0.028***	[0.002]	0.014***	[0.002]	0.028***	[0.002]
Management	0.109***	[0.002]	0.126***	[0.002]	0.151***	[0.003]
Financial	0.099***	[0.002]	0.108***	[0.003]	0.127***	[0.004]
Computer (general)	0.102***	[0.002]	0.103***	[0.003]	0.097***	[0.003]
Female=1 # Cognitive	0.029***	[0.003]	0.007*	[0.003]	0.005	[0.004]
Female=1 # Social	0.004*	[0.002]	0.003	[0.002]	0.010***	[0.002]
Female=1 # Character	−0.037***	[0.002]	−0.038***	[0.002]	−0.040***	[0.002]
Female=1 # Writing/language	0.009**	[0.003]	−0.013***	[0.003]	−0.020***	[0.003]
Female=1 # Customer Service	−0.012***	[0.002]	−0.044***	[0.003]	−0.046***	[0.003]
Female=1 # Management	−0.015***	[0.003]	−0.032***	[0.003]	−0.036***	[0.004]
Female=1 # Financial	−0.044***	[0.003]	−0.053***	[0.004]	−0.064***	[0.004]
Female=1 # Computer (general)	−0.033***	[0.003]	−0.038***	[0.004]	−0.023***	[0.004]
Base controls	X		X		X	
Occupation FEs						
Firm controls						
Firm FEs						
Observations	1202198		1202198		1202198	

Observations weighted by full-time equivalents. Robust standard errors in brackets

** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Decompositions

This section very briefly introduces the procedure of an Oaxaca-Blinder decomposition of the differences in mean hourly wages between women and men before results are reported.[17] First, estimate a standard wage regression separately for women and men:

$$w_i = I_i\beta + \varepsilon_i$$

Where I_i is a matrix of individual characteristics and skill levels and w_i log hourly earnings. Let $\hat{\beta}_M$ denote the estimated vector of coefficients for the subsample of men and $\hat{\beta}_W$ for women. Furthermore, denote $\bar{I}_M$ the sample mean male characteristics and $\bar{I}_W$ the sample mean female characteristics. Let $\widehat{\Delta}^{\mu}_{M-W}$ denote the difference in sample mean hourly wages between men and women. The difference in the sample mean hourly wages between men and women can then be decomposed as follows (see Fortin et al. 2011 for derivations):

$$\widehat{\Delta}^{\mu}_{M-W} = \bar{I}_M\left(\hat{\beta}_W - \hat{\beta}_M\right) + (\bar{I}_M - \bar{I}_W)\hat{\beta}_W$$

Or analogously:

$$\widehat{\Delta}^{\mu}_{M-W} = \bar{I}_W\left(\hat{\beta}_M - \hat{\beta}_F\right) + (\bar{I}_M - \bar{I}_W)\hat{\beta}_M$$

The first term is referred to as the "unexplained" share of the difference in means, as it is due to gender differences in coefficients/returns to the same characteristics, e.g. gender differences in returns to skills. The second term is referred to as the "explained" share, as this is due to level differences in characteristics. Note that two entirely separate wage regressions are estimated for women and men. This means that all coefficients, β, are allowed to differ between women and men, and not just the intercept. The mean gender difference in hourly wages is decomposed using both the male, female and pooled coefficients as reference levels (Fortin et al. 2011). Two decompositions are reported, one which shows each skill's contributions, and one where only the total contribution of skills is shown. The results for the 3-digit stock and flow data are reported in Table 11 and Table 12 together with cluster-robust standard errors.

The decompositions reveal that, after allowing the coefficients on all other variables to vary by gender, gender differences in returns to each individual skill only constitute a relatively small share of the unexplained gender pay gap. An exception is character skills to which gender differences in coefficients account for more than a tenth of the total unexplained gender pay gap across all decompositions. Note also that women's wages correlate positively with cognitive skills, and more so than men's, which reduces the unexplained gender pay gap – again this holds across decompositions. Finally, the total *explained* gender pay gap is relatively small, and gender differences in levels of skills can explain a relatively large share of this in the stock data, but less so in the flow data.

[17] I have applied Oaxaca-Blinder decompostions in two other projects: my bacherlor thesis (2015) and an econometrics exam project (2017). This was, however, on different data and in different contexts, but the notation that follows is somewhat similar.

Table 11: Decompositions of the gender pay gap in mean hourly wages, detailed skills

	3-digit stock data			*3-digit flow data*		
	(1)	(2)	(3)	(1)	(2)	(3)
	Male coef.	Female coef.	Pooled coef.	Male coef.	Female coef.	Pooled coef.
Mean ln(hour wage):						
- Men	5.411***	5.411***	5.411***	5.228***	5.228***	5.228***
	[0.010]	[0.010]	[0.010]	[0.005]	[0.005]	[0.005]
- Women	5.244***	5.244***	5.244***	5.074***	5.074***	5.074***
	[0.007]	[0.007]	[0.007]	[0.003]	[0.003]	[0.003]
- Difference	0.167***	0.167***	0.167***	0.154***	0.154***	0.154***
	[0.007]	[0.007]	[0.007]	[0.004]	[0.004]	[0.004]
Explained						
Total	0.038***	0.029***	0.033***	0.044***	0.036***	0.040***
	[0.006]	[0.006]	[0.006]	[0.003]	[0.003]	[0.003]
- Cognitive	0.002***	0.002***	0.002***	0.001***	0.002***	0.002***
	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]
- Social	0.001	0.001*	0.001	0.001***	0.001***	0.001***
	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]
- Character	0.002***	0.001	0.001**	0.001***	0.000	0.001***
	[0.001]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]
- Writing/language	0.005***	0.005***	0.005***	0.002***	0.002***	0.002***
	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]
- Customer Service	0.007***	0.003***	0.005***	0.003***	0.000	0.002**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
- Management	0.003***	0.003***	0.003***	0.004***	0.003***	0.004***
	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]
- Financial	0.002*	0.001*	0.002*	0.002***	0.001***	0.001***
	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]
- Computer (general)	0.005***	0.004***	0.004***	0.004***	0.003***	0.004***
	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]
- Controls	0.012**	0.009*	0.010*	0.027***	0.022***	0.024***
	[0.005]	[0.005]	[0.005]	[0.003]	[0.003]	[0.003]
Unexplained						
Total	0.129***	0.138***	0.134***	0.109***	0.118***	0.114***
	[0.005]	[0.005]	[0.005]	[0.002]	[0.002]	[0.002]
- Cognitive	-0.002	-0.003	-0.002	-0.003***	-0.004***	-0.004***
	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
- Social	0.008	0.007	0.007	-0.000	-0.000	-0.000
	[0.008]	[0.007]	[0.008]	[0.003]	[0.003]	[0.003]
- Character	0.019***	0.020***	0.019***	0.012***	0.013***	0.012***
	[0.005]	[0.005]	[0.005]	[0.002]	[0.002]	[0.002]
- Writing/language	0.000	0.000	0.000	-0.000	-0.001	-0.000
	[0.002]	[0.003]	[0.002]	[0.001]	[0.001]	[0.001]
- Customer Service	0.004***	0.008***	0.006***	0.003***	0.005***	0.004***
	[0.001]	[0.003]	[0.002]	[0.001]	[0.001]	[0.001]
- Management	0.004	0.005	0.004	0.001	0.001	0.001
	[0.003]	[0.003]	[0.003]	[0.001]	[0.001]	[0.001]
- Financial	0.004*	0.004*	0.004*	0.003***	0.004***	0.003***
	[0.002]	[0.003]	[0.002]	[0.001]	[0.001]	[0.001]
- Computer (general)	0.001	0.001	0.001	0.000	0.001	0.001
	[0.001]	[0.002]	[0.001]	[0.000]	[0.001]	[0.001]
- Controls	-0.087***	-0.084***	-0.085***	-0.105***	-0.100***	-0.102***
	[0.026]	[0.027]	[0.027]	[0.015]	[0.015]	[0.015]
- Constant	0.179***	0.179***	0.179***	0.199***	0.199***	0.199***
	[0.023]	[0.023]	[0.023]	[0.016]	[0.016]	[0.016]
Base controls	X	X	X	X	X	X
Occupation FEs						
Firm controls						
Firm Fes						
Observations	10,313,978	10,313,978	10,313,978	1,202,198	1,202,198	1,202,198

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Decompositions of the gender pay gap in mean hourly wages, grouped skills

	3-digit stock data			*3-digit flow data*		
	(1)	(2)	(3)	(1)	(2)	(3)
	Male coef.	Female coef.	Pooled coef.	Male coef.	Female coef.	Pooled coef.
Mean ln(hour wage):						
- Men	5.411***	5.411***	5.411***	5.228***	5.228***	5.228***
	[0.010]	[0.010]	[0.010]	[0.005]	[0.005]	[0.005]
- Women	5.244***	5.244***	5.244***	5.074***	5.074***	5.074***
	[0.007]	[0.007]	[0.007]	[0.003]	[0.003]	[0.003]
- Difference	0.167***	0.167***	0.167***	0.154***	0.154***	0.154***
	[0.007]	[0.007]	[0.007]	[0.004]	[0.004]	[0.004]
Explained						
Total	0.038***	0.029***	0.033***	0.044***	0.036***	0.040***
	[0.006]	[0.006]	[0.006]	[0.003]	[0.003]	[0.003]
- All skills	0.026***	0.020***	0.024***	0.018***	0.014***	0.016***
	[0.003]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
- Controls	0.012**	0.009*	0.010*	0.027***	0.022***	0.024***
	[0.005]	[0.005]	[0.005]	[0.003]	[0.003]	[0.003]
Unexplained						
Total	0.129***	0.138***	0.134***	0.109***	0.118***	0.114***
	[0.005]	[0.005]	[0.005]	[0.002]	[0.002]	[0.002]
- All skills	0.037***	0.042***	0.039***	0.015***	0.019***	0.017***
	[0.011]	[0.012]	[0.012]	[0.004]	[0.005]	[0.005]
- Controls	-0.087***	-0.084***	-0.085***	-0.105***	-0.100***	-0.102***
	[0.026]	[0.027]	[0.027]	[0.015]	[0.015]	[0.015]
- Constant	0.179***	0.179***	0.179***	0.199***	0.199***	0.199***
	[0.023]	[0.023]	[0.023]	[0.016]	[0.016]	[0.016]
Base controls	X	X	X	X	X	X
Occupation FEs						
Firm controls						
Firm Fes						
Observations	10,313,978	10,313,978	10,313,978	1,202,198	1,202,198	1,202,198

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets
** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

In sum, gender differences in returns to skills do matter in the formation of the unexplained gender pay gap. Before introducing decomposition, it was observed that the unexplained gender pay gap decreased when female*skills interactions were introduced in models 1D-4D in the previous section. Lastly, Table 12 shows that differences in returns to skills can account for a significant share of the unexplained gender pay gap, and this share is particularly large in the stock data.

Limitations

The results above are presented with some caution. As I am the first to utilise the combination of Danish job vacancy data and register data, this is a central contribution of this thesis. Although I use a tried-and-tested method to extract skills from job posts' keywords, a next step would be to validate my skills measures, e.g. using occupation level skills data from O*NET, DOT, European Social Survey, or the OECD PIAAC Survey. This would also give an indication of whether or not workers actually apply the skills that are required in a job post after getting the job. Furthermore, simply creating dummy variables of skills for each job posts is also not ideal. Developing continuous skill measures, which also contain information on skills intensities, is a natural next step. Furthermore, interactions of different skills should also be considered. For now, I only looked at the dual requirement of cognitive and social skills. Other skills may complement each other.

Furthermore, I present only correlational evidence; I do not identify casual effects. However, the correlations do suggest that causal analyses are warranted to further explore the skills dimension of work. To identify casual effects, one could exploit the panel dimension of the Danish register and consider individual-level effects of workers changing from one job to another with a different set of required skills.

Discussion

Keeping the above-mentioned limitations in mind, this section discusses my results in relation to the existing literature. The combination of Danish job vacancy data and individual-level register data is unique to this study. Internationally, only one study has merged job vacancy data with individual-level data, but with much lower match rates (Kettemann et al. 2018). Thus, I am among the first to be able to utilise individual-level variation in characteristics together with data from job vacancies. In this thesis, I choose to focus on variation in skills and in returns to skills across genders.

The data analysis confirms that a large gender pay gap still exists, and that individual characteristics, e.g. educational attainment, and occupational segregation explain some, but far from all of the pay gap (cf. Blau & Kahn 2017; Goldin 2014; Lindley 2016). Furthermore, I also find that women and men are very segregated at the firm*occupation-level, which aligns with findings of high levels of occupational and industrial gender segregation (cf. Jarman et al. 2012; Levanon & Grusky 2016; Ngai & Petrongolo 2017; Olivetti & Petrongolo 2014).

Moving on technological change and returns to skills I confirm that patterns of employment polarisation are observed over the period 1991-2009, and that these patterns are more pronounced for women (cf. Black & Spitz-Oener 2010). In my most basic models, I find positive correlations between wages and cognitive skills, but after adjusting for occupation FEs these coefficients are insignificant. The coefficient on social skills is negative across all models, but insignificant after introducing more comprehensive sets of controls. The female interactions with cognitive and social skills are also generally insignificant. Also, the coefficient on the dual requirement of both social and cognitive skills is insignificant. This contrast a number of studies on returns to skills, which highlight the interaction between social and cognitive skills as particularly important (Deming 2017; Deming & Kahn 2018; Weinberger 2014). The insignificant coefficients on cognitive skills may be explained by a reversal in the demand for cognitive since year 2000 (cf. Beaudry et al. 2016), but a longer time series of skills data would be necessary to confirm this. For the same reason, I cannot confirm whether or not changing skills prices caused a narrowing of the gender pay gap (as pointed out by Bacolod & Blum 2010; Beaudry & Lewis 2014; Rendall 2010; Yamaguchi 2018).

However, my results do indicate that differences in returns to skills contribute to the gender pay gap (cf. Bizopoulou 2017; Fedorets 2014; Lindley 2012). More specifically, I generally find negative and significant coefficients on the female interactions with character, writing/language, customer service, management, financial, and computer skills. Thus, ignoring the gendered dimension of returns to skills would lead to biased results, overestimating the returns to skills for women and underestimating them for men. However, most studies focus on social and cognitive skills, e.g. Deming and Kahn (2018), and for these skills I do not find any gender differences in returns. However, my results still contrast those of Deming and Kahn (2018) as I do not find indications of positive returns to the requirement of cognitive, social or both of these skills.

Conclusions

In this part of my thesis, a combination of Danish job vacancy and register data is operationalised for the first time. I derive skills measures from job posts by a reading of keywords. The vacancy and register data are matched on the firm*occupation- and firm*occupation*start-month-levels, which involves some aggregation. However, a high degree of gender segregation at these levels preserves variation in skills across genders. With this novel combination of data, it is possible to show that similar skills are required by women and men in aggregate. Some gender differences are observed, but they surprisingly small considering the high degree of gender segregation in the labour market. Furthermore, variation in skills cannot be entirely explained by variation in occupations or other sets of covariates. Keeping this in mind, the skills measures are included in wage regressions. All skills, except social skills, correlate positively with hourly wages in the basic models with only individual controls. After including additional control variables in regressions, the general positive wage effect of skills diminishes. However, the coefficients on female*skills interactions tell a different story. Even after including occupation FEs, firm FEs, and a very extensive set of control variables, the female interactions with character, writing/language, customer service, management, financial, and computer skills are generally negative and significant. Thus, ignoring the gendered dimension of returns to skills would lead to biased results. This result is driven by workers in the private sector. Furthermore, I show that the unexplained gender pay gap and the returns to skills tend increase along the wage distribution, whereas the gender differences in returns are relative stable along the distribution. Finally, I apply Oaxaca-Blinder decompositions to show that gender differences in returns to skills do account for some of the unexplained gender pay gap. The presented results are merely correlations; they do, however, serve to warrant future causal analyses.

PART II: Gendered Work(ers)

Introduction

When considering the supply of labour, neoclassical economists tend to model work only as a means to an end. Work generates disutility, but nevertheless, individuals choose to work in order to generate income with which they then can finance consumption (see e.g. Cahuc & Zylberberg 2004). More consumption and more leisure result in higher levels of utility. An hour worked is also an hour of foregone leisure. Thus, work is only worth undertaking because it is income-generating, and the generated income constitutes sole value of labour to worker. Outside the discipline of economics, however, work or labour is perceived as something more complex than an income-generating process.

Philosophers, sociologists, and psychologists all emphasise the centrality of work to human life. Thus, this chapter first explores the interdependence of work, life, and identity. In fact, Hegel, Marx, and Freud all point out that work is exactly what *makes* us humans (Budd 2011:155). Second, I establish that an identity of work relies on the notion of occupations. Social identities are constructed from work, but particularly through occupational categories. Workers are contained in an occupational category, which is different from "the others". Occupations are mutually exclusive groups: you are one or the other. Third, occupations are inherently gendered, and thus, a work identity is also gendered. Workers "do" or "perform" gender when going to work in a gendered occupation. Looking closer, this gendered system is circular and keeps reinforcing itself. The gendered system remains even though the skills required by women and men are similar in aggregate (see Figure 9). Thus, Part II of the thesis aims to answer the following research question:

1. How may a reconceptualisation of work as the utilisation of skills, rather than as mutually exclusive occupational categories, affect the gendered occupational identities of workers?

Conceptualising work as the utilisation of skills with an indefinite number of potential values and combinations is radically different from the current conceptualisation of work as a limited set of mutually exclusive occupational categories. Through an application of the skills framework, workers may realise that they are similar across occupations; they apply the same skills. Another way the skills framework helps undoing occupations is by looking within occupations. Focussing on variation in skills within occupations allows the group identity, and thus the group, to break down; workers realise that they are indeed different from their occupational identity.

It follows that breaking down occupational categories by the use of the skills framework, may in fact also undo gender at work, as the gendered character of work is bound to occupational identities. If occupational categories dissolve, and gendered work identities are undone, the circular gendered system of occupational gender segregation may also break down.

Limiting occupational gender segregation is not only desirable because it can explain a significant share of the gender pay gap (recall the difference between models 1A and 2A above). Occupational gender segregation also creates "tokens" of displaced workers

(Kanter 1977): for example, "a woman in a man's job" (Faulkner 2013). Token workers experience this displacement through a subjective disparity between their individual identity and their occupational identity, often with feelings of being "estranged, alienated" from their labour (Marx 1959:25). Thus, undoing gender segregation may also reduce feelings of estrangement and alienation amongst tokenised workers.

Work and identity

Labour supply can be conceptualised straightforwardly in neo-classical economic models as a solely income-generating process. What matters is the number of hours worked and the wage rate. Other theories of work conceptualise labour supply in a significantly different way, for example by emphasising the nature of the work itself, and how it affects the formation of human identities. First, I revisit some theories of work itself, which proceeds into a discussion of the construction of occupational identities.

Think of a full-time worker, working 40 hours a week and sleeping 8 hours a day. Then a quick calculation reveals that the worker spends more than a third of all woken hours working every week. So even in a naïve and solely quantitative sense, work is a significant part of life for most human beings. Economists acknowledge this too. But the significance of the work itself, not the hours, for human life should not be underestimated. In fact, work can be seen as exactly what defines human life:

> *It is just in his work upon the objective world, therefore, that man really proves himself to be a species-being. This production is his active species-life. Through this production, nature appears as his work and his reality. The object of labor is, therefore, the objectification of man's species-life: for he duplicates himself not only, as in consciousness, intellectually, but also actively, in reality, and therefore he sees himself in a world that he has created. (Marx 1959:24)*

Here, Marx points out that human production goes beyond the production of what is immediately needed to survive, and thus, human production is distinguishable from that of animals. In other words, our work and production is precisely what makes us different from animals. It ia not only a Marxist perspective that work characterises humans and human life. For example, Hegel, Heidegger, Freud, and the Catholic church all emphasise the "centrality of work for humanness" (Budd 2011:154). Because work is so central in human life, workers do also define themselves from the work they do: "We need work, and as adults we find identity and are identified by the work we do." (Gini 1998:714). Here, it becomes apparent that the nature of the work itself matters for life. One can easily see this: Keeping the number of hours worked and the wage fixed, most people can think of a job they would despise, and one they would love. Furthermore, work identity does not form from undertaking a certain set of tasks, but rather from belonging to a certain group of workers; an occupation. Indeed, occupational identity "refers to the conscious awareness of oneself as a worker" and this occupational identity "represents a core, integrative element of identity [...] but also as a major factor in the emergence of meaning and structure in individuals' lives" (Skorikov & Vondracek 2011:169). The conceptualisation of work through occupations establishes group-dependent occupational identities. Occupational categorisation gives workers a sense of "belonging". Social identity theory and social categorisation theory (Tajfel 1978; Tajfel & Turner 1986; Turner 1982) explain how identity is formed in relation to others, by belonging or not belonging with the "others". Here, self-categorisation is an important aspect – people categorise themselves as members of certain groups and non-member of others, which then constitute their social identity:

> *Social categories [...] do not merely systematize the social world; they also provide a system of orientation for* self-reference*: they create and define the individual's place in society. Social groups, understood in this sense, provide their members with an identification of themselves in social terms. These identifications are to a very large extent relational and comparative: they define the individual as similar to or different from, as "better" or "worse" than, members of other groups. It is in a strictly limited sense, arising from these considerations, that we use the term* social identity*. It consists, for the purposes of the present discussion, of those aspects of an individual's self-image that derive from the social categories to which he perceives himself as belonging.* (Tajfel & Turner 2004:283)

Thus, the importance of occupational categories should not be underestimated as they form a point of reference for relationally constructed social identities. In other words, the categorisation assigns workers to social groups which underpin the formation of a social identity of their members. Realising this, sociologists have also studied "occupational classes", as well as the following class inequality and what Marxists would refer to as "class consciousness" (Crompton 2010). Occupations are not equal-ranking, rather they are perceived as hierarchal and indicative of social status:

> *As such, there appears to be a consensus among sociologists that "in modern societies occupation is one of the most salient characteristics to which status attaches," and therefore to which identity attaches. Occupation can also be a major determinant of one's lifestyle, including consumption patterns, leisure interests, and values, and occupation is therefore a major unit of analysis in the sociology of work." (Budd 2011:150)*

In other words, work defines human life, and through occupational categories also social status, identities, and thus, behaviours. Occupational categories unite the members within an occupation, but divide across categories, creating an "us" and "them". This is not only acknowledged by philosophers, sociologists, and psychologists, recently also economists have included identity and social categories in economic models (e.g. Akerlof & Kranton 2000, 2005). Because work is so crucial in the formation of our social and occupational identity "we must be very careful about what we choose to do for a living, for what we do is what we'll become." (Gini 1998:714). In sum, "Occupations shape workers' identity" (Martin 2013a:324). This realisation is the starting point for the following section, where the gendered character of occupations is explored.

The gendered work(er)

As pointed out above, occupational categories have the potential to divide, and to create an "us" and "them". Gender, and particularly gender segregation, is an essential dimension to consider here. From Figure 3, it is clear that women and men, in general, do not tend to mix at the firm*occupation-level. Because of extensive this gender segregation, most occupations are conceived as being either stereotypical female or male. Max Weber even argued that gender segregation followed from natural differences in women's and men's abilities:

> *According to Weber, bureaucracies require men and the stereotypical practices of masculinity, for example, rational-logical thought, means-end calculation, decisive action, and top-down control. Whereas men behave in these ways naturally, women are constitutionally incapable of doing so. Bureaucracies are thus places for men... (Martin 2013b:283–84)*

Although some still adhere to similar gender essentialist explanations of gender segregation (e.g. Hakim 2000), most gender scholars – and particularly feminists – point out that biology have little explanatory power in the context of occupational segregation (Crompton 2007). Instead, feminists emphasise how occupational segregation maintains "sex-typed" or gendered occupational identities (Acker 1990; In & Brinton 1984). The observation of gender segregation itself is paradoxical because, in aggregate, similar skills are required of women and men in the labour market (Figure 9). Even so, gender identities and occupational identities are often aligned:

> *People have a gender, and their gender rubs off on the jobs they do. The jobs in turn have a gender character which rubs off on the people who do them. (Cockburn 1988:38)*

It follows that occupations and organisations within which work takes place are gendered, and "Gender, then, is not simply brought into organizations by members, like a virus that infects an otherwise gender-neutral space." (Orzechowicz & Martin 2013:319). Rather, "the gender identity of jobs and occupations is repeatedly reproduced" (Acker 1990:145), and thus, it becomes apparent that the gendered occupational system is circular. In other words, gendered workers repeatedly enter occupations in a pattern that uphold the notion that women do "women's work" and men do "men's work". Why is this important? Because going to work then becomes a way to "do" or to "perform" gender. West and Zimmerman (1987) emphasises that gender is socially constructed and something that is "done" when people act in ways that are perceived masculine and feminine. In *Gender Trouble*, Butler (2006:34) elaborates: "*gender* is not a noun, but neither is it a set of free-floating attributes, for we have seen that the substantive effect of gender is performatively produced and compelled by the regulatory practices of gender coherence." Here, Butler does not only emphasise the "doing", the performative production of gender, but crucially also its "regulatory practices". Gender is produced through performative actions within "highly rigid regulatory frame" (Butler 2006:45). Work, and particularly occupations, form part of the rigid structures that make up this gendered frame, and hence, going to work becomes a performative action in the construction of gender. As a result, the gendered nature of occupations upholds gendered identities, and vice versa. Acknowledging the performative production of gender at work is crucial because conceptualising both occupations and gender as performatively constructed "suggests that it has no ontological status apart from the various acts which constitute its reality." (Butler 2006:185). In other words, gender is only gender because it is performatively produced, and there is not necessarily a pre-existing "doer behind the deed" (Butler 2006:195). The same goes for occupations and their gendered characters, which similarly can be seen as constructed through rituals and repetitive acts performed by its members. Gender and occupational identities only exist because they are "performed" or "done" by the people who identify themselves as members of each category.

Transgressions

Plenty of evidence shows that gender segregation is not a result of "differences in knowledge, skills, or abilities" (Orzechowicz & Martin 2013:319). Figure 9 supports this: similar skill requirements are observed for women and men in aggregate. Economists, too, have shown that even conditional on similar levels of human capital, women consistently receive lower hourly wages than men, and a significant share of this wage gap can be explained by occupational segregation (e.g. Blau & Kahn 2017). Thus, the gendered character of occupations also causes gender inequality. This is one of the unfortunate outcomes that results from "doing" gender at work: "women's work" is undervalued. Still, it should not be neglected that some individuals defy the gendered character of occupations; a woman enters what may be perceived as a "male" occupation, e.g. engineering:

> *...women engineers are perceived, and can feel themselves, to be not quite 'real engineers' or 'real women'. Men engineers belong more 'naturally' on both fronts, while women have to do additional identity work on both fronts if they are to secure their membership... (Faulkner 2013:181)*

When an individual's social identities contradict each other, additional identity work is necessitated. However, such identity work is strenuous, and thus, these workers face adversities from acting outside the "highly rigid regulatory frame" of gender or occupations (Butler 2006:45):

> *This juxtaposition of one's sex category and the gender of one's job complicates the process of doing gender and often has disappointing outcomes, such as social ostracization by colleagues or failure on the job. (Reid 2013:192)*

As a worker enters an occupation with which their gender identity is not aligned, the worker becomes the "odd one out". In other words, the underrepresented worker becomes a "token" (Kanter 1977). Thus, occupational segregation does not only lead to gender pay differences, but also to tokenisation of the worker who transgresses the gendered character of an occupation. The token worker faces a series of challenges: "Visibility generates performance pressures; polarization leads dominants to heighten their group boundaries; and assimilation leads to the tokens' role entrapment." (Kanter 1977:965). Because the token worker is hyper-visible and thus constantly observed, performance is also always evaluated. The majority workers are less visible, and thus, not constantly assessed. The majority workers seldom face the stress of constant assessment to which the token workers are subjected. Polarisation within an occupational category threatens the coherent group identity, making separate points of reference for the token and majority workers. This again furthers the divide between tokens and the majority. Finally, assimilation makes work an "activity of alienation, of estrangement" for the token worker (Marx 1959:26). Kanter's (1977) concept of the token worker demonstrates the high costs an individual worker faces if entering an occupation with which their gender identity does not align. The high individual costs of becoming a token worker further fortifies occupational segregation. Thus, the costs of occupational segregation are not simply monetary; segregation is also reinforced as the gendered occupations keep outsiders away by imposing additional identity work on token workers. As a result, occupational choices remain gendered; workers realise that there is a price to pay if they enter an occupation dominated by a gender with which they do not identify.

Thus, there are at least two gendered costs of occupational gender segregation. Firstly, gender differences in pay arising from segregation; this is mainly a cost imposed on women. Secondly, occupational choices are, to a high degree, limited by workers' gender identities as they anticipate the costs of tokenisation.

The skills framework and the "undoing" of occupations

Segregation on its own may not be problematic as the "opposite of equality is not difference, but inequality" (Crompton 2007:232). Women and men could potentially do different work without adverse consequences. However, occupational gender segregation and the following sex-typing of work both play a part in the generation of gender pay inequalities and in the tokenisation of workers with alienation as a potential consequence. Thus, an argument against segregation is easy to make. But how could occupational gender segregation come to an end? Many researchers have offered their advice, for example by pinpointing the effects of gendered institutional settings and by proposing relevant altercations (e.g. Browne 2013). In what follows, I do not propose an instrument that has the potential to end segregation. However, I do propose a reconceptualisation of work, which alters the categorisation of workers and, in the long run, this may affect occupational choices, and thus, patterns of segregation.

As pointed out above, occupational categories are formed relationally. Particularly, an occupational group identity is based both on differences from "the others" and on similarities amongst the members of the group in question. Thus, the key to understand the formation of occupational identities is to understand how group differences and similarities are perceived. For most workers, one of these group similarities is their gender identity. Furthermore, occupations are static entities. Workers are assigned an occupation by either its employer or by an institution (e.g. their union, their unemployment insurance fund, or Statistics Denmark). A job can only offer the worker membership of one occupational group; occupations are mutually exclusive. However, a worker may change job and with that also occupation, leaving behind entirely their former occupational membership with the high social costs that follow.

The skills framework applied in the empirical analysis above offers another, more dynamic perspective on work. Firstly, the skills framework makes it apparent that even though women and men are segregated in the labour market, the same skills are required of them in aggregate. Secondly, it also becomes apparent that even within narrowly defined occupations, there is significant variation in both levels and returns to skills. Thus, occupational categories do only partially – in terms of skills - account for differences from "the other" and similarities within the occupational groups. If work instead was conceptualised and perceived as a collection of skills, workers could, potentially, leave the socially constructed categories of occupations behind. With a skills approach to work, one may, for example, realise that the work of a care assistant and car mechanic are not all that different. Their work requires precision and strength, their work is dirty, and they both have the potential to kill their clients if they do not do their work properly. With the current, gendered conceptualisation of work as occupations, however, it is very hard to grasp these similarities. Socially constructed barriers and a "rigid regulatory frame" are in the way (Butler 2006:45). Thus, the skills approach offers a new understanding of work, where workers are not "one" or "the other". Workers' skills overlap between occupations, and thus, similarities across occupations appear, and differences within occupations stand out. With the skills framework, workers may realise that their occupational identity is indeed socially constructed, not exclusive, nor fixed. Such a realisation is not without implications:

> *Paradoxically, the reconceptualization of identity as an* effect*, that is, as* produced *or* generated*, opens up possibilities of "agency" that are insidiously foreclosed by positions that take identity categories as foundational and fixed. (Butler 2006:206)*

Thus, workers' agency may be recovered if occupations are deconstructed into skills, and work reconceptualised as a collection of skills. The care assistant may then choose to become a car mechanic, and vice versa. Skills are, in aggregate, not gendered, and thus, the gendered dimension of work would dissolve together with the deconstruction of occupations. If work *is* skills, rather than occupational categories, work would no longer be "doing gender" (West & Zimmerman 1987). Therefore, if the skills framework breaks down the categorical notion of an occupation, it also enables transgressions of the boundary between the female and male worker. Female work becomes a multi-dimensional collection of skills and so does its male counterpart; it becomes clear that the perceived reality of the gendered worker is only superficial and merely an effect of repeated performative actions within rigidly defined occupational categories.

Conceptualising work through skills rather than occupational categories may appear as a purely theoretical exercise. However, the effects of telling workers that they belong to a certain group, here an occupation, should not be underestimated. Both institutions and colleagues reinforce the assignment of workers occupational categories. Plenty of studies document how people are obedient to such instructions and change their behaviour in response to authorities (e.g. the classics Milgram 1974; Zimbardo et al. 1973). Thus, conceptualising work as skills may, in practice, have behavioural consequences. If Statistics Denmark, employers, unions, and other authorities changed their definition of work from DISCO-codes to a multi-dimensional collection of skills, workers may regain some agency. The gendered occupational categories would, perhaps, become less important for workers' social identities. Such an exercise could potentially help worker to avoid negative implications of tokenism, and thus in the long run, of gender segregation. I took a first step in that direction in the empirical analysis above by documenting gender similarities in aggregate skill demands, variation in skills within and across occupations, and finally gender differences in returns to skills.

Conclusions

An empirical analysis of gender differences in the deployment of and returns to skills does not merely provide further insights into pay gaps between women and men. Equally important, my analysis opened up a discussion of the difference between the female worker and female work, as well as between the male worker and male work. This represents a move away from the essentialist approach taken by many economists where no such distinction is made. The distinction becomes crucial as the skills framework breaks down mutually exclusive occupational categories, which are inherently gendered. Occupational gender segregation is costly, not only because of gender pay inequalities, but also because of the tokenisation of workers whose occupational choice and gender identity do not align. However, reconceptualising work through the skills framework also facilitates transgressions of the boundaries between occupations, and thus, between the female and male work. Female work becomes multi-dimensional and so does its male counterpart; it becomes clear that the perceived reality of the gendered worker is only superficially constructed. This realisation allows a move beyond segregation; a move beyond an occupational division of workers to a multi-dimensional approach to work.

CONCLUDING REMARKS

The gendered implications of a reconceptualisation of work as the utilisation of skills form a common theme for parts I and II. In part I, it becomes apparent that similar skills are required by women and men in aggregate. This is surprising as women and men otherwise are very segregated in the Danish labour market. Additionally, variation in skills cannot be fully explained by other variables that are normally included in wage regressions. Thus, skills add another dimension to the understanding labour market outcomes. One of these labour market outcomes, the gender pay gap, appears to be partly explained by gender differences in returns to skills. The correlational evidence presented here warrants causal studies of the gendered effect of skills on labour market outcomes, including wages.

Part II commences with a discussion of the interdependence between work, occupation and the identity of workers. By revisiting theories of work and identity, it becomes apparent that workers' social identities rely on occupational categorisation. Furthermore, occupations are inherently gendered, and a circular system of occupational gender segregation remains in place. Next, I again utilise the skills framework developed in Part I. A reconceptualisation work as the utilisation of skills may crucially enable workers to notice similarities across occupations and differences within. Thus, the skills framework has the potential to foster transgressions of the gendered boundaries between occupational categories, and thus in the long run, allow a move beyond segregation; a move beyond an occupational division of workers to a multi-dimensional approach to work.

The theoretical discussions in part II rely on the skills framework and the results from part I. In other words, the discussions in part II were enabled by the empirical analyses in part I. Thus, only together, the two methodological approaches allowed me two answer all of my three research questions.

Appendix

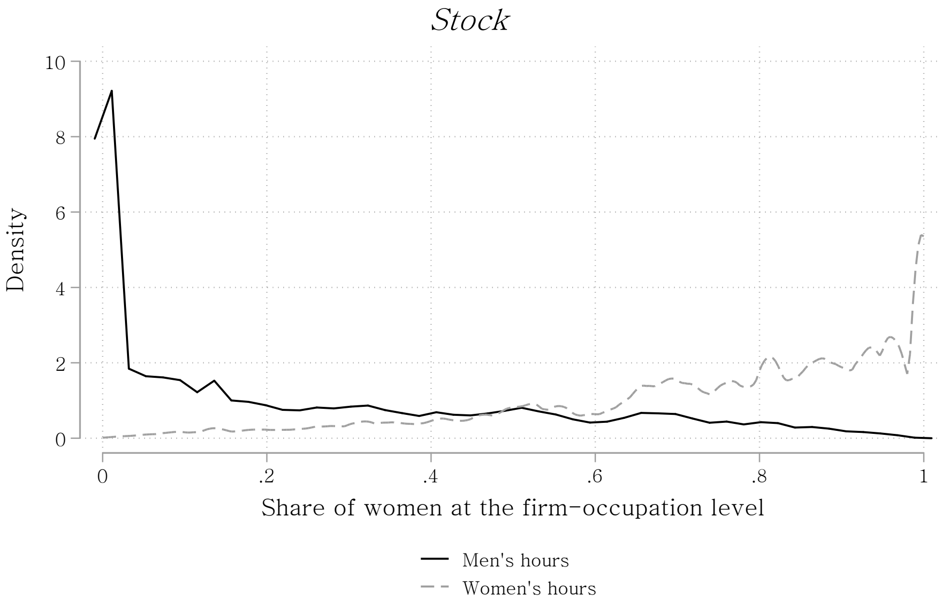
Table 13: 4-digit stock data: Match rates

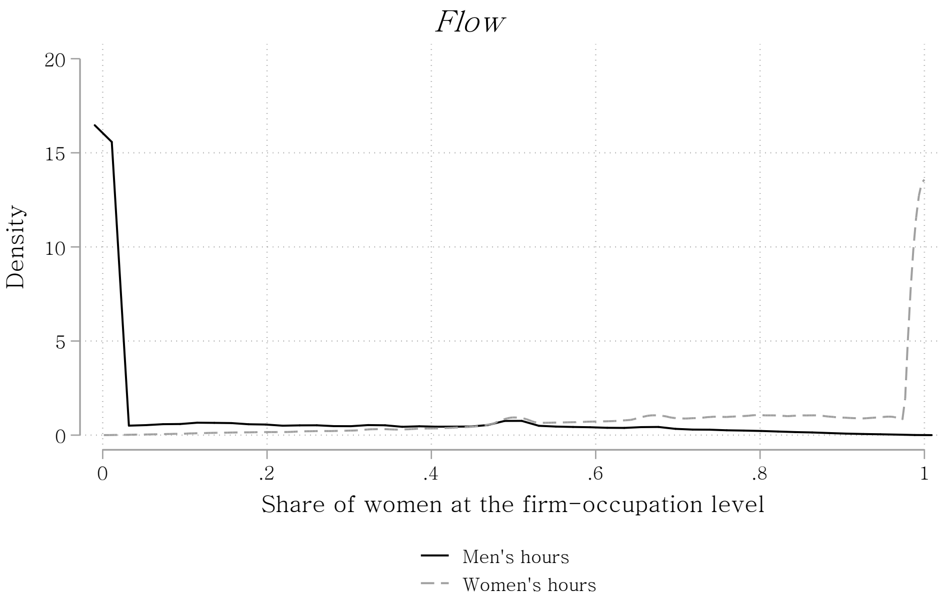
Year	Number of jobs	Matched jobs	% of jobs matched	Number of job posts	Matched posts	% of posts matched
2010	2,790,805	1,236,301	0.44	91,184	51,181	0.56
2011	2,819,170	1,307,372	0.46	96,367	53,121	0.55
2012	2,753,342	1,309,601	0.48	108,950	63,294	0.58
2013	2,749,339	1,339,414	0.49	116,788	71,427	0.61
2014	2,819,786	1,383,059	0.49	124,802	75,131	0.60
2015	2,863,468	1,398,419	0.49	136,320	83,083	0.61
2016	2,902,920	1,395,908	0.48	140,080	81,695	0.58
Total	19,698,830	9,370,074	0.48	814,491	478,932	0.59

Table 14: 4-digit flow data: Match rates

Year	Number of jobs	Matched jobs	% of jobs matched	Number of job posts	Matched posts	% of posts matched
2011	872,320	178,386	0.20	96,367	31,866	0.33
2012	785,000	171,592	0.22	108,950	40,298	0.37
2013	794,574	188,720	0.24	116,788	46,932	0.40
2014	906,366	197,403	0.22	124,802	49,204	0.39
2015	860,198	188,895	0.22	136,320	54,553	0.40
Total	4,218,458	924,996	0.22	583,227	222,853	0.38

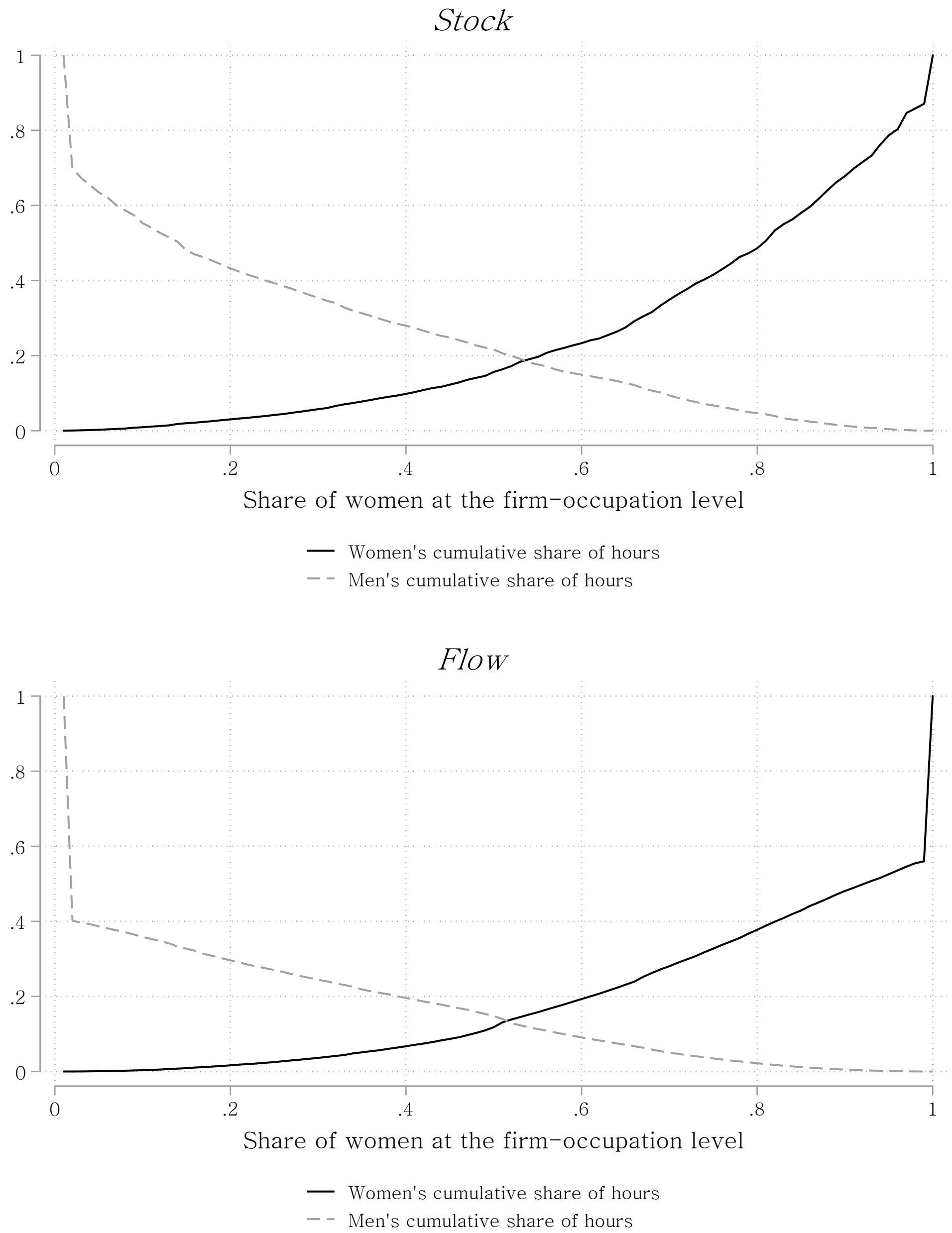
Figure 11: 4-digit data: Density of hours worked





*Source: BFL 2010-2016, excluding observation with missing CVR- or DISCO-codes. Kernel density estimates of hours worked on the share of women in the workers' firm*occupation cell*

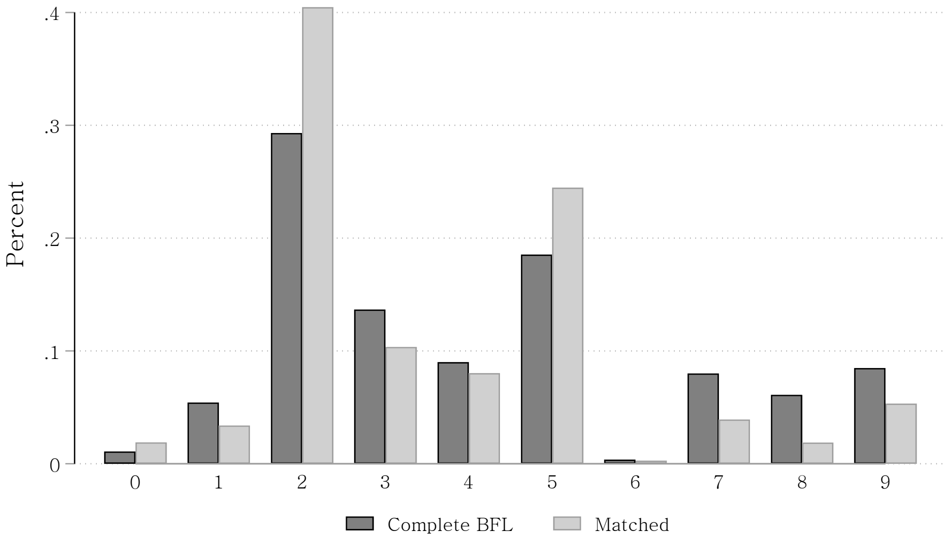
Figure 12: 4-digit data: Cumulative distribution of hours worked



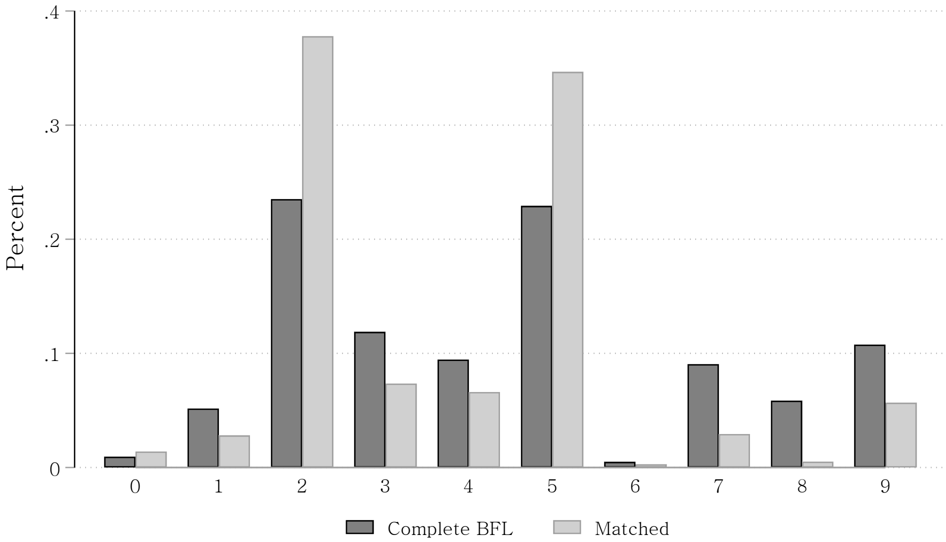
*Source: BFL 2010-2016, excluding observation with missing CVR- or DISCO-codes. Cumulative distribution of hours worked on the share of women in the workers' firm*occupation cell.*

Figure 13: 4-digit data: Distributions of occupations in BFL and matched data

Stock



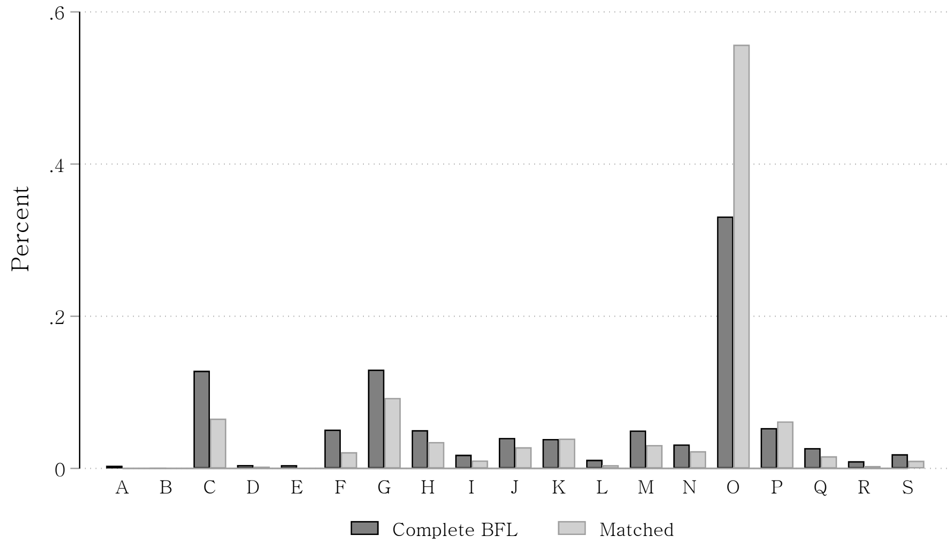
Flow



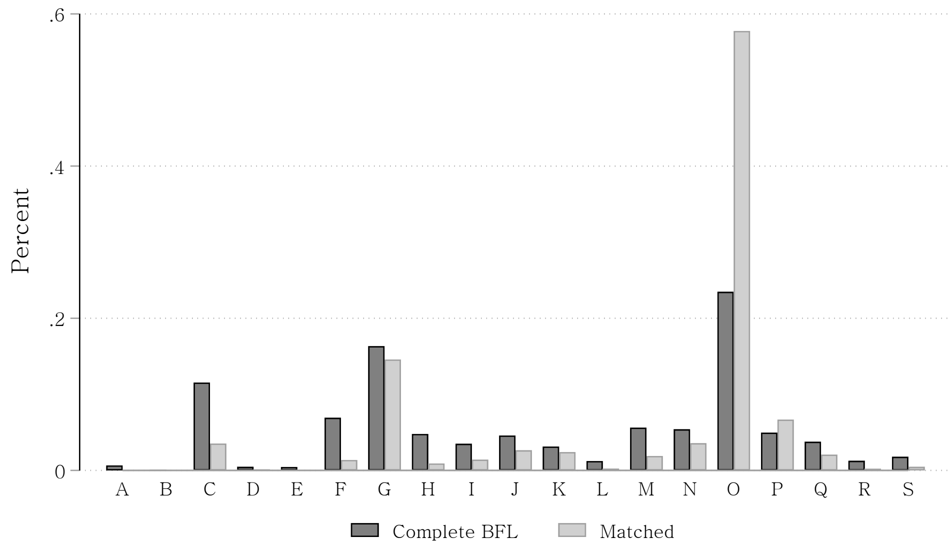
Source: BFL, HBS-Jobindex 2010-2016.
Note: Observations weighted by full-time equivalents.

Figure 14: 4-digit data: Distributions of industries in BFL and matched data

Stock



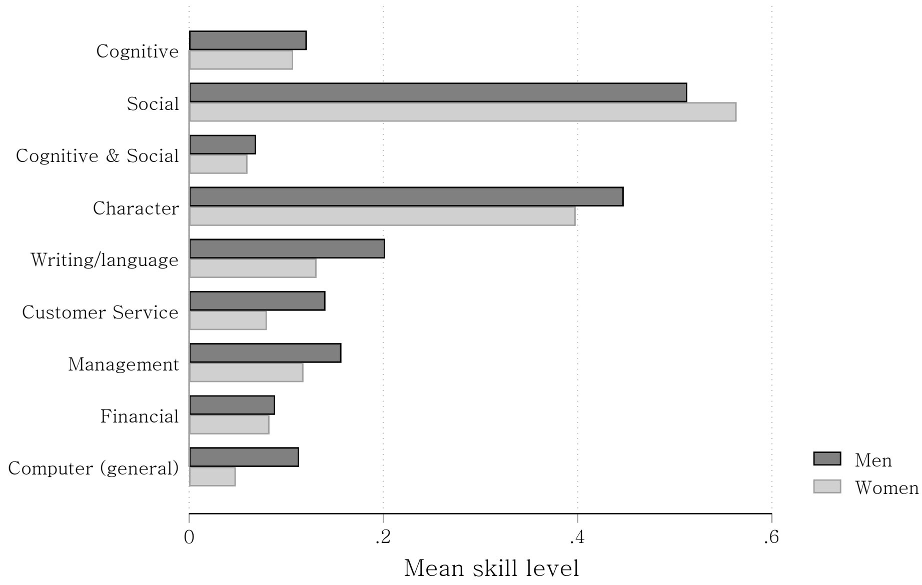
Flow



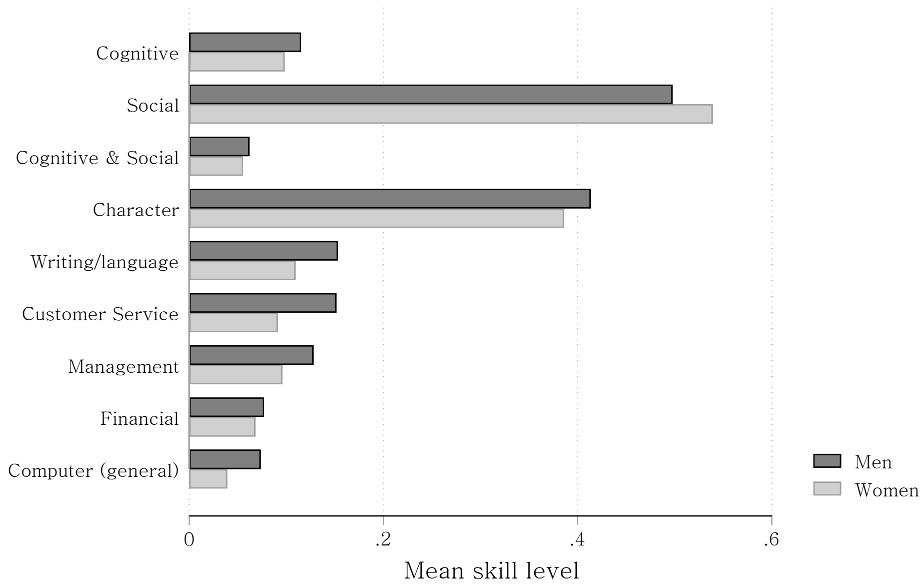
Source: BFL, HBS-Jobindex 2010-2016.
Note: Observations weighted by full-time equivalents

Figure 15: 4-digit data: Mean skill levels by gender

Stock



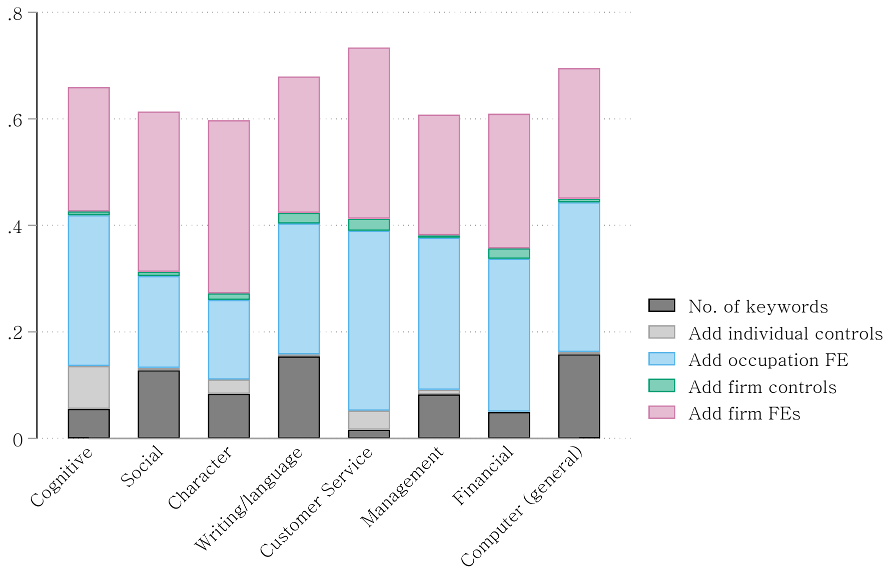
Flow



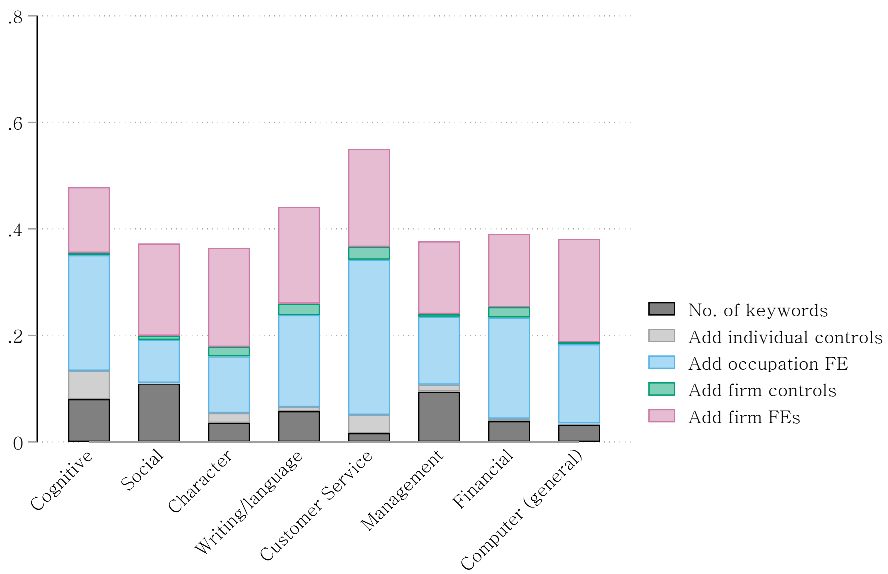
Source: BEF, HBS-Jobindex 2010-2016.
Note: Observations weighted by full-time equivalents

Figure 16: 4-digit data: Adjusted R^2 from regressions of skills level on various controls

Stock



Flow



Source: Various registers, HBS-Jobindex 2010-2016.
Note: Skills are regressed on various sets of controls.
The sets of controls are described in the section on regression models.

Table 15: Occupation Titles

1-digit code	Occupation title[18]
0	Armed Forces Occupations
1	Managers
2	Professionals
3	Technicians and Associate Professionals
4	Clerical Support Workers
5	Services and Sales Workers
6	Skilled Agricultural, Forestry and Fishery Workers
7	Craft and Related Trades Workers
8	Plant and Machine Operators and Assemblers
9	Elementary Occupations

Table 16: Industry Titles

1-letter code	Industry title[19]
A	Agriculture, forestry and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam and air conditioning supply
E	Water supply; sewerage; waste management and remediation activities
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transporting and storage
I	Accommodation and food service activities
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific and technical activities
N	Administrative and support service activities
O	Public administration and defence; compulsory social security
P	Education
Q	Human health and social work activities
R	Arts, entertainment and recreation
S	Other services activities

[18] http://www.ilo.org/public/english/bureau/stat/isco/docs/structure08.docx

[19] http://ec.europa.eu/competition/mergers/cases/index/nace_all.html

Table 17: 4-digit stock data: Regression results

	(1) Model 1A		(2) Model 1B		(3) Model 1C		(4) Model 1D	
Female=1	-0.153^{***}	[0.005]			-0.130^{***}	[0.006]	-0.116^{***}	[0.012]
Cognitive			0.130^{***}	[0.023]	0.125^{***}	[0.023]	0.108^{***}	[0.028]
Social			-0.038^{**}	[0.012]	−0.021	[0.012]	−0.024	[0.017]
Character			0.031^{*}	[0.012]	0.026^{*}	[0.012]	0.042^{**}	[0.016]
Writing/language			0.071^{***}	[0.019]	0.053^{**}	[0.019]	0.033	[0.023]
Customer Service			0.128^{***}	[0.020]	0.096^{***}	[0.020]	0.120^{***}	[0.023]
Management			0.147^{***}	[0.020]	0.127^{***}	[0.021]	0.144^{***}	[0.025]
Financial			0.089^{***}	[0.024]	0.095^{***}	[0.023]	0.131^{***}	[0.027]
Computer (general)			0.068^{*}	[0.033]	0.029	[0.035]	0.015	[0.042]
Female=1 # Cognitive							0.030	[0.019]
Female=1 # Social							0.006	[0.016]
Female=1 # Character							-0.032^{*}	[0.015]
Female=1 # Writing/language							0.044^{*}	[0.019]
Female=1 # Customer Service							-0.059^{***}	[0.017]
Female=1 # Management							-0.042^{*}	[0.018]
Female=1 # Financial							-0.066^{***}	[0.019]
Female=1 # Computer (general)							0.050	[0.033]
Base controls	X		X		X		X	
Occupation FEs								
Firm controls								
Firm FEs								
Observations	9010127		9010127		9010127		9010127	
R^2	0.378		0.386		0.416		0.417	

Table 17 continued

	(5) Model 2A		(6) Model 2B		(7) Model 2C		(8) Model 2D	
Female=1	-0.071^{***}	[0.003]			-0.071^{***}	[0.003]	-0.040^{***}	[0.005]
Cognitive			0.003	[0.007]	0.004	[0.007]	0.007	[0.008]
Social			-0.010^{*}	[0.005]	−0.010	[0.005]	−0.006	[0.006]
Character			0.017^{**}	[0.006]	0.016^{**}	[0.006]	0.030^{***}	[0.007]
Writing/language			0.036^{***}	[0.007]	0.036^{***}	[0.007]	0.039^{***}	[0.007]
Customer Service			0.017	[0.009]	0.015	[0.009]	0.042^{***}	[0.010]
Management			−0.003	[0.008]	−0.002	[0.008]	0.004	[0.008]
Financial			0.012	[0.007]	0.013	[0.007]	0.031^{***}	[0.009]
Computer (general)			0.013	[0.007]	0.012	[0.007]	0.012	[0.008]
Female=1 # Cognitive							−0.010	[0.008]
Female=1 # Social							−0.008	[0.006]
Female=1 # Character							-0.027^{***}	[0.006]
Female=1 # Writing/language							−0.007	[0.007]
Female=1 # Customer Service							-0.067^{***}	[0.009]
Female=1 # Management							−0.013	[0.007]
Female=1 # Financial							-0.034^{***}	[0.009]
Female=1 # Computer (general)							0.000	[0.013]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls								
Firm FEs								
Observations	9010127		9010127		9010127		9010127	
R^2	0.573		0.568		0.574		0.575	

Table 17 continued

	(9) Model 3A		(10) Model 3B		(11) Model 3C		(12) Model 3D	
Female=1	-0.066^{***}	[0.002]			-0.066^{***}	[0.002]	-0.033^{***}	[0.005]
Cognitive			0.001	[0.006]	0.002	[0.006]	0.008	[0.007]
Social			−0.006	[0.005]	−0.006	[0.004]	−0.004	[0.006]
Character			0.009	[0.005]	0.008	[0.005]	0.023^{***}	[0.006]
Writing/language			0.017^{**}	[0.006]	0.017^{**}	[0.006]	0.025^{***}	[0.007]
Customer Service			0.001	[0.008]	−0.001	[0.008]	0.030^{***}	[0.009]
Management			−0.000	[0.007]	0.000	[0.007]	0.008	[0.007]
Financial			0.006	[0.006]	0.007	[0.006]	0.020^{*}	[0.008]
Computer (general)			0.000	[0.006]	−0.001	[0.006]	0.003	[0.008]
Female=1 # Cognitive							−0.014	[0.008]
Female=1 # Social							−0.003	[0.005]
Female=1 # Character							-0.029^{***}	[0.005]
Female=1 # Writing/language							-0.017^{**}	[0.006]
Female=1 # Customer Service							-0.076^{***}	[0.009]
Female=1 # Management							-0.018^{**}	[0.007]
Female=1 # Financial							-0.024^{**}	[0.009]
Female=1 # Computer (general)							−0.007	[0.013]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs								
Observations	9006687		9006687		9006687		9006687	
R^2	0.586		0.581		0.587		0.588	

Table 17 continued

	(13) Model 4A		(14) Model 4B		(15) Model 4C		(16) Model 4D	
Female=1	-0.063^{***}	[0.002]			-0.062^{***}	[0.002]	-0.030^{***}	[0.005]
Cognitive			−0.007	[0.006]	−0.006	[0.006]	0.001	[0.007]
Social			−0.004	[0.004]	−0.004	[0.004]	−0.004	[0.005]
Character			0.019^{***}	[0.005]	0.019^{***}	[0.005]	0.039^{***}	[0.006]
Writing/language			0.005	[0.005]	0.005	[0.005]	0.013^{*}	[0.006]
Customer Service			-0.016^{*}	[0.007]	-0.016^{*}	[0.007]	0.016	[0.008]
Management			−0.010	[0.008]	−0.010	[0.007]	0.000	[0.007]
Financial			0.012	[0.006]	0.012	[0.006]	0.022^{**}	[0.008]
Computer (general)			0.002	[0.007]	0.003	[0.006]	0.012	[0.010]
Female=1 # Cognitive							-0.017^{*}	[0.008]
Female=1 # Social							0.001	[0.005]
Female=1 # Character							-0.034^{***}	[0.005]
Female=1 # Writing/language							-0.017^{*}	[0.007]
Female=1 # Customer Service							-0.069^{***}	[0.009]
Female=1 # Management							-0.020^{**}	[0.006]
Female=1 # Financial							-0.017^{*}	[0.009]
Female=1 # Computer (general)							−0.019	[0.015]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs	X		X		X		X	
Observations	9006687		9006687		9006687		9006687	
R^2	0.614		0.610		0.615		0.615	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: 4-digit flow data: Regression results

	(1) Model 1A		(2) Model 1B		(3) Model 1C		(4) Model 1D	
Female=1	−0.117***	[0.002]			−0.107***	[0.002]	−0.074***	[0.005]
Cognitive			0.119***	[0.009]	0.114***	[0.009]	0.091***	[0.013]
Social			−0.016**	[0.006]	−0.009	[0.006]	−0.008	[0.007]
Character			0.008	[0.006]	0.006	[0.006]	0.029***	[0.007]
Writing/language			0.041***	[0.008]	0.031***	[0.008]	0.034***	[0.010]
Customer Service			0.031**	[0.011]	0.010	[0.010]	0.032**	[0.010]
Management			0.139***	[0.008]	0.129***	[0.008]	0.153***	[0.009]
Financial			0.050***	[0.010]	0.050***	[0.009]	0.093***	[0.012]
Computer (general)			0.083***	[0.010]	0.064***	[0.010]	0.089***	[0.010]
Female=1 # Cognitive							0.041***	[0.011]
Female=1 # Social							−0.001	[0.006]
Female=1 # Character							−0.041***	[0.005]
Female=1 # Writing/language							−0.005	[0.008]
Female=1 # Customer Service							−0.046***	[0.006]
Female=1 # Management							−0.046***	[0.008]
Female=1 # Financial							−0.073***	[0.009]
Female=1 # Computer (general)							−0.054***	[0.010]
Base controls	X		X		X		X	
Occupation FEs								
Firm controls								
Firm FEs								
Observations	1011924		1011924		1011924		1011924	
R^2	0.436		0.442		0.458		0.459	

Table 18 continued

	(5) Model 2A		(6) Model 2B		(7) Model 2C		(8) Model 2D	
Female=1	−0.055***	[0.001]			−0.055***	[0.001]	−0.034***	[0.004]
Cognitive			0.000	[0.005]	0.001	[0.005]	0.002	[0.006]
Social			−0.015***	[0.004]	−0.015***	[0.004]	−0.013**	[0.004]
Character			0.007	[0.004]	0.007	[0.004]	0.019***	[0.004]
Writing/language			0.020***	[0.004]	0.020***	[0.004]	0.024***	[0.005]
Customer Service			0.016**	[0.006]	0.015*	[0.006]	0.033***	[0.007]
Management			0.006	[0.005]	0.006	[0.005]	0.017**	[0.006]
Financial			0.008	[0.006]	0.009	[0.006]	0.029***	[0.007]
Computer (general)			0.009	[0.007]	0.008	[0.007]	0.008	[0.007]
Female=1 # Cognitive							−0.003	[0.005]
Female=1 # Social							−0.005	[0.005]
Female=1 # Character							−0.020***	[0.004]
Female=1 # Writing/language							−0.007	[0.005]
Female=1 # Customer Service							−0.036***	[0.005]
Female=1 # Management							−0.021***	[0.006]
Female=1 # Financial							−0.035***	[0.007]
Female=1 # Computer (general)							0.002	[0.008]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls								
Firm FEs								
Observations	1011924		1011924		1011924		1011924	
R^2	0.583		0.580		0.584		0.584	

Table 18 continued

	(9) Model 3A		(10) Model 3B		(11) Model 3C		(12) Model 3D	
Female=1	-0.051^{***}	[0.001]			-0.051^{***}	[0.001]	-0.031^{***}	[0.004]
Cognitive			−0.003	[0.005]	−0.002	[0.005]	−0.001	[0.005]
Social			-0.011^{**}	[0.004]	-0.012^{**}	[0.004]	-0.011^{**}	[0.004]
Character			0.001	[0.004]	0.001	[0.004]	0.014^{***}	[0.004]
Writing/language			0.002	[0.005]	0.002	[0.005]	0.009	[0.005]
Customer Service			0.000	[0.006]	−0.000	[0.006]	0.017^{**}	[0.006]
Management			0.006	[0.005]	0.005	[0.005]	0.021^{***}	[0.005]
Financial			0.011	[0.007]	0.011	[0.007]	0.028^{***}	[0.007]
Computer (general)			0.002	[0.007]	0.001	[0.007]	0.002	[0.006]
Female=1 # Cognitive							−0.003	[0.005]
Female=1 # Social							−0.001	[0.005]
Female=1 # Character							-0.021^{***}	[0.004]
Female=1 # Writing/language							-0.011^{*}	[0.006]
Female=1 # Customer Service							-0.035^{***}	[0.005]
Female=1 # Management							-0.029^{***}	[0.006]
Female=1 # Financial							-0.029^{***}	[0.007]
Female=1 # Computer (general)							−0.001	[0.008]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs								
Observations	1011707		1011707		1011707		1011707	
R^2	0.595		0.592		0.595		0.596	

Table 18 continued

	(13) Model 4A		(14) Model 4B		(15) Model 4C		(16) Model 4D	
Female=1	-0.048^{***}	[0.001]			-0.048^{***}	[0.001]	-0.033^{***}	[0.003]
Cognitive			-0.012^{*}	[0.005]	-0.011^{*}	[0.005]	-0.011^{*}	[0.005]
Social			-0.014^{**}	[0.004]	-0.015^{***}	[0.004]	-0.017^{***}	[0.005]
Character			0.003	[0.004]	0.004	[0.004]	0.015^{***}	[0.004]
Writing/language			−0.003	[0.005]	−0.003	[0.005]	0.004	[0.005]
Customer Service			−0.009	[0.007]	−0.009	[0.007]	0.006	[0.007]
Management			−0.005	[0.005]	−0.005	[0.005]	0.011^{*}	[0.005]
Financial			−0.001	[0.007]	−0.001	[0.007]	0.010	[0.006]
Computer (general)			−0.003	[0.007]	−0.003	[0.007]	0.001	[0.006]
Female=1 # Cognitive							0.001	[0.004]
Female=1 # Social							0.003	[0.004]
Female=1 # Character							-0.019^{***}	[0.004]
Female=1 # Writing/language							-0.012^{*}	[0.005]
Female=1 # Customer Service							-0.028^{***}	[0.005]
Female=1 # Management							-0.027^{***}	[0.005]
Female=1 # Financial							-0.018^{**}	[0.007]
Female=1 # Computer (general)							−0.008	[0.007]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs	X		X		X		X	
Observations	1011707		1011707		1011707		1011707	
R^2	0.619		0.617		0.619		0.619	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: 3-digit stock data: Regressions results with combined cognitive and social

	(1) Model 1A		(2) Model 1B		(3) Model 1C		(4) Model 1D	
Female=1	−0.158***	[0.005]			−0.134***	[0.005]	−0.101***	[0.010]
Cognitive			0.096***	[0.025]	0.105***	[0.025]	0.091**	[0.030]
Social			−0.045***	[0.013]	−0.017	[0.013]	−0.014	[0.016]
Cognitive & Social			0.040	[0.031]	0.020	[0.030]	0.019	[0.034]
Character			0.044***	[0.012]	0.040***	[0.012]	0.061***	[0.014]
Writing/language			0.101***	[0.016]	0.079***	[0.015]	0.080***	[0.019]
Customer Service			0.112***	[0.017]	0.083***	[0.016]	0.103***	[0.020]
Management			0.128***	[0.018]	0.110***	[0.018]	0.123***	[0.021]
Financial			0.089***	[0.023]	0.088***	[0.023]	0.110***	[0.031]
Computer (general)			0.132***	[0.016]	0.093***	[0.016]	0.100***	[0.018]
Female=1 # Cognitive							0.032	[0.023]
Female=1 # Social							−0.007	[0.012]
Female=1 # Cognitive & Social							−0.002	[0.028]
Female=1 # Character							−0.043***	[0.011]
Female=1 # Writing/language							0.000	[0.016]
Female=1 # Customer Service							−0.046**	[0.016]
Female=1 # Management							−0.026	[0.016]
Female=1 # Financial							−0.041	[0.026]
Female=1 # Computer (general)							−0.015	[0.017]
Base controls	X		X		X		X	
Occupation FEs								
Firm controls								
Firm FEs								
Observations	10313978		10313978		10313978		10313978	
R^2	0.370		0.386		0.417		0.418	

Table 19 continued

	(5) Model 2A		(6) Model 2B		(7) Model 2C		(8) Model 2D	
Female=1	−0.091***	[0.003]			−0.090***	[0.003]	−0.050***	[0.007]
Cognitive			0.004	[0.011]	0.007	[0.010]	0.012	[0.013]
Social			−0.023***	[0.006]	−0.021***	[0.006]	−0.015*	[0.007]
Cognitive & Social			0.006	[0.016]	0.005	[0.015]	−0.001	[0.019]
Character			0.017**	[0.006]	0.017**	[0.006]	0.031***	[0.007]
Writing/language			0.057***	[0.008]	0.053***	[0.008]	0.063***	[0.010]
Customer Service			0.040***	[0.009]	0.034***	[0.008]	0.057***	[0.011]
Management			−0.005	[0.008]	−0.004	[0.008]	0.000	[0.009]
Financial			0.028**	[0.009]	0.026**	[0.009]	0.051***	[0.011]
Computer (general)			0.029**	[0.010]	0.027**	[0.010]	0.034**	[0.011]
Female=1 # Cognitive							−0.012	[0.014]
Female=1 # Social							−0.014	[0.008]
Female=1 # Cognitive & Social							0.012	[0.019]
Female=1 # Character							−0.029***	[0.007]
Female=1 # Writing/language							−0.022	[0.012]
Female=1 # Customer Service							−0.055***	[0.012]
Female=1 # Management							−0.010	[0.008]
Female=1 # Financial							−0.051***	[0.011]
Female=1 # Computer (general)							−0.017	[0.012]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls								
Firm FEs								
Observations	10313978		10313978		10313978		10313978	
R^2	0.537		0.529		0.540		0.541	

Table 19 continued

	(9)		(10)		(11)		(12)	
	Model 3A		Model 3B		Model 3C		Model 3D	
Female=1	−0.083***	[0.003]			−0.083***	[0.003]	−0.042***	[0.006]
Cognitive			0.001	[0.011]	0.003	[0.010]	0.012	[0.012]
Social			−0.011*	[0.005]	−0.010	[0.005]	−0.009	[0.006]
Cognitive & Social			−0.001	[0.014]	−0.001	[0.014]	−0.009	[0.018]
Character			0.002	[0.005]	0.003	[0.005]	0.019**	[0.006]
Writing/language			0.026***	[0.006]	0.025***	[0.006]	0.041***	[0.008]
Customer Service			−0.002	[0.007]	−0.006	[0.007]	0.023*	[0.009]
Management			0.004	[0.007]	0.005	[0.007]	0.011	[0.008]
Financial			0.020*	[0.009]	0.017*	[0.008]	0.038***	[0.010]
Computer (general)			0.012	[0.008]	0.010	[0.008]	0.019*	[0.009]
Female=1 # Cognitive							−0.019	[0.013]
Female=1 # Social							−0.004	[0.007]
Female=1 # Cognitive & Social							0.016	[0.019]
Female=1 # Character							−0.033***	[0.006]
Female=1 # Writing/language							−0.037***	[0.010]
Female=1 # Customer Service							−0.070***	[0.011]
Female=1 # Management							−0.014	[0.007]
Female=1 # Financial							−0.042***	[0.010]
Female=1 # Computer (general)							−0.023*	[0.010]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs								
Observations	10309784		10309784		10309784		10309784	
R^2	0.558		0.550		0.559		0.560	

Table 19 continued

	(13)		(14)		(15)		(16)	
	Model 4A		Model 4B		Model 4C		Model 4D	
Female=1	−0.079***	[0.003]			−0.079***	[0.003]	−0.034***	[0.006]
Cognitive			−0.019	[0.013]	−0.015	[0.012]	−0.002	[0.013]
Social			−0.012*	[0.005]	−0.011*	[0.005]	−0.010	[0.007]
Cognitive & Social			0.022	[0.016]	0.020	[0.015]	0.007	[0.017]
Character			0.015**	[0.006]	0.018**	[0.005]	0.040***	[0.007]
Writing/language			0.006	[0.006]	0.006	[0.006]	0.026**	[0.008]
Customer Service			−0.008	[0.008]	−0.009	[0.008]	0.020	[0.011]
Management			0.005	[0.008]	0.006	[0.007]	0.012	[0.008]
Financial			0.022**	[0.007]	0.019**	[0.007]	0.030**	[0.009]
Computer (general)			0.002	[0.007]	0.004	[0.007]	0.020**	[0.008]
Female=1 # Cognitive							−0.028*	[0.014]
Female=1 # Social							−0.002	[0.007]
Female=1 # Cognitive & Social							0.024	[0.019]
Female=1 # Character							−0.044***	[0.006]
Female=1 # Writing/language							−0.042***	[0.011]
Female=1 # Customer Service							−0.065***	[0.012]
Female=1 # Management							−0.014*	[0.007]
Female=1 # Financial							−0.020	[0.010]
Female=1 # Computer (general)							−0.038***	[0.009]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs	X		X		X		X	
Observations	10309784		10309784		10309784		10309784	
R^2	0.592		0.584		0.592		0.593	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: 3-digit flow data: Regressions results with combined cognitive and social

	(1) Model 1A		(2) Model 1B		(3) Model 1C		(4) Model 1D	
Female=1	−0.129***	[0.002]			−0.114***	[0.002]	−0.094***	[0.004]
Cognitive			0.096***	[0.008]	0.095***	[0.008]	0.067***	[0.010]
Social			−0.029***	[0.006]	−0.017**	[0.005]	−0.021**	[0.006]
Cognitive & Social			−0.020	[0.017]	−0.024	[0.017]	−0.002	[0.024]
Character			0.025***	[0.005]	0.021***	[0.005]	0.041***	[0.006]
Writing/language			0.063***	[0.006]	0.051***	[0.006]	0.053***	[0.008]
Customer Service			0.047***	[0.011]	0.025*	[0.010]	0.040***	[0.009]
Management			0.132***	[0.007]	0.123***	[0.006]	0.132***	[0.008]
Financial			0.088***	[0.008]	0.086***	[0.008]	0.113***	[0.010]
Computer (general)			0.110***	[0.009]	0.089***	[0.009]	0.099***	[0.009]
Female=1 # Cognitive							0.053***	[0.009]
Female=1 # Social							0.006	[0.005]
Female=1 # Cognitive & Social							−0.043*	[0.020]
Female=1 # Character							−0.035***	[0.005]
Female=1 # Writing/language							−0.002	[0.007]
Female=1 # Customer Service							−0.033***	[0.006]
Female=1 # Management							−0.018**	[0.007]
Female=1 # Financial							−0.050***	[0.007]
Female=1 # Computer (general)							−0.023**	[0.008]
Base controls	X		X		X		X	
Occupation FEs								
Firm controls								
Firm FEs								
Observations	1202198		1202198		1202198		1202198	
R^2	0.436		0.446		0.463		0.464	

Table 20 continued

	(5) Model 2A		(6) Model 2B		(7) Model 2C		(8) Model 2D	
Female=1	−0.072***	[0.001]			−0.072***	[0.001]	−0.054***	[0.004]
Cognitive			0.010	[0.009]	0.012	[0.009]	0.007	[0.010]
Social			−0.017***	[0.004]	−0.017***	[0.004]	−0.013**	[0.005]
Cognitive & Social			−0.008	[0.012]	−0.007	[0.012]	−0.003	[0.014]
Character			0.008*	[0.003]	0.008*	[0.003]	0.015***	[0.004]
Writing/language			0.030***	[0.004]	0.030***	[0.004]	0.035***	[0.005]
Customer Service			0.021***	[0.005]	0.019***	[0.005]	0.031***	[0.006]
Management			0.009*	[0.004]	0.009*	[0.004]	0.016**	[0.005]
Financial			0.025***	[0.005]	0.024***	[0.005]	0.043***	[0.006]
Computer (general)			0.035***	[0.008]	0.034***	[0.008]	0.033***	[0.007]
Female=1 # Cognitive							0.009	[0.008]
Female=1 # Social							−0.006	[0.004]
Female=1 # Cognitive & Social							−0.007	[0.012]
Female=1 # Character							−0.013***	[0.004]
Female=1 # Writing/language							−0.011*	[0.005]
Female=1 # Customer Service							−0.027***	[0.005]
Female=1 # Management							−0.014**	[0.005]
Female=1 # Financial							−0.036***	[0.006]
Female=1 # Computer (general)							0.004	[0.007]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls								
Firm FEs								
Observations	1202198		1202198		1202198		1202198	
R^2	0.552		0.548		0.553		0.554	

Table 20 continued

	(9)		(10)		(11)		(12)	
	Model 3A		Model 3B		Model 3C		Model 3D	
Female=1	−0.066***	[0.001]			−0.066***	[0.001]	−0.051***	[0.003]
Cognitive			0.001	[0.007]	0.003	[0.006]	−0.001	[0.007]
Social			−0.011**	[0.004]	−0.011**	[0.004]	−0.012**	[0.004]
Cognitive & Social			−0.001	[0.008]	−0.000	[0.008]	0.003	[0.010]
Character			−0.000	[0.003]	0.000	[0.003]	0.009*	[0.004]
Writing/language			0.008*	[0.004]	0.008*	[0.004]	0.016***	[0.005]
Customer Service			−0.008	[0.005]	−0.009	[0.005]	0.003	[0.005]
Management			0.013**	[0.004]	0.012**	[0.004]	0.022***	[0.005]
Financial			0.020***	[0.005]	0.019***	[0.005]	0.032***	[0.006]
Computer (general)			0.016**	[0.006]	0.015**	[0.006]	0.018**	[0.006]
Female=1 # Cognitive							0.008	[0.007]
Female=1 # Social							0.002	[0.004]
Female=1 # Cognitive & Social							−0.007	[0.010]
Female=1 # Character							−0.015***	[0.003]
Female=1 # Writing/language							−0.016***	[0.004]
Female=1 # Customer Service							−0.025***	[0.005]
Female=1 # Management							−0.020***	[0.005]
Female=1 # Financial							−0.025***	[0.006]
Female=1 # Computer (general)							−0.005	[0.006]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs								
Observations	1201910		1201910		1201910		1201910	
R^2	0.568		0.564		0.569		0.569	

Table 20 continued

	(13)		(14)		(15)		(16)	
	Model 4A		Model 4B		Model 4C		Model 4D	
Female=1	−0.064***	[0.001]			−0.064***	[0.001]	−0.052***	[0.003]
Cognitive			−0.016**	[0.006]	−0.015**	[0.006]	−0.018**	[0.007]
Social			−0.010*	[0.004]	−0.011*	[0.004]	−0.013**	[0.004]
Cognitive & Social			0.018*	[0.008]	0.019*	[0.007]	0.018*	[0.009]
Character			0.004	[0.003]	0.004	[0.003]	0.013***	[0.004]
Writing/language			−0.001	[0.004]	−0.000	[0.004]	0.008*	[0.004]
Customer Service			−0.013*	[0.006]	−0.012*	[0.006]	−0.004	[0.006]
Management			−0.003	[0.004]	−0.003	[0.004]	0.006	[0.004]
Financial			0.009	[0.005]	0.008	[0.005]	0.017**	[0.006]
Computer (general)			0.011	[0.006]	0.011	[0.006]	0.018***	[0.005]
Female=1 # Cognitive							0.006	[0.006]
Female=1 # Social							0.003	[0.004]
Female=1 # Cognitive & Social							0.001	[0.009]
Female=1 # Character							−0.015***	[0.003]
Female=1 # Writing/language							−0.015***	[0.004]
Female=1 # Customer Service							−0.018***	[0.005]
Female=1 # Management							−0.016***	[0.004]
Female=1 # Financial							−0.016**	[0.005]
Female=1 # Computer (general)							−0.014*	[0.006]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs	X		X		X		X	
Observations	1201910		1201910		1201910		1201910	
R^2	0.593		0.589		0.593		0.594	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets
** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Table 21: 4-digit stock data: Regression results with combined cognitive and social

	(1) Model 1A		(2) Model 1B		(3) Model 1C		(4) Model 1D	
Female=1	−0.153***	[0.005]			−0.130***	[0.006]	−0.118***	[0.013]
Cognitive			0.102***	[0.025]	0.106***	[0.024]	0.080**	[0.027]
Social			−0.043***	[0.013]	−0.024	[0.013]	−0.028	[0.018]
Cognitive & Social			0.051	[0.032]	0.034	[0.032]	0.048	[0.040]
Character			0.031*	[0.012]	0.025*	[0.012]	0.041**	[0.016]
Writing/language			0.071***	[0.019]	0.053**	[0.018]	0.033	[0.023]
Customer Service			0.127***	[0.020]	0.095***	[0.020]	0.119***	[0.023]
Management			0.146***	[0.020]	0.127***	[0.021]	0.144***	[0.025]
Financial			0.088***	[0.024]	0.095***	[0.022]	0.130***	[0.026]
Computer (general)			0.068*	[0.033]	0.030	[0.034]	0.015	[0.042]
Female=1 # Cognitive							0.044*	[0.020]
Female=1 # Social							0.009	[0.018]
Female=1 # Cognitive & Social							−0.024	[0.031]
Female=1 # Character							−0.032*	[0.015]
Female=1 # Writing/language							0.045*	[0.019]
Female=1 # Customer Service							−0.059***	[0.017]
Female=1 # Management							−0.042*	[0.018]
Female=1 # Financial							−0.065***	[0.019]
Female=1 # Computer (general)							0.049	[0.032]
Base controls	X		X		X		X	
Occupation FEs								
Firm controls								
Firm FEs								
Observations	9010127		9010127		9010127		9010127	
R^2	0.378		0.386		0.416		0.417	

Table 21 continued

	(5) Model 2A		(6) Model 2B		(7) Model 2C		(8) Model 2D	
Female=1	−0.071***	[0.003]			−0.071***	[0.003]	−0.041***	[0.006]
Cognitive			0.005	[0.009]	0.006	[0.009]	0.006	[0.011]
Social			−0.010	[0.006]	−0.009	[0.006]	−0.006	[0.007]
Cognitive & Social			−0.003	[0.014]	−0.003	[0.014]	0.003	[0.017]
Character			0.017**	[0.006]	0.016**	[0.006]	0.030***	[0.007]
Writing/language			0.036***	[0.007]	0.036***	[0.007]	0.039***	[0.007]
Customer Service			0.017	[0.009]	0.015	[0.009]	0.042***	[0.010]
Management			−0.003	[0.008]	−0.002	[0.008]	0.004	[0.008]
Financial			0.012	[0.007]	0.013	[0.007]	0.031***	[0.009]
Computer (general)			0.013	[0.007]	0.012	[0.007]	0.012	[0.008]
Female=1 # Cognitive							−0.002	[0.011]
Female=1 # Social							−0.006	[0.006]
Female=1 # Cognitive & Social							−0.014	[0.016]
Female=1 # Character							−0.026***	[0.006]
Female=1 # Writing/language							−0.007	[0.007]
Female=1 # Customer Service							−0.066***	[0.009]
Female=1 # Management							−0.013	[0.007]
Female=1 # Financial							−0.034***	[0.009]
Female=1 # Computer (general)							0.000	[0.013]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls								
Firm FEs								
Observations	9010127		9010127		9010127		9010127	
R^2	0.573		0.568		0.574		0.575	

Table 21 continued

	(9) Model 3A		(10) Model 3B		(11) Model 3C		(12) Model 3D	
Female=1	−0.066***	[0.002]			−0.066***	[0.002]	−0.034***	[0.005]
Cognitive			0.004	[0.008]	0.005	[0.008]	0.007	[0.010]
Social			−0.005	[0.005]	−0.005	[0.005]	−0.004	[0.006]
Cognitive & Social			−0.005	[0.012]	−0.005	[0.012]	0.002	[0.015]
Character			0.009	[0.005]	0.008	[0.005]	0.023***	[0.006]
Writing/language			0.017**	[0.006]	0.017**	[0.006]	0.025***	[0.007]
Customer Service			0.001	[0.008]	−0.001	[0.008]	0.030***	[0.009]
Management			−0.000	[0.007]	0.000	[0.007]	0.008	[0.007]
Financial			0.006	[0.006]	0.007	[0.006]	0.020*	[0.008]
Computer (general)			−0.000	[0.006]	−0.001	[0.006]	0.003	[0.008]
Female=1 # Cognitive							−0.005	[0.010]
Female=1 # Social							−0.002	[0.006]
Female=1 # Cognitive & Social							−0.015	[0.014]
Female=1 # Character							−0.028***	[0.005]
Female=1 # Writing/language							−0.017**	[0.006]
Female=1 # Customer Service							−0.076***	[0.009]
Female=1 # Management							−0.018**	[0.007]
Female=1 # Financial							−0.024**	[0.009]
Female=1 # Computer (general)							−0.007	[0.013]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs								
Observations	9006687		9006687		9006687		9006687	
R^2	0.586		0.581		0.587		0.588	

Table 21 continued

	(13) Model 4A		(14) Model 4B		(15) Model 4C		(16) Model 4D	
Female=1	−0.063***	[0.002]			−0.062***	[0.002]	−0.031***	[0.005]
Cognitive			−0.007	[0.008]	−0.007	[0.007]	−0.002	[0.010]
Social			−0.004	[0.004]	−0.004	[0.004]	−0.005	[0.005]
Cognitive & Social			0.001	[0.012]	0.001	[0.012]	0.006	[0.015]
Character			0.019***	[0.005]	0.019***	[0.005]	0.039***	[0.006]
Writing/language			0.005	[0.005]	0.005	[0.005]	0.013*	[0.006]
Customer Service			−0.016*	[0.007]	−0.016*	[0.007]	0.016	[0.008]
Management			−0.010	[0.008]	−0.010	[0.007]	0.000	[0.007]
Financial			0.012	[0.006]	0.012	[0.006]	0.022**	[0.008]
Computer (general)			0.002	[0.007]	0.003	[0.006]	0.012	[0.010]
Female=1 # Cognitive							−0.011	[0.009]
Female=1 # Social							0.002	[0.005]
Female=1 # Cognitive & Social							−0.010	[0.014]
Female=1 # Character							−0.034***	[0.005]
Female=1 # Writing/language							−0.017*	[0.007]
Female=1 # Customer Service							−0.069***	[0.009]
Female=1 # Management							−0.020**	[0.006]
Female=1 # Financial							−0.017*	[0.009]
Female=1 # Computer (general)							−0.020	[0.015]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs	X		X		X		X	
Observations	9006687		9006687		9006687		9006687	
R^2	0.614		0.610		0.615		0.615	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 22: 4-digit flow data: Regression results with combined cognitive and social

	(1)		(2)		(3)		(4)	
	Model 1A		Model 1B		Model 1C		Model 1D	
Female=1	−0.117***	[0.002]			−0.107***	[0.002]	−0.075***	[0.005]
Cognitive			0.151***	[0.011]	0.147***	[0.010]	0.119***	[0.013]
Social			−0.011	[0.006]	−0.003	[0.006]	−0.003	[0.007]
Cognitive & Social			−0.059**	[0.019]	−0.060**	[0.019]	−0.052	[0.028]
Character			0.009	[0.006]	0.006	[0.006]	0.030***	[0.007]
Writing/language			0.041***	[0.008]	0.032***	[0.007]	0.035***	[0.010]
Customer Service			0.030**	[0.011]	0.010	[0.010]	0.032**	[0.010]
Management			0.139***	[0.008]	0.129***	[0.007]	0.153***	[0.009]
Financial			0.051***	[0.010]	0.051***	[0.009]	0.093***	[0.012]
Computer (general)			0.082***	[0.009]	0.063***	[0.009]	0.089***	[0.010]
Female=1 # Cognitive							0.050***	[0.012]
Female=1 # Social							−0.001	[0.006]
Female=1 # Cognitive & Social							−0.014	[0.024]
Female=1 # Character							−0.041***	[0.005]
Female=1 # Writing/language							−0.005	[0.008]
Female=1 # Customer Service							−0.045***	[0.006]
Female=1 # Management							−0.046***	[0.008]
Female=1 # Financial							−0.072***	[0.008]
Female=1 # Computer (general)							−0.054***	[0.010]
Base controls	X		X		X		X	
Occupation FEs								
Firm controls								
Firm FEs								
Observations	1011924		1011924		1011924		1011924	
R^2	0.436		0.442		0.458		0.460	

Table 22 continued

	(5)		(6)		(7)		(8)	
	Model 2A		Model 2B		Model 2C		Model 2D	
Female=1	−0.055***	[0.001]			−0.055***	[0.001]	−0.034***	[0.004]
Cognitive			0.008	[0.008]	0.008	[0.008]	0.010	[0.010]
Social			−0.014**	[0.004]	−0.014**	[0.004]	−0.011*	[0.005]
Cognitive & Social			−0.015	[0.010]	−0.013	[0.009]	−0.014	[0.012]
Character			0.007	[0.004]	0.007	[0.004]	0.019***	[0.004]
Writing/language			0.020***	[0.004]	0.020***	[0.004]	0.024***	[0.005]
Customer Service			0.016**	[0.006]	0.015*	[0.006]	0.032***	[0.007]
Management			0.006	[0.005]	0.006	[0.005]	0.018**	[0.005]
Financial			0.009	[0.006]	0.009	[0.006]	0.029***	[0.007]
Computer (general)			0.009	[0.007]	0.008	[0.007]	0.008	[0.007]
Female=1 # Cognitive							−0.003	[0.009]
Female=1 # Social							−0.005	[0.005]
Female=1 # Cognitive & Social							0.001	[0.012]
Female=1 # Character							−0.020***	[0.004]
Female=1 # Writing/language							−0.007	[0.005]
Female=1 # Customer Service							−0.036***	[0.005]
Female=1 # Management							−0.021***	[0.006]
Female=1 # Financial							−0.035***	[0.007]
Female=1 # Computer (general)							0.002	[0.008]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls								
Firm FEs								
Observations	1011924		1011924		1011924		1011924	
R^2	0.583		0.580		0.584		0.584	

Table 22 continued

	(9) Model 3A		(10) Model 3B		(11) Model 3C		(12) Model 3D	
Female=1	−0.051***	[0.001]			−0.051***	[0.001]	−0.030***	[0.004]
Cognitive			0.002	[0.007]	0.002	[0.007]	0.004	[0.009]
Social			−0.010*	[0.004]	−0.011**	[0.004]	−0.010*	[0.005]
Cognitive & Social			−0.010	[0.009]	−0.009	[0.009]	−0.010	[0.011]
Character			0.001	[0.004]	0.001	[0.004]	0.014***	[0.004]
Writing/language			0.002	[0.005]	0.003	[0.005]	0.009	[0.005]
Customer Service			0.000	[0.006]	−0.000	[0.006]	0.017**	[0.006]
Management			0.006	[0.005]	0.005	[0.005]	0.021***	[0.005]
Financial			0.011	[0.007]	0.011	[0.007]	0.028***	[0.007]
Computer (general)			0.002	[0.007]	0.001	[0.007]	0.002	[0.006]
Female=1 # Cognitive							−0.004	[0.008]
Female=1 # Social							−0.001	[0.005]
Female=1 # Cognitive & Social							0.003	[0.011]
Female=1 # Character							−0.021***	[0.004]
Female=1 # Writing/language							−0.011*	[0.006]
Female=1 # Customer Service							−0.035***	[0.005]
Female=1 # Management							−0.029***	[0.006]
Female=1 # Financial							−0.029***	[0.008]
Female=1 # Computer (general)							−0.001	[0.008]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs								
Observations	1011707		1011707		1011707		1011707	
R^2	0.595		0.592		0.595		0.596	

Table 22 continued

	(13) Model 4A		(14) Model 4B		(15) Model 4C		(16) Model 4D	
Female=1	−0.048***	[0.001]			−0.048***	[0.001]	−0.033***	[0.004]
Cognitive			−0.015*	[0.006]	−0.015*	[0.006]	−0.013	[0.007]
Social			−0.015**	[0.005]	−0.015**	[0.005]	−0.017***	[0.005]
Cognitive & Social			0.005	[0.009]	0.007	[0.009]	0.003	[0.009]
Character			0.003	[0.004]	0.004	[0.004]	0.015***	[0.004]
Writing/language			−0.003	[0.005]	−0.003	[0.005]	0.004	[0.005]
Customer Service			−0.009	[0.007]	−0.009	[0.007]	0.006	[0.007]
Management			−0.005	[0.005]	−0.005	[0.005]	0.011*	[0.005]
Financial			−0.001	[0.007]	−0.001	[0.007]	0.010	[0.006]
Computer (general)			−0.003	[0.007]	−0.003	[0.007]	0.001	[0.006]
Female=1 # Cognitive							−0.003	[0.007]
Female=1 # Social							0.003	[0.005]
Female=1 # Cognitive & Social							0.007	[0.009]
Female=1 # Character							−0.019***	[0.004]
Female=1 # Writing/language							−0.012*	[0.006]
Female=1 # Customer Service							−0.028***	[0.005]
Female=1 # Management							−0.027***	[0.005]
Female=1 # Financial							−0.018**	[0.007]
Female=1 # Computer (general)							−0.008	[0.007]
Base controls	X		X		X		X	
Occupation FEs	X		X		X		X	
Firm controls	X		X		X		X	
Firm FEs	X		X		X		X	
Observations	1011707		1011707		1011707		1011707	
R^2	0.619		0.617		0.619		0.619	

Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets
** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Deming & Kahn's (2018) Table 3

Table 3
Average Wages and Skill Requirements

	Dependent Variable: Log(Mean Wages) in MSA-Occupation Cells					
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive	.113***	−.413***	.245***	.181***	.0792***	.0465***
	(.00908)	(.0166)	(.00784)	(.0139)	(.00873)	(.0122)
Social	.429***	−.0919***	.301***	.236***	.0517***	.0202
	(.0155)	(.0206)	(.0121)	(.0167)	(.00966)	(.0127)
Both required		1.319***		.157***		.0760***
		(.0349)		(.0278)		(.0198)
Years of education	.131***	.129***	.0764***	.0765***	.00865***	.00873***
	(.000770)	(.000763)	(.000844)	(.000844)	(.000995)	(.000995)
Years of experience	.160***	.161***	.0848***	.0849***	.0318***	.0318***
	(.00120)	(.00118)	(.00120)	(.00120)	(.00102)	(.00102)
Base controls			X	X		
Detailed controls					X	X
F-statistic (cognitive and social)	553.1	855.0	1,004	680.4	69.66	51.35
F-statistic (all 10 skills)	1,874	2,054	612.6	560.1	59.93	55.83
MSA-occupation cells	56,611	56,611	56,611	56,611	56,611	56,611
R^2	.702	.710	.846	.846	.940	.941

NOTE.—All regressions control for the share of ads with each of the eight other job skill, education, and experience requirements. Years of education and experience equal 0 if the MSA-occupation cell has no ads that specify requirements. The dependent variable is the log of median hourly earnings in the MSA-occupation cell, obtained from Occupational Employment Statistics data. Base controls include metropolitan statistical area (MSA) characteristics from the American Community Survey, four-digit Standard Occupational Classification (SOC) occupation fixed effects, and the share of ads in the MSA-occupation cell that are in each of the two-digit North American Industry Classification System industries. Detailed controls include MSA and six-digit SOC occupation fixed effects and the industry shares. Observations are from the full sample, weighted by the number of ads in the MSA-occupation cell. See table 1 for skills definitions.

*** $p < .01$.

Bibliography

Acemoglu, Daron and David H. Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." Pp. 1043–1171 in *Handbook of Labor Economics*, vol. 4. Elsevier Inc.

Acker, Joan. 1990. "Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations." *Gender & Society* 4(2):139–58.

Akerlof, George A. and Rachel E. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of Economics* CXV(3):715–53.

Akerlof, George A. and Rachel E. Kranton. 2005. "Identity and the Economics of Organizations." *Journal of Economic Perspectives* 19(1):9–32.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, N.J.: Princeton University Press.

Arulampalam, Wiji, Alison L. Booth, and Mark L. Bryan. 2007. "Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution." *ILR Review* 60(2):163–86.

Atalay, Enghin, Phai Phongthiengtham, Sebastian Sotelo, and Daniel Tannenbaum. 2018a. "New Technologies and the Labor Market." *Journal of Monetary Economics*.

Atalay, Enghin, Phai Phongthiengtham, Sebastian Sotelo, and Daniel Tannenbaum. 2018b. "The Evolving U.S. Occupational Structure." *Working Paper* (March):1–86.

Autor, David H. and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103(5):1553–1597.

Autor, David H., Lawrence F. Katz, and Alan B. Krueger. 1998. "Computing Inequality: Have Computers Changed the Labor Market?" *The Quarterly Journal of Economics* 113(4):1169–1213.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118(4):1279–1333.

Azar, José A., Ioana Marinescu, Marshall I. Steinbaum, and Bledi Taska. 2018. "Concentration in US Labor Markets: Evidence from Online Vacancy Data." *NBER Working Paper Series* (24395).

Bacolod, Marigee P. and Bernardo S. Blum. 2010. "Two Sides of the Same Coin: U.S. 'Residual' Inequality and the Gender Gap." *The Journal of Human Resources* 45(1):197–242.

Beaudry, Paul, David A. Green, and Benjamin M. Sand. 2016. "The Great Reversal in the Demand for Skill and Cognitive Tasks." *Journal of Labor Economics* 34(1):S199–247.

Beaudry, Paul and Ethan Lewis. 2014. "Do Male-Female Wage Differentials Reflect Differences in the Return to Skill? Cross-City Evidence from 1980-2000." *American Economic Journal: Applied Economics* 6(2):178–94.

Bergmann, Barbara R. 1974. "Occupational Segregation, Wages and Profits When Employers Discriminate by Race or Sex." *Eastern Economic Journal* 1(2):103–10.

Berkes, Enrico, Paul Mohnen, and Bledi Taska. 2018. "The Effect of Skill Mismatch on Early Career Outcomes of College Graduates: Evidence from Online Job Postings." *Working Paper*.

Bizopoulou, Aspasia. 2017. "Task Profiles and Gender Wage-Gaps Within Occupations." *Working Paper*.

Black, Sandra E. and Alexandra Spitz-Oener. 2010. "Explaining Women's Success:

Technological Change and the Skill Content of Women's Work." *The Review of Economics and Statistics* 120(120):187–94.
Blau, Francine D. and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55(3):178–865.
Browne, Jude. 2013. "The Default Model: Gender Equality, Fatherhood, and Structural Constraint." *Politics & Gender* 9(2):152–73.
Budd, John W. 2011. *The Thought of Work*. Ithaca, NY: Cornell University Press.
Butler, Judith. 2006. *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
Cahuc, Pierre and André Zylberberg. 2004. *Labor Economics*. Cambridge, MA: MIT Press.
Cameron, A. Colin and Douglas L. Miller. 2015. "A Practitioner ' s Guide to Cluster-Robust Inference." *Journal of Human Resources* 50(2):317–72.
Cameron, Adrian Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.
Cerina, Fabio, Alessio Moro, and Michelle Rendall. 2017. "The Role of Gender in Employment Polarization." *Ssrn* (March):1–39.
Cockburn, Cynthia. 1988. "Workplace Relations and the Reproduction of Sex Segregation." in *Gender segregation at work*, edited by S. Walby. Philadelphia: Open University Press.
Cortes, Guido Matias, Nir Jaimovich, and Henry Siu. 2018. "The 'End of Men' and Rise of Women in the High-Skilled Labor Market."
Crompton, R. 2010. "Class and Employment." *Work, Employment & Society* 24(1):9–26.
Crompton, Rosemary. 2007. "Gender Inequality and the Gendered Division of Labour." Pp. 228–49 in *The Future of Gender*, edited by J. Browne. Cambridge: Cambridge University Press.
Deming, David. 2017. "The Growing Importance of Social Skills in the Labor Market." *Quarterly Journal of Economics* 132(4):1593–1640.
Deming, David and Lisa B. Kahn. 2018. "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals." *Journal of Labor Economics* 36(S1):S337–69.
Deming, David and Kadeem L. Noray. 2018. "STEM Careers and Technological Change." *NBER Working Paper Series* (25065).
Faulkner, Wendy. 2013. "Doing Gender in Engineering Workplace Cultures. II. Gender in/Authenticity and the in/Visibility Paradox." *Engineering Studies* 1(3):169–89.
Fedorets, Alexandra. 2014. "Closing the Gender Pay Gap and Individual Task Profiles: Women's Advantages from Technological Progress." *Working Paper* 1–33.
Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. *Decomposition Methods in Economics*. Elsevier Inc.
Frey, Carl Benedikt and Michael Osborne. 2013. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Oxford Martin Programme on Technology and Employment* (Work. Pap.).
Gini, Al. 1998. "Work, Identity and Self: How We Are Formed by The Work We Do." *Journal of Business Ethics* 17(7):707–14.
Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104(4):1091–1119.
Goos, Maarten and Alan Manning. 2007. "Lousy and Lovely Jobs: The Rising Polarization

of Work in Britain." *The Review of Economics and Statistics* 89(1):118–33.
Goos, Maarten, Alan Manning, and Anna Salomons. 2009. "Job Polarization in Europe." *American Economic Review* 99(2):59–63.
Goos, Maarten, Alan Manning, and Anna Salomons. 2014. "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring." *American Economic Review* 104(8):2509–2526.
Grinis, Inna. 2017. "The STEM Requirements of 'Non-STEM' Jobs: Evidence from UK Online Vacancy Postings and Implications for Skills & Knowledge Shortages." *Working Paper*.
Hakim, Catherine. 2000. *Work-Lifestyle Choices in the 21st Century : Preference Theory*. Oxford: Oxford University Press.
Hersch, Joni. 1998. "Compensating Differentials for Gender-Specific Job Injury Risks." *American Economic Review* 88(3):598–607.
Hershbein, Brad and Lisa B. Kahn. 2018. "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings." *American Economic Review* 108(7):1737–72.
In, Margaret Mooney and Mary C. Brinton. 1984. "Sex Typing in Occupational Socialization." Pp. 192–232 in *Sex Segregation in the Workplace*, edited by B. F. Reskin. Washington, DC: National Academy Press.
Jarman, Jennifer, Robert M. Blackburn, and Girts Racko. 2012. "The Dimensions of Occupational Gender Segregation in Industrial Countries." *Sociology* 46(6):1003–19.
Jensen, Mathias F. 2017. "Reconciling and Quantitative Methodology." *Women, Gender & Research* 1:8–17.
Kanter, Rosabeth Moss. 1977. "Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women." *American Journal of Sociology* 82(5):965–90.
Kettemann, Andreas, Andreas I. Mueller, and Josef Zweimüller. 2018. "Vacancy Durations and Entry Wages: Evidence From Linked Vacancy-Employer-Employee Data." *NBER Working Paper Series* (25118).
Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2018. *Children and Gender Inequality: Evidence from Denmark*.
Kuhn, Peter and Kailing Shen. 2013. "Gender Discrimination in Jobs Ads: Evidence from China." *The Quarterly Journal of Economics* 128(1):287–336.
Levanon, Asaf and David B. Grusky. 2016. "The Persistence of Extreme Gender Segregation in the Twenty-First Century." *American Journal of Sociology* 122(2):573–619.
Lindley, Joanne Kathryn. 2012. "The Gender Dimension of Technical Change and the Role of Task Inputs." *Labour Economics* 19(4):516–26.
Lindley, Joanne Kathryn. 2016. "Lousy Pay with Lousy Conditions: The Role of Occupational Desegregation in Explaining the UK Gender Pay and Work Intensity Gaps." *Oxford Economic Papers* 68(1):152–73.
Lindley, Joanne and Stephen Machin. 2012. "The Quest for More and More Education: Implications for Social Mobility." *Fiscal Studies* 33(2):265–86.
Machin, Stephen. 2001. "The Changing Nature of Labour Demand in the New Economy and Skill-Biased Technology Change." *Oxford Bulletin of Economics and Statistics* 63:753–76.

Machin, Stephen. 2011. “Changes in UK Wage Inequality Over the Last Forty Years.” Pp. 45–66 in *The Labour Market in Winter: The State of Working Britain*, edited by P. Gregg and J. Wadsworth. Oxford: Oxford University Press.

Machin, Stephen and John Van Reenen. 1998. “Technology and Changes in Skill Structure: Evidence from Seven OECD Countries.” *The Quarterly Journal of Economics* 113(4):1215–44.

Manning, Alan and Barbara Petrongolo. 2008. “The Part-Time Pay Penalty for Women in Britain.” *The Economic Journal* 118(526):28–51.

Marinescu, Ioana. 2017. “The General Equilibrium Impacts of Unemployment Insurance: Evidence from a Large Online Job Board.” *Journal of Public Economics* 150:14–29.

Marinescu, Ioana and Roland Rathelot. 2018. “Mismatch Unemployment and the Geography of Job Search.” *American Economic Journal: Macroeconomics* 10(3):42–70.

Marinescu, Ioana and Ronald Wolthoff. 2016. “Opening the Black Box of the Matching Function: The Power of Words.” *Working Paper*.

Martin, Patricia Yancey. 2013a. “Gendered Work Identities.” Pp. 322–24 in *Sociology of Work: An Encyclopedia*, edited by V. Smith. Thousand Oaks: SAGE Publications.

Martin, Patricia Yancey. 2013b. “Sociologists for Women in Society: A Feminist Bureaucracy?: SWS Presidential Address.” *Gender and Society* 27(3):281–93.

Marx, Karl. 1959. *Economic and Philosophic Manuscripts of 1844*. Moscow: Progress Publishers.

Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View*. New York: Harper & Row.

Modestino, Alicia Sasser, Daniel Shoag, and Joshua Ballance. 2016a. “Downskilling: Changes in Employer Skill Requirements over the Business Cycle.” *Labour Economics* 41:333–47.

Modestino, Alicia Sasser, Daniel Shoag, and Joshua Ballance. 2016b. “Upskilling: Do Employers Demand Greater Skill When Workers Are Plentiful?” *Working Paper*.

Ngai, L. Rachel and Barbara Petrongolo. 2017. “Gender Gaps and the Rise of the Service Economy.” *American Economic Journal: Macroeconomics* 9(4):1–44.

Olivetti, Claudia and Barbara Petrongolo. 2014. “Gender Gaps across Countries and Skills: Demand, Supply and the Industry Structure.” *Review of Economic Dynamics* 17(4):842–59.

Olivetti, Claudia and Barbara Petrongolo. 2016. “The Evolution of Gender Gaps in Industrialized Countries.” *Annual Review of Economics* (8):405–34.

Orzechowicz, D. and Patricia Yancey Martin. 2013. “Gendered Organizations.” Pp. 319–21 in *Sociology of Work: An Encyclopedia*, edited by V. Smith. Thousand Oaks: SAGE Publications.

Reid, Erin. 2013. “‘Doing Gender.’” Pp. 191–92 in *Sociology of Work: An Encyclopedia*, edited by V. Smith. Thousand Oaks: SAGE Publications, Inc.

Rendall, Michelle. 2010. “Brain Versus Brawn: The Realization of Women’s Comparative Advantage.” *SSRN Electronic Journal* (491):1–26.

Skorikov, Vladimir B. and Fred W. Vondracek. 2011. “Occupational Identity.” Pp. 693–714 in *Handbook of Identity Theory and Research*, edited by S. J. Schwartz, K. Luyckx, and V. L. Vignoles. New York, NY: Springer New York.

Stinebrickner, Todd R., Ralph Stinebrickner, and Paul J. Sullivan. 2018. “Job Tasks and the Gender Wage Gap among College Graduates.” *NBER Working Paper Series* (24790).

Tajfel, H. 1978. *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations.* edited by H. Tajfel. London: Academic Press.
Tajfel, H. and JC Turner. 1986. "The Social Identity Theory of Intergroup Behaviour." in *Psychology of intergroup relations*, edited by S. Worchel and W. Austin. Chicago, IL: Nelson Hall.
Tajfel, Henri and John C. Turner. 2004. "The Social Identity Theory of Intergroup Behavior." Pp. 276–93 in *Political psychology: Key readings*, edited by J. T. Jost and J. Sidanius. New York, NY: Psychology Press.
Turner, John C. 1982. "Towards a Cognitive Redefinition of the Social Group." Pp. 15–40 in *Social identity and intergroup relations*, edited by H. Tajfel. Cambridge: Cambridge University Press.
Weinberger, Catherine J. 2014. "The Increasing Complementarity between Cognitive and Social Skills." *The Review of Economics and Statistics* 96(5):849–61.
West, Candace and Don. H. Zimmerman. 1987. "Doing Gender." *Gender and Society* 1(2):125–51.
Yamaguchi, Shintaro. 2018. "Changes in Returns to Task-Specific Skills and Gender Wage Gap." *Journal of Human Resources* 53(1):32–70.
Zimbardo, Phillip G., W. C. Banks, Craig Haney, and David Jaffe. 1973. "The Mind Is a Formidable Jailer: A Pirandellian Prison." *The New York Times Magazine* 8 April:38–60.