# Econometric Analysis of the German Wage and Earnings Distribution

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen


vorgelegt von

Matthias Christian Seckler

aus Stuttgart



Tübingen

2019

:

| | |
|---|---|
| Tag der mündlichen Prüfung: | 9. Dezember 2019 |
| Dekan: | Professor Dr. rer. soc. Josef Schmid |
| Erstberichterstatter: | Professor Dr. rer. pol. Martin Biewen |
| Zweitberichterstatter: | Professor Dr. rer. pol. Joachim Grammig |

# Publication notes

Chapter 2 of this doctoral thesis has been published as Biewen, M., and M. Seckler (2019): Unions, Internationalization, Tasks, Firms, and Worker Characteristics: A Detailed Decomposition Analysis of Rising Wage Inequality in Germany. *The Journal of Economic Inequality*. Advance online publication. doi: https://doi.org/10.1007/s10888-019-09422-w.

Chapter 3 has been pre-printed as Seckler, M. (2019): Increasing Inequality in Lifetime Earnings: A Tale of Educational Upgrading and Changing Employment Patterns. *University of Tübingen Working Papers in Economics and Finance No. 119*.

Chapter 4 is based on Biewen, M., Fitzenberger, B., and M. Seckler (2019): Counterfactual Quantile Decompositions with Selection Correction Taking into Account Huber/Melly (2015): An Application to the German Gender Wage Gap. Unpublished manuscript, University of Tübingen and Humboldt-University Berlin.

# Funding

The research for this doctoral thesis was in parts funded through DFG priority program 1764 and DFG project BI767/3-1.

# Danksagung

Die vorliegende Dissertation ist über mehrere Jahre im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Statistik, Ökonometrie und Quantitative Methoden an der Eberhard Karls Universität Tübingen entstanden. Während dieser Zeit habe ich Unterstützung von vielen Seiten erfahren und möchte deshalb an dieser Stelle einigen Menschen danken, ohne die das Gelingen dieser Arbeit nicht vorstellbar gewesen wäre.

Zunächst Martin Biewen, der mir nach dem Abschluss meines Studiums überhaupt erst die Möglichkeit zur Promotion und damit zur Anfertigung dieser Arbeit geboten hat. Herzlichen Dank, Martin, nicht nur für dieses Vertrauen in mich, sondern insbesondere auch für Deine Unterstützung als Doktorvater und Koautor, Deinen Enthusiasmus und die kontinuierliche Förderung während der gesamten Promotionszeit.

Ebenso dankbar bin ich Joachim Grammig, der als Zweitbetreuer nicht nur durch sein Feedback und Fachwissen zu dieser Arbeit beigetragen hat, sondern auch darüber hinaus jederzeit als Ansprechpartner zur Verfügung stand. Vielen Dank, Joachim, für beides.

Während meines Promotionsvorhabens profitierte ich außerdem von der finanziellen Unterstützung sowie dem fachlichen Austausch im Rahmen des DFG Projekts „Accounting for Selection Effects in the Analysis of Wage Inequality in Germany“. In diesem Zusammenhang möchte ich insbesondere Bernd Fitzenberger und Jakob de Lazzer nicht unerwähnt lassen und beiden für die hervorragende Zusammenarbeit danken.

In den letzten Jahren hatte ich zudem die Möglichkeit, die Kapitel dieser Dissertation bei verschiedensten Konferenzen und Workshops vorzustellen. Den Teilnehmern des „6th/7th Internal-network Meeting of the DFG Priority Program 1764“, der „31st/32nd/33rd Annual Conference of the European Society for Population Economics“, des „32nd Annual Congress of the European Economic Association“, der „6th Annual Conference of the International

Association of Applied Econometrics“, der „Summer School on Topics and Methods in Labor Market Research“ (2018), des gemeinsamen Doktorandenseminars in Ökonometrie der Universitäten Hohenheim, Konstanz, Friedrichshafen und Tübingen (2016, 2017, 2018) sowie des „THE Christmas Workshops“ (2017) danke ich für spannende Diskussionen und Feedback.

Ebenfalls zu Dank verpflichtet bin ich meinen Kollegen an den Lehrstühlen für Statistik und Ökonometrie. Besonders hervorheben möchte ich dabei Lea Eiting und Sylvia Bürger, die bei verschiedensten administrativen Aufgaben eine unerlässliche Hilfe waren. Nicht weniger dankbar bin ich Gideon May, Stefanie Seifert, Madalina Thiele, Jakob Schwerter, Miriam Sturm, Thomas Dimpfl, Jantje Sönksen, Johannes Bleher, Dalia Elshiaty und Constantin Hanenberg für spannende Diskussionen, unzählige gemeinsame Mittags- und Kaffeepausen sowie eine unvergessliche Zeit.

Zu guter Letzt war auch der bedingungslose Rückhalt, welchen ich von Familie und Freunden erfahren habe, ein entscheidender Faktor beim Abschluss dieser Arbeit. Vielen Dank für Eure liebevolle Unterstützung, Euren Zuspruch sowie Eure Geduld, nicht nur während der Promotionszeit, sondern auch auf dem Weg dorthin. Dies gilt umso mehr für Sabine, die mich vom Beginn meines Promotionsvorhabens durch alle Höhen und Tiefen begleitet hat.

Matthias Seckler, Tübingen im Juli 2019

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Dissertation Introduction

*"Labor was the first price, the original purchase money that was paid for all things."*

Adam Smith

The compensation of labor is one of the most fundamental questions in economics. Therefore, it is not surprising that economists have also been interested in the distribution of labor earnings, and more generally income, for a long time, with descriptions of their distributions dating at least back to the late 17th century. At the time, the British statistician Gregory King published his *social tables*, a quantitative description of the social conditions in pre-industrialised England, which the author described as a *Scheme of the Income and Expenses of Several Families of England*.[1] Despite its long tradition, the study of income distributions had not been at the center of economic research for a long time, presumably due to its close relationship with normative questions (Sandmo, 2015). Against the background of a sharp rise in wage and income inequality in the U.S. and other countries around the globe in

[1]In an earlier, though less known work, similar calculations were already performed in the mid 17th century by William Petty. For a more detailed description, see, e.g., Holmes (1977) and Slack (2004).

recent decades, the question gained considerable attention and resulted in a rapidly growing literature (see Katz and Autor, 1999, and Acemoglu and Autor, 2011 for wages, and Piketty and Saez, 2014 for income). Simultaneously, a more general and controversial debate on the distribution of income and wealth has emerged among both policymakers and the general public.

For the analysis in this doctoral thesis, it is crucial to distinguish the different forms of income, most importantly labor income or earnings as opposed to capital income. The present work is solely dedicated to the first aspect and focuses on gross labor earnings, i.e. redistributive effects of the tax and transfer system are not part of the analysis. The second important distinction for the further reading is the one between the terms *wages* and *earnings*. Throughout this thesis, *wages* refer to the compensation received per hour worked, whereas *earnings* refer to the total compensation received over a longer time period, e.g. a day, year or up to a certain age. A more in-depth definition of the wage and earnings measures used is also provided in the different chapters. At this point, it is important to note that the present thesis is restricted to a rigorous econometric analysis of the German wage and earnings distribution as well as the ongoing distributional changes. Consequently, it does not attempt to answer any of the normative questions, e.g. on the socially optimal level of inequality, commonly raised in the public discussion.

In recent years, increasing wage and earnings inequality have been accompanied by declining labor income shares in many OECD countries including Germany. Nevertheless, the share of labor income, defined as the ratio of total labor compensation to GDP, still amounted to about 60 percent in Germany in 2016 and remained the by far most important form of income (OECD, 2018). At the same time, previous studies showed rising inequality in labor earnings to be the decisive factor in explaining increasing income inequality (see Daly and Valletta, 2006 for the U.S. and Biewen and Juhasz, 2012 for the German case). The literature further shows that the distribution of income has an impact on outcomes beyond individual consumption possibilities. For example, existing studies point towards a positive relationship between income and health (von Gaudecker and Scholz, 2007, Chetty et al., 2015). From a macro perspective, high levels of income inequality are potentially linked to weak economic growth (Ostry et al., 2014) as well as political polarization (Duca and Saving, 2016, Winkler, 2019). Although some of these studies consider different measures of income, e.g. also including non-labor income or measuring income at the household level, they underline the importance of wage and earnings distributions, given that labor earnings constitute the primary source of income for a major part of the population.

As this doctoral thesis exclusively uses data from Germany, it is insightful to start with a description of the most important trends in the German wage and earnings distribution as well as a number of special features related to the German labor market. Germany has traditionally been characterized by a rather stable wage distribution and an in international comparison relatively low level of wage inequality. In the 1980s, inequality started to rise but only at the at top of the distribution (Dustmann et al., 2009). Contrary to the U.S., the bottom half of the wage distribution remained stable during the 1980s as a likely consequence of a traditionally strong role of unions played in the wage setting process (Fitzenberger, 1999). Though delayed in international comparison, inequality in the lower half of the wage distribution equally started to increase in the early 1990s. Overall, wage inequality in Germany has increased sharply in all parts of the distribution in recent decades (Dustmann et al., 2009, Card et al., 2013), but seems to have stagnated or even moderately declined after 2010 (Felbermayr et al., 2016, Möller, 2016). Nevertheless, inequality levels in Germany remain significantly lower compared to the U.S. (Krueger et al., 2010). The latter equally observed a sharp and even more pronounced increase in the dispersion of wages starting already in the late 1970s in all parts of the distribution. Contrary to Germany, inequality in the upper part of the U.S. wage distribution has continued to increase to the day (e.g., Autor, 2014, Antonczyk et al., 2018 and, for recent numbers, Gould, 2019). At the same time, the pattern of distributional changes also differ between both countries. Changes in the U.S. during the 1990s were characterized by an increasing polarization in the sense that both the upper and the lower end of the distribution gained relative to the middle (Acemoglu and Autor, 2011). Despite some evidence in favor of job polarization, the literature on Germany does not find similar changes in the distribution of wages. Instead, these were characterized by a monotonic development with gains at higher and losses at lower quantiles. In this regard, they rather resembled the distributional changes in the U.S. during the 1980s (Goos et al., 2009, Dustmann et al., 2009).

Simultaneously with this development, the German labor market experienced a sharp increase in the dispersion of lifetime earnings, which nearly doubled between men born in the early 1960s and their parental generation and hence, exceeded the increase in cross-sectional wage or earnings inequality (Bönke et al., 2015a). A similar and even more dramatic development has also been described by Guevenen et al. (2017) for the U.S. Taking on a cohort perspective and looking at lifetime as opposed to cross-sectional earnings is very insightful from a theoretical perspective. This is due to the fact that cross-sectional wages reflect a single point-in-time observation within each individual's employment biography. However, from the perspective of consumption theory and under the assumption of perfect capital markets, individual consumption possibilities are rather determined by lifetime than short-term earnings as outlined in more detail by Corneo (2015). Against this background, the literature on lifetime earnings inequality describes another important dimension of distributional changes in the German wage and earnings distribution.

A distinguishing feature of the German labor market is a persistent West-East difference in the standard of living as a consequence of the German division. Despite major political efforts and some convergence between the formerly separated parts, significant differences in wages, earnings and unemployment rates remain to the day (Smolny, 2009, Schnabel, 2016). At the same time, East Germany equally faced a strong catch-up in terms of wage inequality. Starting from very low levels as a consequence of the egalitarian wage policy followed in the former German Democratic Republic, inequality levels (among men) roughly reached the level of West Germany by the turn of the millennium (Franz and Steiner, 2000).

Besides increasing inequality, the German labor market was characterized by very modest increases in real wages between the late 1990s and the outbreak of the financial crisis in 2008, also referred to as the *German wage moderation* (see, e.g., Bofinger, 2015). In its course, Germany even experienced a period of falling real median wages in the early 2000s as shown, for example, in Dustmann et al. (2014). In recent years, real wages have been increasing though, given the fast recovery and strong employment growth following the financial crisis, at a moderate pace (Deutsche Bundesbank, 2018). Moreover, the German labor market has been subject to profound institutional reforms (the so-called Hartz reforms) which, against the background of weak GDP growth and high unemployment rates, aimed at increasing Germany's international competitiveness via a flexibilization of the labor market. In fact, both factors are seen as a main reason for the dramatic reduction in unemployment since the early 2000s - a time when Germany was considered the *sick man of Europe* (Dustmann et al., 2014). The literature argues that the previously described wage moderation also contributed significantly to the exceptional robustness of the German labor market against the consequences of the financial crisis (Boysen-Hogrefe and Groll, 2010, Burda and Hunt, 2011). At the same time, both factors favored the emergence of a large low-wage sector comparable to Anglo-Saxon market economies such as the U.K. or Ireland (Kalina and Weinkopf, 2017). This development played an important role in the public debate on minimum wages which ultimately resulted in the introduction of a statutory minimum wage of 8.50 euro per hour in January 2015. In consequence of its recent introduction and the fact that both the administrative and official data used in the following studies are only available with considerable delay, the distributional effects of the minimum wage are not subject of the analysis in this thesis. However, recent work by Caliendo et al. (2017, 2018) provides first evidence for positive short-term wage effects at the bottom of the distribution, which were accompanied by minor employment losses only.

As it is common practice in the literature, the vast majority of the cited studies focuses on the wage and earnings distribution among men. An important reason for this is the significantly lower labor market participation rate among women, which, from the viewpoint of an econometrician, potentially causes non-random selection into the labor market and the estimation sample. For example, it is well-known that employment rates among women tend to increase with the level of education and equally vary over the life-cycle. Moreover, these selection patterns are also changing over time due to an increasing labor market participation among women of later birth cohorts, which inherently complicates any inter-temporal comparison. To underline the necessity to account for this type of selectivity, it is insightful to take a closer look at women's labor market participation in Germany. Traditionally, women are less likely to participate in the labor force compared to men not only in Germany, but around the globe. Despite a significant catching-up of women in recent decades, labor force participation rates continue to differ between both genders with overall numbers for 2013 amounting to 73% and 63%, respectively. At the same time, part-time employment rates among German women are high in international comparison. About half of all women employed work part-time, whereas the same holds true for only one out of nine men (Brenke, 2015). As a consequence, large differences in full-time employment rates persist until today. For this reason, two of the subsequent studies (chapters 2 and 3) on changes in the wage and lifetime earnings distribution are restricted to men only. The third study (chapter 4), on the other hand, explicitly addresses the effect of women's selection into full-time employment on the distributional gender pay gap.

From a statistical point of view, the availability of sophisticated decomposition methods is essential for an accurate description of differences in the distribution of wages and earnings.[2] For a long time, most studies were restricted to explaining differences in mean outcomes, building on the widely used decomposition method by Oaxaca (1973) and Blinder (1973). It has remained popular to the day, likely due to it's simplicity and easy interpretability. The technique assumes the conditional expectation $E(Y|X)$ to be linear in parameters, allowing an estimation by OLS. It equally enables a detailed decomposition, i.e. changes in mean

[2] I only outline some of the most important methods here. A more comprehensive discussion is provided in the literature, e.g. in Fortin et al. (2011).

outcomes can be linked to individual covariates and can be further split up into a part due to changes in the distribution of individual covariates (composition effect) as well another one due to changes in their returns (wage structure effect). Variance decomposition extends this idea by using the well-known statistical property of the variance being separable into a within and a between group component but does not provide a comparable detailed decomposition. Also, general inequality measures like the variance cannot explain changes in specific parts of a distribution which is often particularly relevant from a policy perspective (Fortin et al., 2011).

In the light of rising U.S. wage and earnings inequality and a stronger focus of the literature on distributional aspects, a number of new decomposition techniques have emerged during the last three decades that overcome some of the previous restrictions. Most importantly, more recent decomposition techniques (e.g. Juhn et al., 1993, DiNardo et al., 1996, Machado and Mata, 2005, Melly, 2005, Firpo et al., 2009, 2018, Chernozhukov et al., 2013) enable an analysis of inequality measures beyond the mean, including quantile differences. Chapters 2 and 4 of the present work provide two applications of the from my viewpoint still underused *Recentered-Influence-Function (RIF) decomposition* (Firpo et al., 2009, 2018). The method represents a generalization of the Oaxaca-Blinder decomposition allowing for a detailed decomposition, both in terms of composition and wage structure effects, of any distributional statistic of interest that does not depend on the ordering of factors. In this regard, the method is superior to the other commonly applied techniques which are restricted in different ways. The methods by Juhn et al. (1993), Machado and Mata (2005) and Melly (2005) do not allow for a detailed decomposition of distributional changes, which limits their usefulness for research questions intending to link observed changes to specific covariates of interest. This problem is overcome by the decomposition technique suggested in DiNardo et al. (1996). The authors use semi-parametric inverse probability weighting which allows to determine detailed effects by sequentially changing the distribution of individual covariates. As the major disadvantage relative to RIF decomposition, the results depend on the sequential ordering of the covariates included and the interpretation of results is only valid for a specific sequence. In principle, this problem can be moderated by calculating average effects for the different groups of covariates over all possible sequences. This idea is known as the Shapley decomposition (Shorrocks, 2013) and has, among others, been applied by Biewen and Plötze (2019) for a small number of groups. However, such an approach easily becomes computationally very costly as the number of groups increases. For example, the analysis in chapter 2 studies the impact of seven groups of covariates resulting in a total of 5040 potential sequences. Moreover, the DiNardo et al. (1996) method cannot be extended to a detailed decomposition of wage structure effects which are potentially a similar important source of distributional changes. Both arguments equally apply to the method suggested more recently in Chernozhukov et al. (2013). On the downside, RIF decomposition only yields a local approximation of the underlying effects given the linear specifications of the RIFs. This is potentially problematic in long-run distributional analyses where observed changes are typically large. To overcome this shortcoming, chapter 2 suggests a step-wise implementation which is aggregated to the overall long-run effect. To conclude this introduction, I want to give an overview of the three studies this doctoral thesis comprises.

*Unions, Internationalization, Tasks, Firms, and Worker Characteristics: A Detailed Decomposition Analysis of Rising Wage Inequality in Germany*

In chapter 2, we study changes in the wage distribution among German men. The main focus of the analysis is a quantitative assessment of competing factors associated with the rise in male wage inequality in Europe's largest economy over the period 1995-2010. The analysis uses a unified framework which considers an extensive set of explanatory factors including personal characteristics, measures of internationalization, task composition, union coverage, industry, region and firm characteristics. In contrast to most of the literature on Germany, the analysis relies on four waves of the German Structure of Earnings Surveys (GSES) which allows to focus on hourly wages (rather than daily earnings) which are not censored at the social contributions threshold. Another unique feature of the data is the inclusion of information on unionization at the individual as opposed to the firm-level information provided in all other data sets for Germany. These differences are carefully evaluated and we show that previous studies on German wage inequality most likely underestimated the role of de-unionization as important parts of de-unionization occur within establishments. As the second most important factor, we identify compositional effects of personal characteristics such as age and education. This finding is in line with the SBTC hypothesis in the sense that the increasing demand for higher skills was matched by a rising supply due to both educational upgrading as well as the aging of the workforce. At the same time, we find only moderate effects linked to internationalization, firm heterogeneity, task changes and regional convergence. The econometric analysis uses RIF Decomposition (Firpo et al., 2009, 2018) which is, following the arguments above, particularly well-suited to simultaneously control for a large number of covariates and to quantify their relative importance. Moreover, the chapter addresses some issues regarding the practical implementation of RIF decomposition and provides a supplementary analysis of unobserved firm heterogeneity.

*Increasing Inequality in Lifetime Earnings: A Tale of Educational Upgrading and Changing Employment Patterns*

Taking on a cohort perspective, chapter 3 aims at explaining the rising inequality in lifetime earnings in Germany by means of a detailed decomposition analysis. The study adds to a still comparably small but growing literature which documented rising lifetime earnings inequality for both Germany and the U.S. (Bönke et al., 2015a, Guevenen et al., 2017), but still lacks an in-depth analysis of the underlying causes. To address this blind spot, the study uses the high-quality administrative data of the *Sample of Integrated Employment Biographies (SIAB)* which are used to construct individual employment biographies. The empirical analysis is restricted to West German men born between the years 1955 and 1974 as the *SIAB* does not provide any information on East Germany prior to the year 1991. The analysis uses the concept of up-to-age X earnings (UAX) known from the literature and mostly focuses on earnings until the age of 40, which have been shown to be a good proxy for lifetime earnings (as well as changes in their distribution) and, at the same time, allow to include a reasonable number of birth cohorts. The findings suggest that significant parts of rising lifetime earnings inequality among West German men can be attributed to a lower labor market participation as a consequence of longer periods of both part-time and non-employment as well as the educational expansion among later cohorts. However, they can only partly explain losses at the bottom of the distribution and therefore suggest some similarities with the U.S., where the major part of earnings losses among later cohorts occurred conditional on working (Guevenen et al., 2017). The study also points towards potentially important changes in the penalty linked to employment interruptions but only finds a moderate impact of skill-biased technological change (SBTC) beyond educational upgrading. Finally, the analysis equally reveals that later cohorts did not only face an increase in inequality, but also a stagnation in earnings for a major part of their career. The latter trend is found not to be a direct consequence of a delayed labor market entry (due to educational upgrading), but is even stronger when looking at the development within education groups.

*Counterfactual Quantile Decompositions with Selection Correction Taking into Account Huber/Melly (2015): An Application to the German Gender Wage Gap*

Finally, chapter 4 proposes a novel technique for the estimation of counterfactual quantile decompositions with selection correction. In doing so, we directly build on previous contributions by Buchinsky (1998, 2001) as well as Albrecht et al. (2009). Using a control function approach, the Buchinsky method allows to control for both selection on observables and unobservables in the context of conditional quantile regression and is commonly used in applied research. A particularly important application is the Albrecht et al. (2009) decomposition method for the estimation of counterfactual distributions with selection correction, which builds itself on the decomposition by Machado and Mata (2005). A recent paper by Huber and Melly (2015) has criticized the use of the

Buchinsky correction by pointing out that it assumes a form of conditional independence that may be easily violated in applications. Therefore, we propose to transform the original quantile model in order to eliminate potential problems with the conditional independence assumption. We demonstrate theoretically and empirically that this transformation can make the Huber/Melly test pass when it rejects in the untransformed model. Applying the technique to the gender pay gap in Germany, we find evidence for a strong positive selection of women into full-time employment in line with previous evidence on other countries (see, e.g., Albrecht et al., 2009 for the Netherlands and Chzhen and Mumford, 2011 for the U.K.). The results also point towards important effects linked to unobservable selection which have not been documented in the previous literature on Germany and potentially bias estimates of the gender pay gap downwards. Moreover, we also show that unobserved selection may in fact be underestimated in applications where conditional independence is violated.

# Chapter 2

# Unions, Internationalization, Tasks, Firms, and Worker Characteristics: A Detailed Decomposition Analysis of Rising Wage Inequality in Germany*

## 2.1 Introduction

An extensive literature has documented a steady increase in wage inequality since the early 1980s in many countries around the world (see Katz and Autor, 1999, and Acemoglu and Autor, 2011 for surveys, and Dustmann et al., 2009, for the German case). Different explanations have been proposed for this trend. The most prominent explanation are changes

*This chapter is based on:
Biewen, M., and M. Seckler (2019): Unions, Internationalization, Tasks, Firms, and Worker Characteristics: A Detailed Decomposition Analysis of Rising Wage Inequality in Germany. *The Journal of Economic Inequality*. Advance online publication. doi: https://doi.org/10.1007/s10888-019-09422-w.

in demand and supply across skill groups connected to skill-biased technological change (Katz and Murphy, 1992, Juhn et al., 1993, Katz and Autor, 1999, Goldin and Katz, 2008, among others). Observing that more recent changes in the US wage distribution were not uniformly favoring higher skills, the basic SBTC hypothesis was refined by the task-based approach (Autor et al., 2003, 2008, Acemoglu and Autor, 2011). This more refined version of the SBTC hypothesis explains further inequality increases by falling demand for non-manual routine occupations in the middle of the distribution which fall back when compared to manual routine occupations at the bottom and non-manual analytic occupations at the top of the distribution. At the same time, a number of researchers have criticized the focus on the SBTC hypothesis suggesting that compositional and institutional changes such as de-unionization and changes in the minimum wage account for a substantial part of the inequality increase (DiNardo et al., 1996, Card and DiNardo, 2002, Lemieux, 2006). The third line of explanation, international trade, was identified as less important in earlier studies (e.g. Katz and Murphy, 1992) but has been taken up again as a potentially important factor more recently (Autor et al., 2014, Ebenstein et al., 2014, Firpo et al., 2014). Finally, a number of recent contributions have emphasized the potential role of growing heterogeneity between firms for the rise in wage inequality (e.g., Card et al., 2013, Barth et al, 2016, Baumgarten et al., 2018, and Song et al., 2019).

In order to evaluate these explanations in a more general sense, it is important to look at the relevance of these factors for a range of countries. A particular interesting case is Germany, given its large degree of integration in the world economy and its relative economic importance within the European Union. Adding to previous research on the German wage distribution (Dustmann et al., 2009, Antonczyk et al., 2010, Card et al., 2013, see more detailed literature review below), this paper makes the following contributions. First, we use a different data set than most of the other studies that have examined the German wage distribution. We use data from the mandatory *German Structure of Earnings Surveys (GSES)* conducted by the German Federal Statistical Office, which, compared to the widely used administrative data sets provided by the Institute for Employment Research (IAB), includes information on hourly wages (instead of daily or monthly earnings), is not censored at the social security contributions threshold and contains a richer set of covariates.[1] In particular, our data set includes information on unionization at the individual rather than at the firm level, which makes it different from all other data sets for Germany known to us. We show that this feature makes a substantial difference for the results, suggesting that previous studies have substantially underestimated the strong role de-unionization played for the rise of German wage inequality.

Our second contribution is that we simultaneously consider a very large set of potential explanatory factors for changes in the wage distribution, larger than in previous contributions. Our set of explanatory factors covers all the major explanations for rising wage inequality that have been put forward in the literature including rich information on worker characteristics, firms, union coverage, information on the task composition of occupations as well different measures of internationalization. As in other contexts, considering as many potential factors at the same time as possible is important to rule out spurious findings and to pin down the quantitative importance individual factors. It is very clear that, if important factors of distributional change are omitted from the analysis, then their impact will be spuriously picked up by the factors included. For our econometric analysis, we use a powerful tool for distributional analysis, the RIF regression approach (Firpo et al., 2009, 2018), which is particularly well-suited to control for a large number of explanatory variables. We also address some issues in the empirical implementation of RIF regressions, which may be of interest to researchers who want to apply this method.

To preview our results, we find that the most salient factors behind the recent rise in German wage inequality were a dramatic decline of union coverage and compositional changes of

[1]The *LIAB* was used by Dustmann et al., 2009, Card et al., 2013, and Baumgarten et al., 2018, among many others. The only other study using the *GSES* we are aware of is Antonzcyk et al., 2010., who use only a subset of the waves considered by us.

the work force with respect to age and education. These results hold after simultaneously controlling for an extensive list of other determinants of the wage distribution including information on job tasks, firm characteristics, measures of internationalization, regional wage differences and the sector composition of the economy. Using information on individual union coverage, we demonstrate that the use of firm-level information on unionization does not fully capture the substantial effect de-unionization had on the German wage distribution. Our results point to the fact that the mere shrinking of the part of the economy in which wages were more compressed was to a large part responsible for the trend towards increasing wage inequality. As the second most important factor, we identify compositional effects related to personal characteristics, especially workers' age and education. Such effects are consistent with the hypothesis that the increasing demand for higher skills due to SBTC was matched by rising supply for such skills due to educational upgrading and population aging. This follows from the insight that, in the absence of rising demand due to SBTC, rising supply of high skills would have depressed the wage premia for such skills which cannot be observed in the data. We do measure some wage structure effects related to internationalization, firm heterogeneity, task changes, and regional wage convergence, but these were very moderate compared to the dominating effects of de-unionization and compositional changes of the workforce.

The rest of the paper is structured as follows. Section 2.2 provides a review of some related literature. In sections 2.3 and 2.4, we describe our data and econometric methods. Section 2.5 presents our empirical results. In section 2.6, we discuss these results and provide some conclusions.

## 2.2 Literature review

In this section, we provide a selective review of contributions dealing with changes in the German wage structure and with effects of the factors considered by us on the wage structure in other countries. Based on administrative data derived from social security records, Dustmann et al. (2009) analyzed changes in the distribution of daily earnings in West Germany up to 2004. They showed that inequality increases first started in the 1980s at the top, and then in the 1990s at the bottom of the distribution, about a decade later than in the US. Their analysis suggests that compositional effects of personal characteristics account for a substantial part of inequality changes in the upper half of the distribution and nonnegligible shares at the bottom, while compositional changes of de-unionization explain considerable changes at the bottom but only some changes in the upper part of the distribution. Dustmann et al. (2009) consider each of the factors mentioned above separately and do not provide a break-down of the quantitative importance of each factor controlling for all other factors.

Based on a different data base, Antonzcyk et al. (2009) examined polarization effects of task changes on the distribution of monthly wages. They find that changes in task assignment reduced rather than increased wage inequality. Antonzcyk et al. (2010) used two waves of the data base we also use in this article in order to study changes in the West German wage distribution and the gender wage gap between 2001 and 2006. Their results suggest that changes in firm-level characteristics other than those related to union bargaining were the most important determinants of rising inequality, while changes in unionization did not have much explanatory power when other firm-level characteristics were controlled for. Antonzcyk et al. (2010) use a different methodology and a less extensive set of explanatory variables than we do. Moreover, there were substantial developments in some covariates outside the period 2001 to 2006 (esp. de-unionization), which are not covered by their analysis.

Also using administrative data, Card et al. (2013) studied the effects of fixed person and firm effects on the distribution of daily wages. They conclude that both increasing dispersion in person and in firm effects, as well as increasing assortative matching of high person to high firm fixed effects contributed to increasing wage inequality. Based on linked employer-employee data, Ohlert (2016) studies determinants of establishment heterogeneity in Germany. His study concludes that increasing differences in firm size and workforce composition contributed to rising inequality, while changes in union coverage played no important role. Also based on administrative data but without information on union coverage, Rinawi and Backes-Gellner (2015) examine task-composition effects on the wage structure. At odds with Antonczyk et al. (2009), they find that task effects explain up to one third of the rise in wage inequality. Using the same data, Ehrl (2017) attributes most of the increase in German wage inequality to differences in returns to characteristics and identifies occupation-specific skills as the most important single factor.

In a setup very similar to that used in our study, but based on the administrative data sets used in the other contributions, Baumgarten et al. (2018) aim to disentangle between-plant and within-plant sources of wage inequality in Germany. Similar to what we find in this study, Baumgarten et al. (2018) identify an important role for de-unionization for rising wage inequality in Germany. They also estimate significant effects due to shifts between industries, which we do not find in our study. While the set of firm-level variables in Baumgarten et al. (2018) is more informative than ours, we have access to more detailed information on individual-specific covariates than available in administrative data. In particular, we have information on union coverage at the individual rather than at the firm level. We analyze these differences in more detail below.

There is a small number of articles that empirically address aspects of internationalization for wages in Germany. Schank et al. (2007) and Klein et al. (2013) investigated the exporter wage premium, while Geishecker and Görg (2008) and Baumgarten et al. (2013) studied wage penalties associated with offshoring. These articles contain useful information on the effects of internationalization on wages but do not provide a full distributional analysis that quantifies the magnitude of these effects on the overall wage distribution. A full distributional analysis of the exporter wage premium is given in Baumgarten (2013), who finds that these effects are rather small when individual and firm characteristics are controlled for. Baumgarten et al. (2018) also find moderate effects of exporting on the wage structure in Germany.

Our short review of previous contributions on changes in the German wage distribution demonstrates that the literature has not reached a consensus about its main driving forces. As a related article for the US, we would like to point out the study by Firpo et al. (2014) who have analyzed the influence of detailed task measures and measures of offshorability on changes of the US wage distribution. Consistent with Autor et al. (2008), they find that, while distributional change in the 1980s was very monotonic (high quantiles gained, lower quantiles lost), this pattern became U-shaped in the 1990s and 2000s. They further show that (in contrast to what we find for Germany) recent inequality increases were associated with wage structure rather than composition effects and that offshorability became a more influential factor in the 1990s and 2000s.

## 2.3 Data and descriptive statistics

The empirical analysis in this paper uses information from four waves (1995, 2001, 2006, 2010) of the *German Structure of Earnings Surveys (GSES)* provided by the German Federal Statistical Office. The *GSES* are linked employer-employee data, which allow us to to consider a rich set of covariates both at the person and the firm level. We use the minimally anonymized version of the *GSES* which is only accessible onsite at the German Statistical Offices.[2] From a technical point of view, the *GSES* are the result of a two-stage random

[2] This study uses the *GSES* as well as supplementary data from the Linked-Employer-Employee Dataset (LIAB) of the Institute for Employment Research (IAB), Nürnberg, and the BIBB Labor Force Survey of the Federal Institute for Vocational Training (BIBB).

sample. The first stage represents a draw from all German establishments with at least ten employees subject to social insurance contributions. The second stage is a random draw from all employees working in the selected establishment. We use appropriate sample weights in all our analyses to ensure that our results are representative for the population of firms and workers studied by us. The information in *GSES* is highly reliable due to the fact that firms' participation is compulsory under German law.

Our data differ from the widely used administrative data sets provided by the Institute for Employment Research (IAB) in that they contain information on hourly wages (rather than daily earnings) and that the wage information is not censored at the social security contribution ceiling. Hourly wages more directly reflect the prices paid in the labor market than monthly or daily earnings and are thus better suited to test the theories described in the introduction. Focussing on hourly wages also makes results more comparable to those for the US for which most studies have used hourly wages (e.g., DiNardo et al., 1996, Lemieux, 2006, Autor et al., 2008, Firpo et al., 2014). Another advantage is that the GSES include a larger and more reliable set of covariates at the individual level than available in administrative data.

The disadvantages of the *GSES* are that it is not a panel study and that its coverage of the economy was incomplete in the early waves. In order to ensure comparability over time, we have to restrict our analysis to the 24 sectors listed in table A1. A comparison of our sample with other data sources suggest that our choice of sectors covers around 70 percent of the German economy with an emphasis on the traditionally strong manufacturing sector. In order to assess the implications of this limitation, we have checked in the alternative *SIAB* data (Sample of Integrated Labor Market Biographies) how imposing our sector restrictions influences measured inequality trends. As expected, overall trends are very similar, but imposing our sector restrictions implies somewhat lower inequality levels (because we miss part of the service sector). It turns out that these differences are entirely due to the lower half of the distribution, while numbers for the upper half are practically identical to the ones in the SIAB (see figures A1 to A3 in the appendix).

In order to enhance the information content of our data set further, we merge complementary information from two additional data sets, the *LIAB* (for firms' export status) and the *BIBB-IAB* (for occupational task measures) as well as aggregate information provided by the Federal Statistical Office (see next section for more details). Finally, we restrict our sample to prime age (20-60 years) men working full-time (i.e. at least 30 hours per week). In line with the existing literature, we do not include women in the present analysis, given their much lower participation rate in full-time work and given the potential difficulties of sample selection bias.

### 2.3.1 Hourly wages

Our hourly wage measure is defined as October earnings including additional payments from overtime and bonuses from shift work, divided by paid working hours in October including overtime. We inflate price levels in 1995, 2001, 2006 to the 2010 level using the German consumer price index (CPI). For reasons of plausibility, we exclude a small number of wage observations with less than 4 euros per hour as well as those associated with a monthly working time of more than 349 hours. Although the wage information in the *GSES* is largely uncensored, a censoring threshold at 25,000 DM (approxmiately 12,782 Euro) applied in 1995. In order to ensure comparability over time, we extend this censoring threshold to all other years adjusting for changes in the price level (for example, the implied censoring point for 2010 amounts to 15.879 Euro). We argue that we are still able to provide a comprehensive picture of the overall distribution of male hourly wages, as this censoring affected only about 200 (approximately 0.03% ) of the observations for 1995 and a similar, though slightly increasing number of observations in the other waves (2001: 0.05%, 2006: 0.16%, 2010: 0.18%). Ultimately, our sample selection criteria lead to a total number of 1,923,542 observations used in our analysis (1995: 592,198 employees in 23,668 firms, 2001: 359,495 employees in 15,438 firms, 2006: 533,497 employees in 15,477 firms, 2010: 438,352 employees in 13,285 firms).

### 2.3.2 Explanatory factors

Our analysis considers the following explanatory variables which we combine into seven different subgroups representing the different factors whose influence on the wage distribution we study in our decomposition analyses. We label the different subgroups as *Unionization*, *Personal*, *Tasks*, *Internationalization*, *Firm*, *Sector*, *Region*. Descriptive statistics on these variables and their change over time are given in table A1 in the appendix.

#### Unionization

In contrast to other data sets for Germany, our data includes information about union coverage *at the employee level*. This means firms report for each individual separately whether or not a given worker was paid according to some union agreement (in the original data, firms report the id of the exact union agreement used to determine the pay of the employee). This is in contrast to the broader firm-level information available in other data sets for Germany in which one only observes very broadly whether or not the firm takes part in specific forms of union bargaining, but not to what extent the pay of a given employee is determined by a union agreement.

In Germany, there are different variants of union bargaining. *Sectoral bargaining* refers to the case in which unions and employers form an agreement at the sector level. Workers need not be union members in order to be covered by sectoral union agreements. Similary, not all employees of the firm are necessarily paid according to the sectoral agreement. *Firm bargaining* represents the case in which unions and employers reach an agreement at the firm level. Similarly, such an agreement will typically (but not always) also apply to employees

**Table 2.1** – Development of unionization (individual- vs. firm-level information)

| Unionization (firm level) | | *Unionization (individual level)* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | No coverage | | | | Sectoral coverage | | | | Firm coverage | | | |
| | | 1995 | 2001 | 2006 | 2010 | 1995 | 2001 | 2006 | 2010 | 1995 | 2001 | 2006 | 2010 |
| | No coverage | 86.40 | 89.81 | 97.16 | 99.80 | 13.14 | 9.55 | 2.50 | 0.19 | 0.45 | 0.64 | 0.34 | 0.01 |
| | Sectoral coverage | 8.32 | 12.05 | 20.95 | 24.20 | 91.65 | 87.83 | 78.79 | 75.46 | 0.03 | 0.12 | 0.26 | 0.34 |
| | Firm coverage | 12.02 | 13.75 | 19.46 | 24.15 | 11.92 | 7.17 | 1.01 | 1.66 | 76.06 | 79.08 | 79.53 | 74.20 |
| | *Total* | 26.50 | 38.80 | 55.01 | 61.00 | 69.70 | 56.91 | 40.91 | 35.72 | 3.79 | 4.29 | 4.08 | 3.28 |

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.
Relative frequencies are reported within rows of each cell.

in the given firm who are not union members. It is the owners or the management of the firm who decide which bargaining regime to take part in. In particular, firms may decide not to engage in union bargaining, to leave existing agreements, or to deviate from existing agreements for individual workers. This includes the possibility of paying lower wages for new hires than for incumbents after having opted out of existing agreements. There may also be 'opening clauses' that exempt certain employees from union coverage. For more information on the varieties of union coverage in Germany, see Antonczyk et al. (2010), Brändle et al. (2011), Fitzenberger et al. (2011, 2013) and Dustmann et al. (2014).

It turns out that the distinction between the firm's general coverage status and that of the individual worker is quite significant. In table 2.1, we show that union coverage dropped generally, but that a substantial part of this drop was due to the fact that fewer and fewer workers in firms who reported to take part in union bargaining were actually paid according to a union agreement. In addition, even in firms that generally reported not to take part in union bargaining, some 14 percent of workers were paid according to a union agreement in 1995. This proportion dropped to zero by 2010. In general, the drop in the number of individuals paid according to union agreements was massive: from 1995 to 2010, the proportion of uncovered workers increased from 26.5 percent to 61 percent, while that of individuals paid according to sector agreements fell from 69.7 to 35.7 percent. By contrast, the group of individuals covered by firm contracts stayed approximately constant.

#### Personal characteristics

In this subgroup, we include the individual's age (8 categories), tenure (6 categories), educational qualification (6 categories) and occupational position (3 categories). Note that our education variable is more detailed and more reliable than in the administrative data where it is often missing or unreliable as it is not needed for the administrative purpose (see Fitzenberger et al., 2006). As evident from table A1, these variables followed some notable trends over the period under consideration. In particular, there was some aging of the German labor force as evident from the declining population shares of age groups below 40 years and the rising shares of those above 40 years. We observe a slightly rising share of higher tenure groups at the expense of the lowest tenure bracket (0-5 years). There was also considerable educational upgrading which is reflected in the declining share of individuals with lower/middle secondary schooling with or without vocational training, and the rising share of individuals with an upper secondary degree (with or without vocational training) and with tertiary education. Finally, there was a compositional shift from skilled blue collar work to white collar work, while non-skilled blue collar work stayed constant or even increased slightly.

### Tasks

For modeling occupational tasks, we exploit the information in the commonly used *German Qualification and Career Survey of Employees (BIBB-IAB)*, jointly provided by the Federal Institute for Vocational Training (BIBB) and the Institute for Employment Research (IAB). These data allow us to construct measures for the analytical, interactive and manual task content of individuals' jobs. More precisely, we use three independent cross sections, each covering 20,000-30,000 individuals from the years 1998/99, 2006 and 2012, which come closest to our sample period. Given some inconsistencies in how the task questions were asked in these surveys over time, we follow the common practice in the literature and consider time-constant task measures per occupation (Baumgarten et al., 2013, Firpo et al., 2014, Böhm et al., 2016). In order to make the task information independent of time, we pool the information from all the three surveys.

Table A2 in the appendix documents the mapping of the different activities into the three task-groups, i.e. *analytical*, *manual* and *interactive*. In doing so, we closely follow Gathmann and Schönberg (2010). The share of a certain task-group $g$ is defined as the number of activities in group $g$ performed by an individual $i$ divided by the total number of tasks performed by the same individual, i.e.

$$Task_{ig} = \frac{number\ of\ activities\ in\ group\ g\ performed\ by\ i}{total\ number\ of\ activities\ in\ all\ groups\ performed\ by\ i}. \tag{2.1}$$

As common in the task literature (e.g. Spitz-Oener, 2006), these shares are first calculated at the person-level and then averaged at the level of 2-digit occupations. In figure A4 in the appendix, we document that the share of analytical and interactive tasks increased over the period 1995-2001, while that of manual tasks decreased.

### Internationalization

This group of covariates is intended to represent three different aspects of internationalization: the exporting behavior of firms on the one hand, and the pressure on 2-digit occupations exerted by offshoring and import competition on the other. As the *GSES* data lack a firm-level variable on export behavior, we impute this information from the *LIAB* using an ordered logit model for the categories *No Exports*, *Export share 1-25%*, *Export share 26-50%* and *Export share 51-100%*, where export share represents exports in total sales. For this imputation, we exploit a large number of individual and firm characteristics that are available in both data sets in order to predict the export share category for each observation in the *GSES*.[3] Our predicted export share variable displays very similar patterns as in the original *LIAB* data. As shown in the summary statistics in table A1, we observe a steeply increasing trend for the share of the predicted *Export share 51-100%* category at the expense of the lower categories, which was partly reversed after the financial crisis in 2008. By contrast, the share of observations in the *No Exports* category stayed relatively

[3]Our model includes education (7 categories), a polynomial in age and tenure, occupational status (4 categories), sector (20 categories), and firm size (7 categories).

constant with minor fluctuations.

In addition, we use information from the German national accounts (Federal Statistical Office of Germany, 1999-2014) at the 2-digit industry level in order to derive measures of wage pressure on occupations due to offshoring and imports of consumer goods. We differentiate between 77 occupations and 24 industries.[4] Following Baumgarten et al. (2013) and Ebenstein et al. (2014), we first consider the share of intermediate input imports coming from the same industry abroad as an indicator for offshoring at the industry level. In order to arrive at a measure reflecting the wage pressure on occupation $k$ due to trends in offshoring activities across industries, we compute the average of these offshoring intensities across all industries in which workers with occupation $k$ work (using the employment shares of occupation $k$ in industry $j$ as weights). Consequently, our measure of wage pressure on the 2-digit occupation $k$ in year $t$ due to offshoring is given by

$$Offs_{kt} = \sum_{j=1}^{J} \frac{L_{kjt}}{L_{kt}} Offs_{jt} \tag{2.2}$$

where $Offs_{jt}$ denotes the industry-level offshoring intensities and $\frac{L_{kjt}}{L_{kt}}$ is the employment share of occupation $k$ in industry $j$ in year $t$.

For imports of consumer goods, we proceed analogously. Let $Imports_{jt}$ be the share of imports of consumer goods in industry $j$ in year $t$. Our measure of wage pressure on occupation $k$ in year $t$ due to imports of consumption goods in the sectors this occupation is employed in is then defined as

$$Imports_{kt} = \sum_{j=1}^{J} \frac{L_{kjt}}{L_{kt}} Imports_{jt}. \tag{2.3}$$

[4] The data include the *Classification of Occupations (KldB)* at the 2-digit level, i.e. *KldB75* in 1995 and 2001, *KldB88* in 2006 and 2010. For reasons of time consistency minor aggregations were required leading to a total number of 77 occupations. At the industry level, we consider the 24 sectors of the economy listed in table A1, see next section for more details.

**Firms, sector, region**

Under the label *Firm* we include information on firmsize (7 categories) and information on whether corporate management is influenced by the state. The distribution of these characteristics was relatively stable over the period 1995 to 2010 (see table A1). In order to address changes in the composition of the economy over time and changes in inter-industry wage differentials, we include under the label *Sector* categorial dummies for 24 different sectors of the economy based on the *German Classification of Economic Activities (WZ)*, which we harmonized over time.[5] There were generally no big shifts in the sectoral composition between 1995 and 2010. Notable exceptions were a sizable decline of the construction sector and a moderate growth of wholesale trade (table A1). Finally, we include information on the federal state in which a person worked under the label *Region* (16 categories). Including this information is potentially important as there are sizable differences in mean wages paid in different federal states, especially if one compares East and West German states.

## 2.4 Econometric methods

In order to study the quantitative importance of the different sets of covariates on the changes of the wage distribution over the period 1995 to 2010, we apply RIF-regressions (Firpo et al., 2009, 2018). The *Recentered-Influence-Function (RIF)* decomposition is based on the recentered influence function defined as $RIF\,(y, \nu) = \nu + IF\,(y; \nu)$ which integrates to the statistic of interest $\nu\,(F_y) = \int RIF\,(y; \nu)\,dF(y) = E(RIF\,(y; \nu))$, where $F_y$ is

[5] Our sector classification is derived from the 2-digit *German Classification of Economic Activities (WZ)*. The *German Classification of Economic Activities (WZ)* changed between the waves of 1995, 2001 (WZ93) and 2006 (WZ03) as well as 2010 (WZ08). While the change from WZ93 and WZ03 should not affect our results at the 2-digit level, we acknowledge that the latter change might give rise to minor inconsistencies for the period 2006-2010.

the distribution function of the dependent variable (log hourly wage in our case). In the simplest form, the RIF is modeled as a linear function of the explanatory variables, i.e. $E\left[RIF\left(Y;\nu\right)|X\right] = X\gamma$, where $\gamma$ can be estimated by means of simple OLS. The statistic of interest is then obtained as $\nu\left(F_y\right) = E\left(E\left[RIF\left(Y;\nu\right)|X\right]\right) = E(X)\gamma$, using the sample counterparts estimated by OLS (i.e. $\hat{\nu} = \bar{X}\hat{\gamma}$).

As shown in Firpo et al. (2009), the coefficients $\gamma$ of the RIF regression represent the effects of marginal shifts in the distribution of the components of $X = (X_1, \ldots, X_k)$ on the statistic of interest. For example, if $\nu\left(F_y\right)$ is an unconditional quantile of $y$ and $X_j$ union coverage status, then $\gamma_j$ will reflect how much quantile $\nu\left(F_y\right)$ of the unconditional distribution of wages $y$ is increased or decreased if the share of unionized workers is marginally increased. The RIF regression follows the classic division into 'composition' and 'wage structure' effects. It uses the idea that the distribution of wages will change whenever the composition of the workforce changes, even if wages paid for given characteristics stay constant ('composition effect'). On the other hand, it may change if the composition of the workforce stays constant, but wages paid for given characteristics change ('wage structure' effect). We will use this method to carry out detailed decomposition analyses for different inequality measures $\nu\left(F_y\right)$ based on quantiles, such as the 85-15, 85-50 and 50-15 log wage gap, the Gini coefficient and the variance of log wages.

Specifically, we use a refinement of this method suggested by Firpo et al. (2014, 2018), i.e. the RIF decomposition is combined with the semi-parametric reweighting approach introduced by DiNardo et al. (1996). This is done to avoid bias in case the linear specification for the RIFs described above is not sufficiently precise, as the linear specification is only valid locally. The basic idea underlying this approach is to create an artificial time period 01, in which the distribution of $X$ in period 0 is reweighted to that of period 1. Using these three periods, two separate Oaxaca-Blinder decompositions are run on the recentered influence function, leading to

$$\Delta_O^{\nu} = \underbrace{\left(\bar{X}_{01} - \bar{X}_0\right)\hat{\gamma}_0^{\nu}}_{\Delta_{X,p}^{\nu}} + \underbrace{\bar{X_{01}}\left(\hat{\gamma}_{01}^{\nu} - \hat{\gamma}_0^{\nu}\right)}_{\Delta_{X,c}^{\nu}} + \underbrace{\bar{X}_1\left(\hat{\gamma}_1^{\nu} - \hat{\gamma}_{01}^{\nu}\right)}_{\Delta_{S,p}^{\nu}} + \underbrace{\left(\bar{X}_1 - \bar{X}_{01}\right)\hat{\gamma}_{01}^{\nu}}_{\Delta_{S,c}^{\nu}}. \tag{2.4}$$

In this equation, the detailed composition effects $\Delta_{X,p}^{\nu}$ reflect the contribution of changes in the distribution of particular covariates (or groups of covariates) to the overall change of the distributional statistic. For example, suppose that there are wage differentials *between* sectors covered and those not covered by unions in the sense that union coverage is associated with nonnegative wage premia.[6] In addition, it may be the case that inequality *within* the sectors covered by unions differs from inequality in sectors not covered (e.g., unions compress wages in the sectors covered by them). Now assume that union coverage in the economy declines. The overall compositional effect of this decline on wage inequality may be positive or negative depending on whether the decrease in inequality between sectors dominates the increase in inequality due to the declining share of sectors with low levels of within-inequality. The specification error $\Delta_{X,c}^{\nu}$ in (2.4) reflects the differences in estimated RIF coefficients in the sample of period 0 and the coefficients estimated in the sample of period 0 whose distribution was reweighted to that of period 1.

The wage structure effect $\Delta_{S,p}^{\nu}$ represents the contributions of changes in the effects $\gamma$ individual covariates (or groups of covariates) have on the distribution of wages. This includes effects on pay inequality *between* and *within* subgroups. In the above example, this would include changes in the magnitude of wage differentials *between* sectors covered and those not covered by unions, as well as changes in the amount of wage compression *within* sectors resulting from changes in union policies (e.g., unions might increase or loose their ability to compress wages). Finally, the reweighting error $\Delta_{S,c}^{\nu}$ reflects differences in the means of covariates in sample period 1 and those in sample period 0 whose distribution was reweighted to that of sample period 1. The reweighting error will be close to zero if

[6] The following discussion closely follows Firpo et al. (2009).

reweighting is successful in changing the distribution of covariates in sample period 0 to that of sample period 1.

To our best knowledge, the RIF-OLS decomposition is the only method known that is capable of providing a detailed, path-independent decomposition of arbitrary distributional statistics into composition and wage structure effects. Other decomposition methods are either confined to particular distributional statistics (e.g. based on variance decompositions, Juhn et al., 1993), provide no detailed decomposition results (Machado and Mata, 2005, Melly, 2005, Chernozhukov et al., 2013), or provide detailed decomposition results that depend on some ordering of factors (DiNardo et al., 1996, Antonczyk et al., 2010). For more details, see the general discussion in Fortin et al., 2011.

As described in the literature, detailed decompositions of wage structure effects for a set of categorical variables depend on the choice of the omitted reference group (Fortin et al., 2011). This also applies to the RIF decomposition described above. In preliminary estimations, we found that the detailed wage structure effects estimated by us sometimes considerably depended on which reference groups for the various sets of our categorial variables we chose. This is not surprising as the intercept of a regression always represents the average outcome for a very specific reference individual (i.e. an individual with the base level of education, age, tenure, sector, firmsize, region etc.). The intercept of the regression (and hence the exact value of all other regression coefficients) will therefore depend to a large extent on how the position of the reference individual changes over time. In order to make our regression results independent of the choice of the reference individual, we normalize the RIF regression coefficients within sets of categorial variables such that they sum up to zero, i.e. $\sum_{j \in J} \gamma_j = 0$, where $J$ is a set of categorial dummy variables summing up to one (e.g. age categories). Gardezabal and Ugidos (2004) discuss this kind of normalization for the case of the standard Oaxaca-Blinder decomposition. An advantage of the normalization is that it only shifts the intercept of the regression, leaving the relative differences between coefficients intact.

Applying this normalization will not only make results independent of the choice of a reference group but will also facilitate the general interpretation of RIF decomposition results. Given that the RIF regression coefficients for groups of categorial variables are normalized to sum up to zero, information about the general level of the statistic modeled by the RIF regression (e.g. a quantile) will be shifted to the intercept of the regression, while differences in regression coefficients will only reflect deviations of individual categories from this general level.[7] The intercept of the RIF regression will therefore capture general changes in unconditional quantiles (or other inequality measures) that are not related to pure relative changes within groups of categorial variables and which therefore cannot be attributed in a detailed way to individual regressors. They may still reflect changes in the relative importance of groups of categorial variables (e.g. the importance of age vs. education effects), but such changes cannot be attributed to individual variables or groups of variables. They should therefore be summarized in the intercept as a general contribution to wage structure effects. Finally, changes in the intercept will also incorporate general changes in unconditional quantiles (or other inequality measures) that are due to factors not included as observables in the analysis.

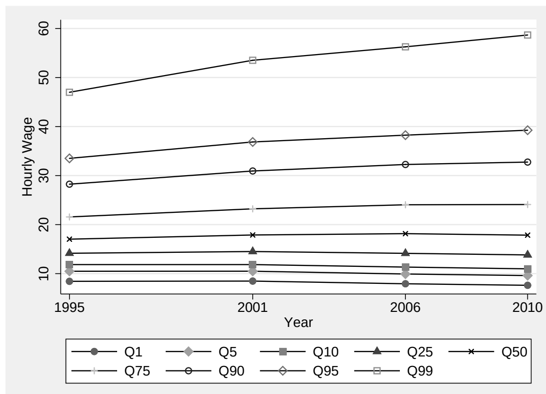
It is important to point out that RIF decomposition (as most other statistical decomposition techniques) ignores general equilibrium effects. It correspond to the hypothetical thought experiment of changing the distributions of observed covariates without changing the wage structure (see Fortin et al., 2011, for a more detailed discussion). Similarly, we emphasize that the RIF decomposition results should certainly not be interpreted as causal effects. However, they do represent a rich and informative description of how distributional change is related to changes in observables. Even if these relationships are not causal, it is important to identify variables through which distributional change is mediated, or with which it is correlated.

[7]We illustrate this kind of normalization for the case of a mean regression below.

## 2.5 Empirical results

### 2.5.1 Development of inequality

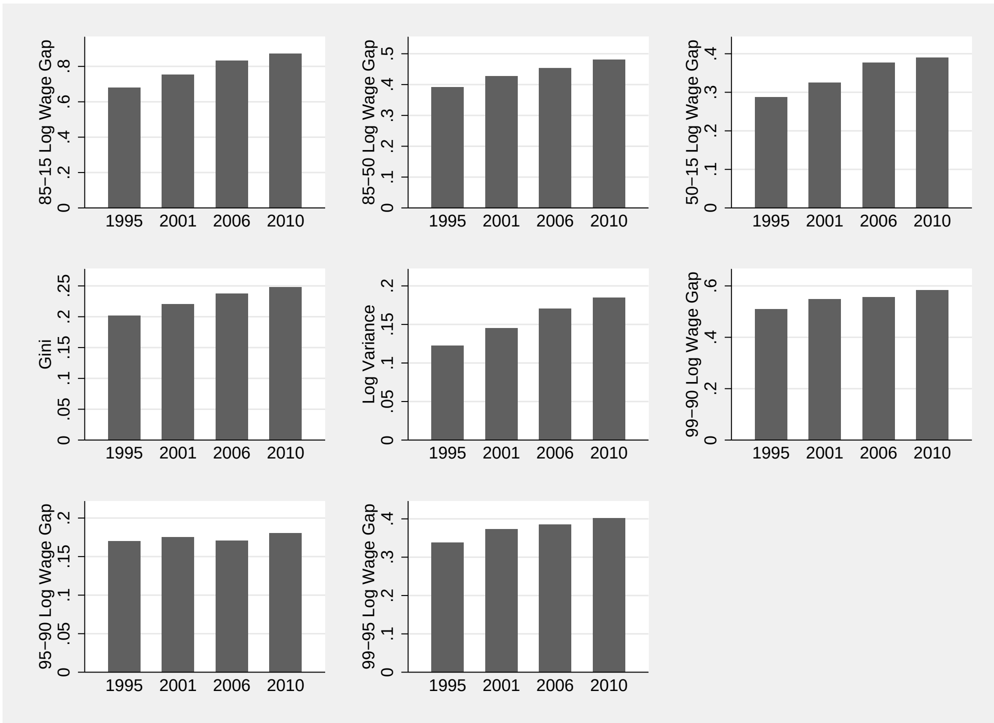
**Figure 2.1** – Quantiles of real hourly wage, 1995-2010




Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

The general development of the distribution of real hourly wages between 1995 and 2010 is displayed in figure 2.1. For the period as a whole, quantiles near or above the median gained whereas quantiles below the median lost. This general trend is confirmed by the results for the Gini coefficient and the variance of log wages. The results for the 50th to 15th percentile and the 85th to 15th percentile in figure 2.2 suggest that the inequality increases were steeper in the lower than in the upper part of the distribution, except at the very end of the observation period. As to the top of the distribution, there were only moderate inequality increases between the 95th and the 90th percentile but more pronounced increases between the 99th and the 95th percentile. Although our data excludes developments at the very top of the distribution of hourly wages, our findings are consistent with the view that changes in the upper part of the German wage distribution were relatively modest when compared to other, especially Anglo-Saxon countries (see Atkinson et al., 2011, Piketty and Saez, 2014, Bartels and Jenderny, 2015).

**Figure 2.2** – Development of inequality, 1995-2010




Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

### 2.5.2 Trends in between-group inequality

In order to prepare the distributional RIF-analysis presented below, we first examine trends in the relationship between observed characteristics and hourly wages as measured by OLS regressions of log hourly wages on our long list of covariates (table 2.2). This regression represents trends in between-group inequality, i.e. in average wage differentials between narrowly defined cells of workers with identical observed characteristics. In order to facilitate interpretation, we apply the normalization described above, i.e. we center estimated coefficients around zero within groups of categorial regressors. This will provide wage differentials with respect to a mean level of returns normalized to zero.[8]

[8]For example, the estimated coefficients for the age categories indicate that in 1995, being in the age group 20 to 25 years was associated with a wage penalty of 13.6 percentage points compared to the mean

**Table 2.2** – OLS regressions of log hourly wage on covariates

| | 1995 | | 2001 | | 2006 | | 2010 | |
|---|---|---|---|---|---|---|---|---|
| | Coeff. | Std. | Coeff. | Std | Coeff. | Std | Coeff. | Std |
| Age 20-25 | -0.136 | 0.002 | -0.162 | 0.003 | -0.184 | 0.004 | -0.173 | 0.004 |
| Age 26-30 | -0.058 | 0.001 | -0.072 | 0.002 | -0.099 | 0.003 | -0.093 | 0.003 |
| Age 31-35 | -0.006 | 0.001 | -0.005 | 0.002 | -0.016 | 0.002 | -0.022 | 0.002 |
| Age 36-40 | 0.021 | 0.001 | 0.031 | 0.001 | 0.04 | 0.002 | 0.033 | 0.002 |
| Age 41-45 | 0.035 | 0.001 | 0.041 | 0.002 | 0.065 | 0.002 | 0.070 | 0.002 |
| Age 46-50 | 0.045 | 0.001 | 0.047 | 0.002 | 0.064 | 0.002 | 0.073 | 0.002 |
| Age 51-55 | 0.054 | 0.002 | 0.056 | 0.002 | 0.066 | 0.002 | 0.061 | 0.002 |
| Age 56-60 | 0.046 | 0.002 | 0.062 | 0.003 | 0.063 | 0.003 | 0.051 | 0.003 |
| Variance age coefficients (×100) | 0.383 | | 0.541 | | 0.777 | | 0.709 | |
| Tenure 0-5 | -0.069 | 0.002 | -0.078 | 0.002 | -0.086 | 0.003 | -0.093 | 0.003 |
| Tenure 6-10 | -0.006 | 0.001 | -0.020 | 0.003 | -0.023 | 0.002 | -0.021 | 0.002 |
| Tenure 11-15 | 0.010 | 0.001 | 0.005 | 0.002 | 0.009 | 0.002 | 0.007 | 0.003 |
| Tenure 16-20 | 0.010 | 0.001 | 0.028 | 0.002 | 0.025 | 0.002 | 0.021 | 0.002 |
| Tenure 21-25 | 0.021 | 0.001 | 0.029 | 0.002 | 0.039 | 0.003 | 0.036 | 0.002 |
| Tenure >25 | 0.034 | 0.002 | 0.035 | 0.003 | 0.036 | 0.003 | 0.049 | 0.003 |
| Variance tenure coefficients (×100) | 0.110 | | 0.156 | | 0.191 | | 0.221 | |
| Lower/middle secondary w/o vocational training | -0.109 | 0.003 | -0.129 | 0.003 | -0.130 | 0.004 | -0.140 | 0.004 |
| Lower/middle secondary w/ vocational training | -0.061 | 0.002 | -0.073 | 0.003 | -0.079 | 0.003 | -0.091 | 0.003 |
| Upper secondary (German high school equiv.) | -0.017 | 0.004 | -0.013 | 0.005 | -0.017 | 0.005 | -0.009 | 0.006 |
| University of Applied Science (Fachhochschule) | 0.092 | 0.003 | 0.101 | 0.004 | 0.099 | 0.005 | 0.105 | 0.004 |
| University | 0.169 | 0.004 | 0.202 | 0.005 | 0.210 | 0.006 | 0.224 | 0.005 |
| Missing information | -0.074 | 0.005 | -0.088 | 0.007 | -0.083 | 0.009 | -0.090 | 0.006 |
| Variance education coefficients (×100) | 0.973 | | 1.348 | | 1.404 | | 1.621 | |
| Non-skilled blue collar | -0.106 | 0.002 | -0.099 | 0.003 | -0.107 | 0.004 | -0.117 | 0.003 |
| Skilled blue collar and foremen | -0.008 | 0.001 | -0.018 | 0.002 | -0.014 | 0.003 | -0.003 | 0.003 |
| White collar | 0.114 | 0.002 | 0.117 | 0.003 | 0.121 | 0.004 | 0.119 | 0.004 |
| Variance occupational position coefficients | 0.809 | | 0.794 | | 0.876 | | 0.929 | |
| Offshoring (0-100%) | 0.005 | 0.001 | 0.006 | 0.001 | 0.008 | 0.001 | 0.011 | 0.001 |
| Imports of consumption goods (0-100%) | -0.002 | 0.000 | -0.002 | 0.000 | -0.002 | 0.000 | -0.002 | 0.000 |
| No Exports | -0.028 | 0.004 | -0.028 | 0.005 | -0.036 | 0.006 | -0.045 | 0.007 |
| Export share 1-25% | -0.008 | 0.002 | 0.001 | 0.003 | 0.003 | 0.006 | -0.021 | 0.004 |
| Export share 26-50% | 0.009 | 0.003 | 0.017 | 0.004 | 0.004 | 0.005 | 0.023 | 0.005 |
| Export share 51-100% | 0.027 | 0.003 | 0.010 | 0.005 | 0.029 | 0.005 | 0.043 | 0.007 |
| Variance export coefficients (×100) | 0.041 | | 0.029 | | 0.054 | | 0.121 | |
| Mining and other quarring | -0.058 | 0.013 | -0.118 | 0.051 | -0.006 | 0.019 | 0.047 | 0.029 |
| Food products, beverages, tobacco | -0.047 | 0.005 | -0.071 | 0.008 | -0.081 | 0.009 | -0.044 | 0.011 |

level of returns to age, while individuals between 51 and 55 years received a premium of 5.4 percentage points above this mean level.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Textiles | -0.086 | 0.008 | -0.073 | 0.016 | -0.081 | 0.012 | -0.124 | 0.017 |
| Wood | -0.026 | 0.008 | -0.041 | 0.012 | -0.070 | 0.011 | -0.090 | 0.010 |
| Paper | -0.006 | 0.008 | 0.012 | 0.008 | 0.030 | 0.011 | 0.030 | 0.009 |
| Printing | 0.144 | 0.007 | 0.130 | 0.009 | 0.123 | 0.009 | 0.078 | 0.013 |
| Coke and petroleum products | 0.099 | 0.025 | 0.146 | 0.021 | 0.171 | 0.021 | 0.220 | 0.056 |
| Chemicals | 0.038 | 0.006 | 0.033 | 0.007 | 0.041 | 0.008 | 0.036 | 0.009 |
| Rubber, plastic | -0.025 | 0.006 | -0.036 | 0.007 | -0.047 | 0.011 | -0.035 | 0.010 |
| Non-metallic products | 0.001 | 0.005 | -0.017 | 0.006 | -0.034 | 0.010 | -0.018 | 0.013 |
| Basic metals | 0.040 | 0.007 | 0.054 | 0.013 | 0.042 | 0.013 | 0.038 | 0.011 |
| Fabricated metal products | 0.018 | 0.005 | -0.011 | 0.007 | -0.009 | 0.009 | -0.026 | 0.011 |
| Computer, electronic, optical products | -0.002 | 0.006 | -0.004 | 0.008 | 0.007 | 0.007 | 0.008 | 0.010 |
| Electrical equipment | 0.002 | 0.006 | 0.000 | 0.008 | -0.014 | 0.010 | -0.005 | 0.010 |
| Machinery and equipment | 0.011 | 0.005 | 0.012 | 0.006 | -0.002 | 0.009 | -0.006 | 0.011 |
| Motor vehicles, trailers | 0.115 | 0.008 | 0.098 | 0.009 | 0.091 | 0.010 | 0.077 | 0.009 |
| Other transport equipment | 0.005 | 0.008 | 0.071 | 0.026 | 0.018 | 0.013 | 0.090 | 0.012 |
| Furniture etc | -0.024 | 0.008 | -0.048 | 0.011 | -0.083 | 0.012 | -0.035 | 0.008 |
| Electricity, water, recycling | 0.084 | 0.008 | 0.110 | 0.011 | 0.114 | 0.015 | 0.073 | 0.012 |
| Construction | 0.058 | 0.004 | 0.018 | 0.005 | 0.014 | 0.006 | 0.017 | 0.008 |
| Trade of vehicles | -0.040 | 0.015 | -0.037 | 0.007 | -0.032 | 0.008 | -0.076 | 0.015 |
| Wholesale trade | -0.078 | 0.007 | -0.062 | 0.009 | -0.018 | 0.009 | -0.037 | 0.010 |
| Retail trade | -0.161 | 0.008 | -0.141 | 0.012 | -0.175 | 0.024 | -0.243 | 0.015 |
| Finance and insurance | -0.060 | 0.007 | -0.023 | 0.012 | 0.001 | 0.010 | 0.025 | 0.013 |
| Variance sector coefficients (×100) | 0.457 | | 0.522 | | 0.546 | | 0.732 | |
| Firmsize 10-19 | -0.076 | 0.004 | -0.077 | 0.005 | -0.089 | 0.006 | -0.067 | 0.007 |
| Firmsize 20-49 | -0.051 | 0.004 | -0.062 | 0.004 | -0.065 | 0.006 | -0.056 | 0.006 |
| Firmsize 50-99 | -0.035 | 0.004 | -0.035 | 0.004 | -0.041 | 0.005 | -0.031 | 0.006 |
| Firmsize 100-199 | -0.005 | 0.003 | -0.001 | 0.005 | -0.014 | 0.005 | -0.010 | 0.007 |
| Firmsize 200-499 | 0.031 | 0.003 | 0.025 | 0.004 | 0.023 | 0.006 | 0.015 | 0.007 |
| Firmsize 500-999 | 0.057 | 0.003 | 0.065 | 0.006 | 0.081 | 0.011 | 0.046 | 0.007 |
| Firmsize >1000 | 0.078 | 0.004 | 0.086 | 0.006 | 0.104 | 0.006 | 0.103 | 0.007 |
| Variance firmsize coefficients (×100) | 0.285 | | 0.332 | | 0.456 | | 0.309 | |
| State-owned | -0.021 | 0.004 | -0.044 | 0.008 | -0.027 | 0.007 | -0.026 | 0.006 |
| Schleswig-Holstein | 0.100 | 0.006 | 0.079 | 0.008 | 0.061 | 0.008 | 0.068 | 0.011 |
| Hamburg | 0.175 | 0.007 | 0.191 | 0.021 | 0.151 | 0.011 | 0.154 | 0.014 |
| Lower Saxony | 0.082 | 0.005 | 0.084 | 0.005 | 0.050 | 0.007 | 0.039 | 0.007 |
| Bremen | 0.136 | 0.011 | 0.054 | 0.012 | 0.071 | 0.010 | 0.083 | 0.017 |
| North Rhine-Westphalia | 0.128 | 0.004 | 0.116 | 0.005 | 0.095 | 0.006 | 0.097 | 0.006 |
| Hesse | 0.106 | 0.004 | 0.127 | 0.007 | 0.122 | 0.007 | 0.106 | 0.008 |
| Rhineland-Palatinate | 0.096 | 0.006 | 0.089 | 0.006 | 0.08 | 0.009 | 0.080 | 0.007 |
| Baden-Wuerttemberg | 0.139 | 0.003 | 0.142 | 0.005 | 0.148 | 0.005 | 0.132 | 0.007 |
| Bavaria | 0.103 | 0.004 | 0.098 | 0.005 | 0.099 | 0.007 | 0.098 | 0.007 |
| Saarland | 0.073 | 0.008 | 0.084 | 0.012 | 0.068 | 0.011 | 0.064 | 0.011 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Berlin | 0.058 | 0.008 | 0.032 | 0.009 | 0.006 | 0.013 | 0.022 | 0.014 |
| Brandenburg | -0.190 | 0.010 | -0.178 | 0.009 | -0.157 | 0.010 | -0.157 | 0.013 |
| Mecklenburg-West Pomerania | -0.221 | 0.009 | -0.214 | 0.010 | -0.189 | 0.010 | -0.190 | 0.013 |
| Saxony | -0.267 | 0.007 | -0.249 | 0.009 | -0.22 | 0.010 | -0.219 | 0.011 |
| Saxony-Anhalt | -0.239 | 0.008 | -0.227 | 0.009 | -0.191 | 0.010 | -0.163 | 0.012 |
| Thuringia | -0.278 | 0.009 | -0.229 | 0.008 | -0.195 | 0.010 | -0.214 | 0.010 |
| Variance federal states Coefficients (x100) | 2.700 | | 2.322 | | 1.775 | | 1.728 | |
| Share of analytical tasks | 0.114 | 0.012 | 0.106 | 0.014 | 0.120 | 0.025 | 0.106 | 0.018 |
| Share of interactive tasks | 0.110 | 0.012 | 0.112 | 0.017 | 0.111 | 0.027 | 0.159 | 0.021 |
| Share of manual tasks | -0.224 | 0.005 | -0.218 | 0.007 | -0.231 | 0.007 | -0.266 | 0.009 |
| Variance task coefficients (x100) | 2.510 | | 2.377 | | 2.669 | | 3.576 | |
| No union coverage | 0.033 | 0.003 | 0.021 | 0.005 | -0.009 | 0.005 | -0.019 | 0.005 |
| Sectoral bargaining | -0.016 | 0.003 | -0.012 | 0.004 | -0.012 | 0.005 | -0.006 | 0.005 |
| Firm bargaining | -0.017 | 0.005 | -0.010 | 0.008 | 0.021 | 0.009 | 0.025 | 0.009 |
| Variance unionization coefficients (x100) | 0.054 | | 0.023 | | 0.022 | | 0.034 | |
| Constant | 2.901 | 0.005 | 2.92 | 0.009 | 2.931 | 0.009 | 2.964 | 0.009 |
| Root MSE | 0.220 | | 0.246 | | 0.267 | | 0.276 | |
| $R^2$ | 0.607 | | 0.583 | | 0.582 | | 0.588 | |

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

Standard errors clustered at establishment level.

Coefficients within groups of categorial regressors are centered around zero.

We briefly summarize the effects of the different covariates. In general, there was a general trend towards widening wage differentials between worker subgroups, reflected in the rising variance of OLS coefficients within subgroups of regressors (e.g. age). We observe a moderate widening of the returns to age, tenure, education and occupational position. The association of wages with offshoring was positive, i.e. occupations that were more affected by offshoring in the different sectors of the economy did not suffer but gain from these activities. On the other hand, we obtain a slightly negative effect of imports in consumption goods on the wages of occupations employed in the respective sectors, suggesting an import pressure effect. There were no changes of these wage differentials over time. By contrast, the exporter wage premium substantially increased over time.[9] Wage differentials across

[9]Note the plausible magnitude and high statistical precision of the estimated coefficients for our variables based on external imputations. This proves that our imputation introduces additional information into our data which turns out as highly significant effects in our regressions.

sectors moderately widened over our observation period. As regards the returns to firm size, we observe that the premia at very large firms increased, while those of medium-sized firms tended to fall. There was considerable convergence of wages across federal states (the variance of the regional coefficients dropped from 2.700 to 1.728). For the different task inputs, we observe a stable positive relative return for analytical tasks, while the return to interactive tasks increased at the expense of that to manual tasks. Note that these task premia and all other coefficients represent ceteris paribus effects holding constant education and a long list of further covariates.

There are interesting trends in wage differentials between workers covered and those not covered by union bargaining (last rows of table 2.2). Over the period under consideration, we observe a continuous trend of a deteriorating position of uncovered workers relative to covered workers. In 1995, not being covered by some union agreement was associated with a slightly higher pay than if the person was covered by a pay scheme negotiated by unions. This relationship was reversed from the mid 2000s onwards. Our interpretation of this pattern is that in the years 1995 and 2001, i.e. when union coverage was generally very high, individual non-coverage was mainly used to pay higher wages to highly productive workers. Towards the end of the observation period however, employers more and more used either individual or collective non-coverage in order to limit or even reduce the wages of uncovered workers.

Given the important role unions play in the German labor market, we also carried out the above wage regressions separately for individuals covered and those not covered by union agreements (tables A3 and A4 in the appendix). The results confirm the expectation that unions considerably compressed wage differentials across practically all observable covariates (reflected in the much lower variances of regression coefficients for the different sets of covariates) and within narrowly defined groups of workers with identical observable characteristics (as reflected in a lower estimate for the residual variance of the regression). It is especially these strong differences in inequality between covered and uncovered workers that suggest potentially important composition effects as a result of the secular decline in union coverage identified in the previous section (i.e. overall inequality will increase if the more compressed part of the economy shrinks).

### 2.5.3 RIF decomposition

Given the local nature of the RIF methodology, our strategy is to apply RIF decompositions separately to our three subperiods 1995-2001, 2001-2006, 2006-2010, and to aggregate (i.e. add up) the contributions of the different factors over the subperiods.[10] We start with a graphical analysis of the effects changes of our covariates have on unconditional quantiles. Figure 2.3 shows that the change of the distribution of log hourly wages between 1995 and 2010 was such that unconditional quantiles below the 35th percentile fell, while those above the 35th increased. This pattern is distinctively different from the changes in the US wage distribution over similar periods which featured a U-shaped pattern, i.e. especially middle quantiles lost in comparative terms, while lower and upper quantiles gained (Autor et al., 2008, Firpo et al., 2014).

Decomposing the overall change into composition and wage structure effects, we find that the pattern of composition effects shows the same monotonic behavior as the overall change, but that additional wage structure effects played some role in the upper middle range of the distribution. In the detailed plot of individual composition effects (figure 2.4), the most striking effect is that of de-unionization. The shrinking share of workers paid according to union pay schedules both substantially depressed quantiles at the lower end of the distri-

[10]The RIF-regression approximates the effects of marginal changes in the distribution of covariates on inequality measures. The approximation error will be the larger the bigger changes in the distribution of covariates are. In order to keep approximation errors small, it is therefore best to consider changes in covariates over the smallest subperiods available. In a previous version of this paper, Biewen and Seckler (2017), we report more details on individual subperiods. For our estimates, we provide bootstrapped standard errors based on 100 resamples. The resamples are a simultaneous draw from all four years and take account of the clustering at the firm level.

**Figure 2.3** – Aggregate decomposition 1995-2010

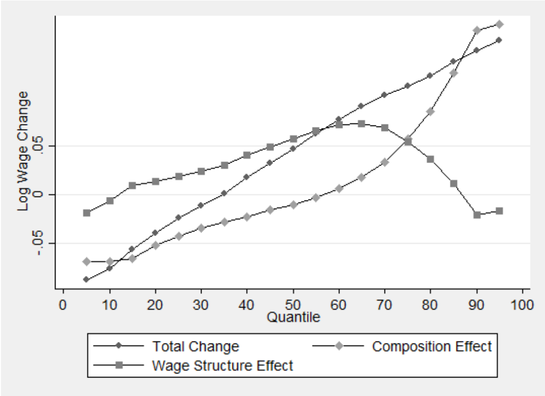



**Figure 2.4** – Composition effects 1995-2010

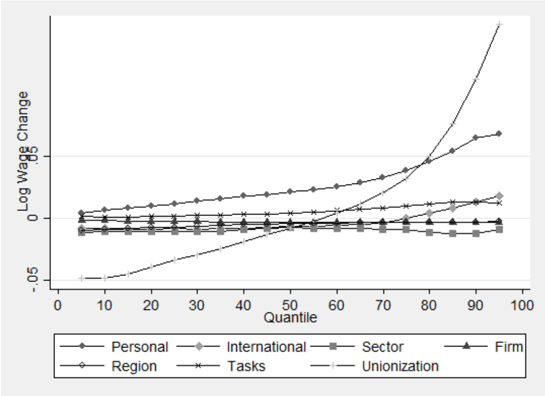
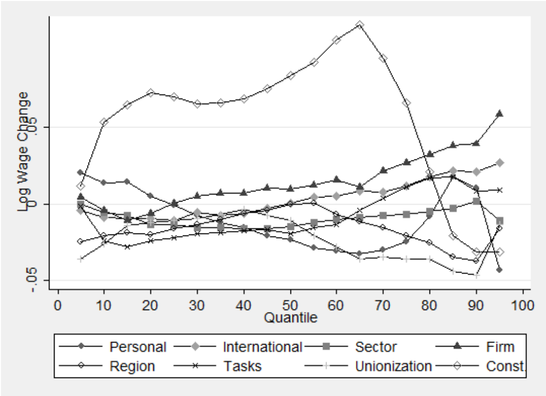



**Figure 2.5** – Wage structure effects 1995-2010




Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

bution and lifted quantiles at the upper end. These effects were so strong that they have the potential to account for much of the overall inequality change. Second to effects of de-unionization, compositional effects of changes in personal characteristics also played an important role. This was particularly true for quantiles in the upper half of the distribution of hourly wages, which significantly gained. This result is consistent with the population aging and educational upgrading described in the previous section. The composition effects of all other groups of covariates were relatively modest, although we observe some increases in unconditional quantiles in the upper quarter of the distribution associated with internationalization, and very modest changes in the upper half of the distribution related to task compositions.

Figure 2.5 provides the break-down of wage structure effects that are related to the different groups of covariates considered by us. These effects are less smooth than the composition effects, and some of them counteract each other. In particular, wage structure effects related to firm characteristics and internationalization tended to favor higher quantiles, while those related to region, unionization and personal characteristics were detrimental for higher quantiles. For tasks, we observe small effects whose patterns are consistent with the polarization hypothesis, i.e. the middle of the distribution lost compared to the bottom and the top of the distribution. Importantly, all of these effects were dominated by general wage structure effects represented by the constant of the RIF regression. As discussed above, the regression constant represents changes in the wage structure that cannot be attributed to particular groups of covariates or that may be related to factors not included as covariates in the analysis. According to figure 2.5, these general wage structure effects were such that the upper middle part of the distribution gained, while the very top part suffered losses.

**Table 2.3** – Aggregated RIF-decompositions 1995-2010

| Inequality measure | 85-15 | 85-50 | 50-15 | Gini | Logvar | 99-90 | 99-95 | 95-90 |
|---|---|---|---|---|---|---|---|---|
| *Total change* | 19.21*** | 8.93*** | 10.28*** | 4.61*** | 6.21*** | 7.26*** | 6.23*** | 1.03** |
| | (0.93) | (0.73) | (0.64) | (0.20) | (0.27) | (1.00) | (0.89) | (0.45) |
| *Total Composition* | 19.83*** | 13.68*** | 6.14*** | 6.09*** | 7.06*** | 5.50*** | 0.06 | 5.44*** |
| | (0.67) | (0.55) | (0.48) | (0.17) | (0.21) | (0.82) | (0.68) | (0.39) |
| Personal | 4.63*** | 3.32*** | 1.32*** | 1.38*** | 1.61*** | 0.70** | 0.41* | 0.29** |
| | (0.26) | (0.18) | (0.13) | (0.06) | (0.08) | (0.27) | (0.24) | (0.15) |
| International | 1.72*** | 1.49*** | 0.24 | 0.47*** | 0.51*** | -0.46 | -0.94** | 0.49*** |
| | (0.38) | (0.33) | (0.26) | (0.09) | (0.11) | (0.49) | (0.44) | (0.18) |
| Sector | -0.13 | -0.45** | 0.31 | 0.06 | 0.09 | 1.29*** | 1.00*** | 0.29* |
| | (0.27) | (0.23) | (0.29) | (0.07) | (0.09) | (0.41) | (0.33) | (0.15) |
| Firm | -0.13 | -0.02 | -0.11 | -0.01 | -0.03 | 0.27 | 0.29 | -0.02 |
| | (0.15) | (0.09) | (0.16) | (0.03) | (0.04) | (0.24) | (0.21) | (0.07) |
| Region | 0.48** | 0.08 | 0.40** | 0.08* | 0.14** | -0.05 | -0.12 | 0.06 |
| | (0.22) | (0.10) | (0.17) | (0.05) | (0.07) | (0.15) | (0.12) | (0.06) |
| Task | 1.18*** | 0.86*** | 0.31*** | 0.30*** | 0.31*** | -0.03 | 0.03 | -0.06 |
| | (0.20) | (0.14) | (0.08) | (0.05) | (0.06) | (0.13) | (0.11) | (0.06) |
| Unionization | 12.08*** | 8.41*** | 3.67*** | 3.82*** | 4.44*** | 3.77*** | -0.61 | 4.38*** |
| | (0.42) | (0.31) | (0.18) | (0.11) | (0.13) | (0.53) | (0.44) | (0.25) |
| *Total Wage Structure* | 0.15 | -3.21*** | 3.36*** | 0.04 | 0.55** | 6.85*** | 5.52*** | 1.33** |
| | (0.82) | (0.67) | (0.63) | (0.18) | (0.23) | (1.52) | (1.41) | (0.59) |
| Personal | 0.30 | 4.11*** | -3.80*** | -0.68** | -1.11*** | -8.82*** | -3.48 | -5.34*** |
| | (1.03) | (0.91) | (0.66) | (0.28) | (0.36) | (2.39) | (2.29) | (1.02) |
| International | 3.10*** | 2.11*** | 0.99 | 0.65*** | 0.81*** | 0.14 | -0.42 | 0.56 |
| | (0.85) | (0.66) | (0.61) | (0.22) | (0.29) | (1.59) | (1.33) | (0.63) |
| Sector | 0.41 | 1.17 | -0.76 | -0.04 | -0.11 | -4.04** | -2.77** | -1.28 |
| | (0.89) | (0.82) | (0.55) | (0.25) | (0.31) | (1.64) | (1.27) | (0.86) |
| Firm | 4.90** | 2.85 | 2.05 | 1.26** | 1.59** | 5.72** | 3.74 | 1.97 |
| | (2.47) | (1.86) | (1.48) | (0.57) | (0.78) | (2.73) | (2.57) | (1.20) |
| Region | -1.60* | -3.41*** | 1.81*** | 0.01 | 0.02 | 4.62*** | 2.55** | 2.07*** |
| | (0.88) | (0.70) | (0.65) | (0.19) | (0.26) | (1.35) | (1.13) | (0.52) |
| Tasks | 4.56*** | 3.74*** | 0.82 | 0.73*** | 0.77*** | -2.83* | -2.90* | 0.07 |
| | (1.03) | (0.90) | (0.69) | (0.23) | (0.29) | (1.69) | (1.49) | (0.80) |
| Unionization | -2.98** | -3.34** | 0.37 | -0.36 | -0.26 | 5.25*** | 1.99 | 3.27*** |
| | (1.42) | (1.51) | (1.11) | (0.29) | (0.39) | (1.45) | (1.22) | (0.86) |
| Constant | -8.54** | -10.42*** | 1.88 | -1.53* | -1.15 | 6.82 | 6.81 | 0.00 |
| | (3.52) | (3.05) | (2.19) | (0.81) | (1.11) | (6.15) | (4.97) | (2.43) |
| *Specification Error* | -0.85 | -0.18 | -0.66* | -1.29*** | -1.27*** | -3.35*** | 1.50 | -4.85*** |
| | (0.61) | (0.50) | (0.36) | (0.12) | (0.15) | (1.02) | (1.07) | (0.55) |
| *Reweighting Error* | 0.08 | -1.36*** | 1.44*** | -0.23*** | -0.13 | -1.73*** | -0.84*** | -0.89*** |
| | (0.39) | (0.29) | (0.21) | (0.09) | (0.11) | (0.30) | (0.23) | (0.12) |

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.
Log wage differentials×100. ***/**/* statistically significant at 1%/5%/10%-level.
Bootstrapped standard errors clustered at establishment level in parentheses (100 replications).

In table 2.3, we provide a detailed break-down of the importance of the different factors for the overall change in inequality. Consistent with the graphical analysis, the numbers show that composition effects fully accounted for the overall inequality change, while wage structure effects compensated each other, resulting in a combined effect of zero. For our main inequality measure, the 85-15 log wage differential, the strongest composition effects came from de-unionization (12.08 out of 19.83 points) and from personal characteristics (4.63 out of 19.83 points). Some smaller compositional effects were contributed by internationalization (1.72 out of 19.83 points) and by shifts in occupational tasks (1.18 out of 19.83 points). Turning to wage structure effects, there were inequality increasing wage structure effects coming from internationalization (3.10 points), firm differences (4.90 points) and tasks (4.56 points). However, these were fully compensated by inequality reducing wage structure effects related to union pay schemes (-2.98 points), regional convergence (-1.60 points), and the RIF constant (-8.54 points). As explained above, the latter represent general wage structure effects that cannot be attributed to any particular group of covariates. The results for the Gini coefficient and the variance of logs generally reproduce the results for the 85-15 log wage gap (columns five and six of table 2.3).

Distinguishing between effects on the upper half (85-50 log wage differential) and on the lower half of the distribution (50-15 log wage differential), we find that the same groups of covariates generally turn out significant, but that the effects in the upper half of the distribution generally dominate. This is also true of the strong compositional effects of de-unionization, contrary to the results in Dustmann et al. (2009, 2014), who found that de-unionization affected mainly the bottom of the distribution. Below, we investigate differences between Dustmann et al. (2009, 2014) and our results in more detail, and provide an explanation why de-unionization affected the whole of the distribution rather than just the bottom.

The last three columns of table 2.3 display the results for top 10 percent of the distribution. As shown earlier, most of the inequality increase occurred at the very top, i.e. within the top 5 percent. Compared to the rest of the distribution, we find weaker composition effects and much stronger unexplained wage structure effects. Overall, the patterns found for the top 10 percent of the distribution look more erratic and less precisely estimated. Also, specification errors are larger than in the main part of the distribution. The main conclusion is that the factors responsible for changes in the main part of the distribution do not explain changes at the top.

### 2.5.4 How do data features drive the results?

One of the contributions of this paper is to use a data set that is different from most of the data sets used in the literature on changes in the German wage distribution. As explained above, our data have a number of features that are not available in the administrative data often used for Germany. These include i) the availability of hourly wages, ii) no top censoring at the social security contributions threshold, and iii) the availability of information on union-determined pay at the individual level. The purpose of this section is to artificially impose in our data set the restrictions present in the administrative data in order to see how this influences the results. This will allow us to assess differences between our results and those reported in the literature. As most of the literature has focussed on West Germany, we also include a variation that uses West German workers only. Taken together, we consider the following variations: i) we consider daily earnings (as in the administrative data) rather than hourly wages, ii) we consider daily earnings and in addition artificially censor our wage information at the social security contributions ceiling,[11] iii) we only use the West German part of our sample, and finally, iv) we include union coverage status at the *firm* rather than at the *individual* level.

[11] For this variation, we proceed exactly as it is done in the literature using administrative data, i.e. we impute wages above the social security contributions ceiling based on the procedure described in Gartner (2005).

**Table 2.4** – Aggregated RIF-decompositions 1995-2010, alternative specifications (main specification in bold face)

| Inequality measure | 85-15 | 85-50 | 50-15 | Gini | Logvar | 99-90 | 99-95 | 95-90 |
|---|---|---|---|---|---|---|---|---|
| **Total composition** | **19.83** | **13.68** | **6.14** | **6.09** | **7.06** | **5.50** | **0.06** | **5.44** |
| Spec. Daily earnings | 18.14 | 12.85 | 5.29 | 5.63 | 6.44 | 4.37 | -0.51 | 4.88 |
| Spec. Daily earnings censored | 18.29 | 13.00 | 5.29 | 4.71 | 5.50 | -3.36 | -2.46 | -0.90 |
| Spec. West Germany | 18.75 | 14.33 | 4.42 | 5.89 | 6.40 | 4.76 | 0.22 | 4.54 |
| Spec. Firm-level unionization | 12.46 | 7.83 | 4.63 | 3.69 | 4.32 | 4.13 | 1.51 | 2.62 |
| **Composition Personal** | **4.63** | **3.32** | **1.32** | **1.38** | **1.61** | **0.70** | **0.41** | **0.29** |
| Spec. Daily earnings | 4.32 | 2.97 | 1.34 | 1.30 | 1.50 | 0.70 | 0.15 | 0.56 |
| Spec. Daily earnings censored | 4.38 | 3.03 | 1.34 | 1.38 | 1.58 | 1.16 | 0.40 | 0.76 |
| Spec. West Germany | 5.49 | 3.88 | 1.61 | 1.56 | 1.72 | 0.56 | 0.58 | -0.02 |
| Spec. Firm-level unionization | 4.97 | 3.56 | 1.41 | 1.49 | 1.74 | 0.91 | 0.42 | 0.49 |
| **Composition International** | **1.72** | **1.49** | **0.24** | **0.47** | **0.51** | **-0.46** | **-0.94** | **0.49** |
| Spec. Daily earnings | 1.36 | 1.33 | 0.03 | 0.43 | 0.45 | -0.02 | -0.67 | 0.65 |
| Spec. Daily earnings censored | 1.37 | 1.34 | 0.03 | 0.39 | 0.41 | -0.19 | -0.22 | 0.03 |
| Spec. West Germany | 1.23 | 1.46 | -0.23 | 0.37 | 0.38 | -0.47 | -0.85 | 0.38 |
| Spec. Firm-level unionization | 2.27 | 1.95 | 0.32 | 0.67 | 0.73 | -0.23 | -0.97 | 0.74 |
| **Composition Sector** | **-0.13** | **-0.45** | **0.31** | **0.06** | **0.09** | **1.29** | **1.00** | **0.29** |
| Spec. Daily earnings | -0.10 | -0.20 | 0.10 | 0.06 | 0.08 | 1.29 | 1.07 | 0.22 |
| Spec. Daily earnings censored | -0.12 | -0.22 | 0.10 | 0.01 | 0.02 | 0.57 | 0.37 | 0.20 |
| Spec. West Germany | -0.31 | -0.32 | 0.01 | 0.02 | 0.01 | 1.28 | 0.85 | 0.43 |
| Spec. Firm-level unionization | 0.01 | -0.27 | 0.28 | 0.12 | 0.17 | 1.46 | 1.03 | 0.43 |
| **Composition Firm** | **-0.13** | **-0.02** | **-0.11** | **-0.01** | **-0.03** | **0.27** | **0.29** | **-0.02** |
| Spec. Daily earnings | -0.13 | -0.02 | -0.11 | -0.01 | -0.02 | 0.31 | 0.27 | 0.04 |
| Spec. Daily earnings censored | -0.14 | -0.03 | -0.11 | -0.03 | -0.04 | 0.16 | 0.11 | 0.04 |
| Spec. West Germany | -0.17 | -0.02 | -0.15 | -0.02 | -0.04 | 0.41 | 0.42 | -0.01 |
| Spec. Firm-level unionization | 0.11 | 0.22 | -0.12 | 0.10 | 0.08 | 0.43 | 0.21 | 0.23 |
| **Composition Region** | **0.48** | **0.08** | **0.40** | **0.08** | **0.14** | **-0.05** | **-0.12** | **0.06** |
| Spec. Daily earnings | 0.34 | 0.00 | 0.34 | 0.06 | 0.11 | -0.03 | -0.04 | 0.01 |
| Spec. Daily earnings censored | 0.33 | -0.01 | 0.34 | 0.07 | 0.12 | 0.21 | 0.11 | 0.10 |
| Spec. West Germany | 0.10 | 0.01 | 0.09 | 0.01 | 0.02 | -0.05 | -0.04 | -0.01 |
| Spec. Firm-level unionization | 0.44 | 0.07 | 0.37 | 0.07 | 0.13 | -0.12 | -0.16 | 0.03 |
| **Composition Task** | **1.18** | **0.86** | **0.31** | **0.30** | **0.31** | **-0.03** | **0.03** | **-0.06** |
| Spec. Daily earnings | 1.10 | 0.77 | 0.33 | 0.29 | 0.30 | -0.01 | -0.09 | 0.08 |
| Spec. Daily earnings censored | 1.10 | 0.77 | 0.33 | 0.21 | 0.22 | -0.81 | -0.52 | -0.30 |
| Spec. West Germany | 1.33 | 0.86 | 0.47 | 0.36 | 0.36 | 0.06 | 0.13 | -0.06 |
| Spec. Firm-level unionization | 1.28 | 0.94 | 0.34 | 0.35 | 0.36 | 0.09 | 0.04 | 0.05 |
| **Composition Unionization** | **12.08** | **8.41** | **3.67** | **3.82** | **4.44** | **3.77** | **-0.61** | **4.38** |
| Spec. Daily earnings | 11.26 | 8.00 | 3.26 | 3.51 | 4.04 | 2.12 | -1.20 | 3.32 |
| Spec. Daily earnings censored | 11.37 | 8.11 | 3.26 | 2.67 | 3.18 | -4.45 | -2.72 | -1.74 |
| Spec. West Germany | 11.08 | 8.47 | 2.61 | 3.60 | 3.95 | 2.97 | -0.87 | 3.84 |
| Spec. Firm-level unionization | 3.37 | 1.36 | 2.01 | 0.90 | 1.12 | 1.59 | 0.93 | 0.65 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Total wage structure** | **0.15** | **-3.21** | **3.36** | **0.04** | **0.55** | **6.85** | **5.52** | **1.33** |
| Spec. Daily earnings | 0.77 | -2.02 | 2.78 | 0.56 | 1.12 | 8.18 | 6.39 | 1.80 |
| Spec. Daily earnings censored | -1.06 | -3.85 | 2.78 | 1.84 | 2.54 | 14.49 | 9.30 | 5.19 |
| Spec. West Germany | 0.85 | -4.46 | 5.31 | 0.15 | 0.80 | 6.76 | 6.27 | 0.49 |
| Spec. Firm-level unionization | 7.08 | 2.18 | 4.90 | 1.65 | 2.57 | 6.06 | 5.15 | 0.91 |
| **Wage structure Personal** | **0.30** | **4.11** | **-3.80** | **-0.68** | **-1.11** | **-8.82** | **-3.48** | **-5.34** |
| Spec. Daily earnings | -1.53 | 0.71 | -2.23 | -0.99 | -1.45 | -5.81 | -1.73 | -4.08 |
| Spec. Daily earnings censored | -1.67 | 0.56 | -2.23 | -1.84 | -2.72 | -5.62 | 1.02 | -6.64 |
| Spec. West Germany | 2.82 | 4.80 | -1.99 | -0.26 | -0.54 | -9.38 | -5.80 | -3.58 |
| Spec. Firm-level unionization | -3.59 | 0.42 | -4.02 | -0.57 | -1.53 | -3.90 | -3.12 | -0.78 |
| **Wage structure International** | **3.10** | **2.11** | **0.99** | **0.65** | **0.81** | **0.14** | **-0.42** | **0.56** |
| Spec. Daily earnings | 1.94 | 1.35 | 0.59 | 0.48 | 0.59 | 0.96 | 0.36 | 0.60 |
| Spec. Daily earnings censored | 2.12 | 1.53 | 0.59 | 0.25 | 0.34 | -1.45 | -1.32 | -0.14 |
| Spec. West Germany | 1.67 | 1.26 | 0.41 | 0.37 | 0.46 | 0.23 | -0.04 | 0.26 |
| Spec. Firm-level unionization | 4.73 | 3.61 | 1.12 | 1.13 | 1.35 | -1.19 | -1.61 | 0.43 |
| **Wage structure Sector** | **0.41** | **1.17** | **-0.76** | **-0.04** | **-0.11** | **-4.04** | **-2.77** | **-1.28** |
| Spec. Daily earnings | -0.17 | 0.52 | -0.70 | -0.17 | -0.26 | -3.90 | -2.46 | -1.44 |
| Spec. Daily earnings censored | -0.30 | 0.40 | -0.70 | 0.16 | 0.07 | 1.02 | 1.09 | -0.07 |
| Spec. West Germany | 1.69 | 1.19 | 0.50 | 0.19 | 0.20 | -2.83 | -1.26 | -1.57 |
| Spec. Firm-level unionization | -0.59 | 0.32 | -0.91 | -0.10 | -0.27 | -1.72 | -2.05 | 0.34 |
| **Wage structure Firm** | **4.90** | **2.85** | **2.05** | **1.26** | **1.59** | **5.72** | **3.74** | **1.97** |
| Spec. Daily earnings | 3.01 | 1.48 | 1.53 | 0.94 | 1.15 | 4.01 | 2.23 | 1.78 |
| Spec. Daily earnings censored | 3.04 | 1.51 | 1.53 | 0.67 | 0.93 | -4.48 | -5.00 | 0.52 |
| Spec. West Germany | 2.00 | 1.29 | 0.71 | 0.78 | 1.06 | 6.94 | 5.30 | 1.65 |
| Spec. Firm-level unionization | 5.38 | 2.92 | 2.46 | 1.07 | 1.55 | 2.07 | 0.24 | 1.83 |
| **Wage structure Region** | **-1.60** | **-3.41** | **1.81** | **0.01** | **0.02** | **4.62** | **2.55** | **2.07** |
| Spec. Daily earnings | -1.51 | -1.06 | -0.45 | -0.06 | -0.23 | 4.02 | 2.58 | 1.44 |
| Spec. Daily earnings censored | -1.23 | -0.78 | -0.45 | 0.09 | -0.10 | 0.68 | -0.32 | 1.00 |
| Spec. West Germany | -3.63 | -3.80 | 0.17 | -0.60 | -0.65 | 3.22 | 2.19 | 1.03 |
| Spec. Firm-level unionization | -0.46 | -1.59 | 1.13 | 0.26 | 0.14 | 2.99 | 1.95 | 1.05 |
| **Wage structure Task** | **4.56** | **3.74** | **0.82** | **0.73** | **0.77** | **-2.83** | **-2.90** | **0.07** |
| Spec. Daily earnings | 2.54 | 2.82 | -0.28 | 0.54 | 0.46 | -2.26 | -1.82 | -0.44 |
| Spec. Daily earnings censored | 1.71 | 1.99 | -0.28 | 0.04 | -0.15 | 2.04 | 2.62 | -0.58 |
| Spec. West Germany | 3.96 | 3.06 | 0.89 | 0.88 | 1.02 | -2.11 | -1.91 | -0.21 |
| Spec. Firm-level unionization | 4.13 | 3.30 | 0.82 | 0.86 | 0.62 | -1.60 | -2.08 | 0.48 |
| **Wage structure Unionization** | **-2.98** | **-3.34** | **0.37** | **-0.36** | **-0.26** | **5.25** | **1.99** | **3.27** |
| Spec. Daily earnings | -1.65 | -1.12 | -0.53 | -0.17 | -0.04 | 2.89 | 1.28 | 1.61 |
| Spec. Daily earnings censored | -1.41 | -0.87 | -0.53 | 0.34 | 0.61 | 1.92 | 0.27 | 1.64 |
| Spec. West Germany | -3.94 | -1.97 | -1.97 | -0.49 | -0.58 | 4.04 | 1.79 | 2.25 |
| Spec. Firm-level unionization | 3.99 | 2.42 | 1.57 | 0.51 | 0.91 | -4.48 | -2.63 | -1.84 |
| **Wage structure Constant** | **-8.54** | **-10.42** | **1.88** | **-1.53** | **-1.15** | **6.82** | **6.81** | **0.00** |
| Spec. Daily earnings | -1.86 | -6.72 | 4.86 | -0.02 | 0.89 | 8.27 | 5.94 | 2.33 |
| Spec. Daily earnings censored | -3.33 | -8.19 | 4.86 | 2.13 | 3.55 | 20.39 | 10.94 | 9.45 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Spec. West Germany | -3.71 | -10.29 | 6.58 | -0.73 | -0.15 | 6.66 | 6.01 | 0.65 |
| Spec. Firm-level unionization | -6.51 | -9.22 | 2.71 | -1.51 | -0.20 | 13.87 | 14.46 | -0.59 |
| **Specification Error** | **-0.85** | **-0.18** | **-0.66** | **-1.29** | **-1.27** | **-3.35** | **1.50** | **-4.85** |
| Spec. Daily earnings | -1.40 | -0.63 | -0.77 | -1.31 | -1.28 | -2.59 | 1.61 | -4.20 |
| Spec. Daily earnings censored | 0.26 | 1.04 | -0.77 | -1.12 | -1.12 | 0.81 | 1.80 | -0.99 |
| Spec. West Germany | -1.59 | -0.41 | -1.17 | -1.29 | -1.19 | -1.54 | 1.54 | -3.08 |
| Spec. Firm-level unionization | -1.32 | -0.37 | -0.95 | -0.71 | -0.82 | -0.72 | 0.82 | -1.54 |
| **Reweighting Error** | **0.08** | **-1.36** | **1.44** | **-0.23** | **-0.13** | **-1.73** | **-0.84** | **-0.89** |
| Spec. Daily earnings | 0.18 | -0.69 | 0.86 | -0.12 | -0.07 | -1.50 | -0.95 | -0.55 |
| Spec. Daily earnings censored | 0.20 | -0.67 | 0.86 | -0.04 | 0.03 | -0.51 | -0.15 | -0.36 |
| Spec. West Germany | 0.56 | -1.23 | 1.79 | -0.12 | -0.00 | -1.90 | -0.87 | -1.03 |
| Spec. Firm-level unionization | 0.99 | -0.71 | 1.70 | -0.02 | 0.13 | -2.20 | -1.25 | -0.96 |

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

The results are shown in table 2.4. A first important conclusion from this table is that the practice of the previous literature to focus on daily earnings instead of hourly wages does not change the results in any important way.[12] Similarly, introducing in addition artificial censoring at the social security contributions threshold in combination with an imputation procedure above this threshold, does not change the results in important ways, as long as one only considers the distribution up to the 85th wage percentile. However, the right hand columns of table 2.4 warn that this is not true for the range above the 85th percentile or if one uses inequality measures that include the whole range of the distribution such as the Gini.[13] As a next variation, we restrict our estimates to West Germany. Considering only West Germany induces some smaller changes in the wage structure effects (esp. for personal characteristics), but does not challenge in any way the strong composition effects contributed by de-unionization and personal characteristics.

Finally, we consider the variation of using *firm level* instead of *individual level* union coverage. This produces substantial differences. In the specification with firm level union coverage, the compositional de-unionization is drastically reduced (from 12.08 to 3.37 points), and

[12]The only substantial difference between the analysis of hourly vs. daily wages is the RIF-regression constant. In addition, there are minor differences in wage structure effects in the lower and the upper half of the distribution (esp. for personal characteristics).

[13]Additional graphical evidence presented in the appendix suggests that the imputation procedure produces nonsensical patterns for distributional analysis above the 85 percentile (tables A5 to A8).

the wage structure effects for personal characteristics and union coverage are reversed. Note that it is quite plausible that reducing information on union coverage to the firm level shifts explanatory power to coefficients on personal characteristics. Switching to firm-level union status also practically *eliminates* the strong compositional effects of de-unionization found for *the upper half* of the distribution in our original specification (column three of table 2.4).

How can the substantial differences between the results with individual-level and firm-level coverage status be explained? If firm-level union status is used, all individuals in the same firm are assigned an identical union effect, ignoring that not all workers in a given firm are paid according to union pay schemes. In table 2.1, we showed that such workers exist and that their share increased over time. Indeed, the subgroup of workers not paid according to union pay schemes is very diverse, including both high-productivity workers for whom non-coverage is used to pay higher wages, and low-productivity workers for whom non-coverage is used to pay particularly low wages. As shown earlier, both between- and within-group wage differentials are much more pronounced in the group of uncovered than in the group of covered workers (tables A3 and A4 in the appendix). Increasing the very heterogenous portion of workers not covered by union pay schemes will therefore mechanically increase inequality across the whole distribution, not only in the lower part.

We are now in the position to contrast our results with results reported in the literature and to explain observed differences. Using administrative data on daily earnings, Dustmann et al. (2009, 2014) also obtained the result that de-unionization was a leading factor for rising wage inequality in Germany, along with compositional changes in personal characteristics. We have shown in this paper that the use of daily earnings and censored wage information in the administrative data does not compromise the validity of their findings. Moreover, we provide a further validation of some of their conclusions by showing that the compositional effects of de-unionization and personal characteristics also hold in a multivariate setting, i.e. when controlling for a large set of factors at the same time (Dustmann et al., 2009, 2014, only considered one factor at a time, not controlling for other factors). Finally, we show that Dustmann et al. (2009, 2014) underestimate and partly misinterpret the effects of de-unionization by considering only firm-level unionization status which does not show the full extent of the erosion of union coverage, and which misses effects of de-unionization in the upper part of the wage distribution.

Antonczyk et al. (2010) used two waves of the same data set we use in this study and a less extensive set of explanatory factors. Employing a methodology based on sequentially introducing explanatory factors in quantile regressions, they found no leading role for effects of de-unionization. Apart from differences in methodology, we show above that their use of firm-level union information tends to underestimate the full effect of de-unionization. Moreover, they considered only the waves 2001 and 2006, while our results suggest that there were important effects in the other waves not covered by their analysis. Ohlert (2016) analyzed administrative wage data combined with rich firm survey data (the *LIAB* data). He also found no important role for de-unionization but for firm characteristics. Ohlert (2016) employed the regression-based decomposition methodology introduced by Fields (2003) which also controls for many factors at the same time, but which does not make the classical distinction between 'composition' and 'wage structure' effects. His results are therefore hard to compare to ours.

Ehrl (2017) used a very similar method but based on administrative wage data. He concludes that occupational characteristics are important for rising wage inequality, but his analysis does not include information on union coverage, neither at the firm nor at the individual level. Finally, in an analysis similar to ours but based on administrative data, Baumgarten et al. (2018) find important effects for de-unionization but also for sectoral change. The latter result is different from ours, which is probably due to the fact that we observe a narrower range of sectors in our data set than they do in theirs. Our analysis suggests that Baumgarten et al. (2018), just as all the previous literature on wage inequality in Germany, grossly underestimate the leading role of de-unionization as their data does not include information on union coverage at the individual level.

### 2.5.5 What role is left for unobserved firm heterogeneity?

In light of recent contributions focussing on firm and establishment effects (Card et al., 2013, Ohlert, 2016, Barth et al. 2016, Song et al. 2019), one might ask what role is left for between-firm differences that go beyond the differences in observable firm characteristics already included in our analysis. In order to address this question, we carry out the following procedure. First, we obtain cross-sectional firm effects by regressing log hourly wages on our list of observable covariates and a full set of firm dummies. Because of partitioning properties of OLS, this is equivalent to taking the residuals from wage regressions as in table 2.2 and computing average residuals at the firm level. We then consider the distribution of these firm-specific wage effects. In order to assess to what extent rising heterogeneity in firm-specific wage effects contributed to rising wage inequality, we assign to each individual in the wage distribution of a base year the corresponding firm effect in the distribution of the target year, assuming that the individual keeps working at a firm in the same percentile of the distribution of firm effects. We are aware that we are unable to capture changes in sorting of workers to firms as in Card et al. (2013) in this way. Still, our procedure will be informative about the quantitative importance of changes in heterogeneity between firms not captured by our firm level observables (net of additional sorting effects).

**Table 2.5** – Effect of unobserved firm heterogeneity 1995-2010

| Inequality measure | 85-15 | 85-50 | 50-15 | Gini | Logvar |
|---|---|---|---|---|---|
| Total change | 18.43 | 8.83 | 9.59 | 4.44 | 6.35 |
| Unobserved firm heterogeneity | 1.72*** | 0.79*** | 0.93*** | 0.47*** | 0.64*** |
| | (0.17) | (0.09) | (0.10) | (0.05) | (0.07) |

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.
Log wage differentials×100. ***/**/* statistically significant at 1%/5%/10%-level.
Bootstrapped standard errors clustered at establishment level (100 replications).

Table 2.5 shows that assigning workers in 1995 their (more heterogenous) firm effects of 2010 increases the 85-15 wage gap by a moderate 1.72 log percentage points. This accounts for some 10 percent of the overall inequality change of 18.43 log percentage points. Overall, we conclude that rising heterogeneity between firms beyond the factors explicitly included in our analysis mattered for rising wage inequality but that its contribution was limited compared to the effects explicitly analyzed in the previous sections.

## 2.6 Summary and discussion

This paper has analyzed the quantitative importance of a large set of explanatory factors for the evolution of the German wage distribution over the period 1995 to 2010. A distinguishing feature of our analysis is that we simultaneously take into account most of the factors considered in the literature so far, and that we base our analysis on different data than used in most of the prominent studies on wage inequality in Germany. In contrast to the administrative data sets usually analyzed for Germany, our data include information on hourly wages instead of daily earnings, is not top-coded, and contain richer information at the individual level, esp. information on individual union coverage.

We explicitly analyze the differences induced by these data features. Our results suggest that analyses based on administrative data are not compromised by the fact that these data typically only report daily earnings rather than hourly wages (which are the best measure of relative prices in the labor market). Similarly, top-coding does in general not invalidate distributional results, as long as they do not include information above the 85th percentile. Using our largely uncensored wage information, we also do not find extreme movements in the upper part of the German wage distribution that is subject to top-coding in administrative data, in contrast to what is known for other countries. This is good news for the users of administrative data which often suffer from limitations due to their administrative purpose. However, an important conclusion from our analysis is that using firm-level information on unionization may miss an important part of the erosion of union coverage if this erosion also takes place *within* firms.

In substantive terms, our study suggests that compositional effects due to de-unionization were by far the most important factor behind recent rises in wage inequality in Germany. Our analysis suggests that the inequality increasing effects of de-unionization were mainly due to the fact that de-unionization shrank the part of the economy in which wages were more compressed. We document this by showing that wages among workers covered by union agreements were much more compressed than among uncovered workers, both along observable characteristics (such as age or education), and along unobservable characteristics (within groups of workers with identical observable characteristics). It is therefore more than plausible that a shift of more than 34 percentage points from the covered to uncovered portion of the workforce led to substantially higher wage inequality. As the second most important factor for changes in the distribution, we measure compositional effects related to personal characteristics, especially workers' age and education. Such effects are consistent with the hypothesis that the increasing demand for higher skills due to SBTC were matched by rising supply for such skills in the form of educational upgrading and population aging (higher age represents higher human capital in the form of richer work experience and acquired skills). This is because in the absence of rising demand due to SBTC, rising supply of high skills would have depressed the wage premia paid for such skills, which one does not observe.[14]

Taken together, our analysis suggests that a large part of changes in the German wage distribution can be explained by compositional changes of the workforce (around 60 percent by de-unionization and around 25 percent by compositional changes in personal characteristics). We do measure some wage structure effects related to internationalization, firm heterogeneity, task changes, and regional wage convergence, but these are smaller in magnitude and tend to compensate each other.

We emphasize that our estimates are certainly not to be interpreted as causal effects. This

[14] In the previous version of this paper, Biewen and Seckler (2017), we present some evidence for excess demand for higher skills.

is for several reasons, one being that the factors in our analysis might be dynamically related to each other. For example, de-unionization might have been a consequence of internationalization (e.g. Dreher and Gaston, 2007). In a previous version of this paper (Biewen and Seckler, 2017), we explicitly considered this possibility by placing union coverage at the end of a sequential conditioning scheme using the method of DiNardo et al. (1996), with the result that it robustly remained the most important explanatory factor. Even if a factor like de-unionization was itself a consequence of another factor, it would still be relevant to see that changes in the distribution were largely mediated by this factor. In a broader perspective and in line with Dustmann et al. (2014), de-unionization might have been a way for the German economy to arrive at a wage structure consistent with the needs of the economy. Our finding that the decline in union coverage was a major determinant of the recent rise in wage inequality is also consistent with the fact that de-unionization substantially slowed down towards the end of our observation period, and that newer data for Germany indicate no further increases in wage inequality after 2011 (Möller, 2016).

# Appendix A

**Table A1** – Descriptive statistics

| Variable | 1995 | | 2001 | | 2006 | | 2010 | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std | Mean | Std | Mean | Std | Mean | Std |
| Personal | | | | | | | | |
| Age 20-25 | 0.081 | 0.273 | 0.071 | 0.256 | 0.069 | 0.253 | 0.069 | 0.254 |
| Age 26-30 | 0.152 | 0.359 | 0.106 | 0.308 | 0.095 | 0.294 | 0.100 | 0.300 |
| Age 31-35 | 0.176 | 0.381 | 0.169 | 0.375 | 0.122 | 0.327 | 0.108 | 0.311 |
| Age 36-40 | 0.150 | 0.357 | 0.190 | 0.393 | 0.178 | 0.382 | 0.132 | 0.338 |
| Age 41-45 | 0.132 | 0.339 | 0.161 | 0.368 | 0.191 | 0.393 | 0.184 | 0.387 |
| Age 46-50 | 0.109 | 0.312 | 0.134 | 0.340 | 0.158 | 0.365 | 0.183 | 0.387 |
| Age 51-55 | 0.121 | 0.326 | 0.106 | 0.308 | 0.121 | 0.326 | 0.141 | 0.348 |
| Age 56-60 | 0.078 | 0.268 | 0.063 | 0.242 | 0.066 | 0.249 | 0.083 | 0.276 |
| Tenure 0-5 | 0.403 | 0.491 | 0.406 | 0.491 | 0.360 | 0.480 | 0.349 | 0.477 |
| Tenure 6-10 | 0.185 | 0.388 | 0.188 | 0.391 | 0.205 | 0.404 | 0.189 | 0.392 |
| Tenure 11-15 | 0.114 | 0.317 | 0.143 | 0.351 | 0.138 | 0.345 | 0.143 | 0.350 |
| Tenure 16-20 | 0.100 | 0.300 | 0.084 | 0.278 | 0.115 | 0.319 | 0.117 | 0.322 |
| Tenure 21-25 | 0.085 | 0.279 | 0.076 | 0.265 | 0.071 | 0.257 | 0.086 | 0.280 |
| Tenure >25 | 0.113 | 0.316 | 0.103 | 0.303 | 0.110 | 0.313 | 0.116 | 0.321 |
| Lower/middle secondary w/o vocational training | 0.140 | 0.347 | 0.127 | 0.333 | 0.104 | 0.305 | 0.097 | 0.296 |
| Lower/middle secondary w/ vocational training | 0.711 | 0.453 | 0.680 | 0.467 | 0.663 | 0.473 | 0.644 | 0.479 |
| Upper secondary (German high school equiv.) | 0.026 | 0.158 | 0.039 | 0.195 | 0.051 | 0.219 | 0.053 | 0.224 |
| University of Applied Science (Fachhochschule) | 0.043 | 0.203 | 0.050 | 0.218 | 0.052 | 0.221 | 0.052 | 0.222 |
| University | 0.032 | 0.177 | 0.045 | 0.207 | 0.051 | 0.220 | 0.055 | 0.227 |
| Missing information | 0.048 | 0.213 | 0.059 | 0.235 | 0.080 | 0.272 | 0.100 | 0.300 |
| Non-skilled blue collar | 0.218 | 0.413 | 0.236 | 0.424 | 0.221 | 0.415 | 0.236 | 0.424 |
| Skilled blue collar and foremen | 0.462 | 0.499 | 0.396 | 0.489 | 0.389 | 0.487 | 0.381 | 0.486 |
| White collar | 0.321 | 0.467 | 0.369 | 0.482 | 0.390 | 0.488 | 0.383 | 0.486 |
| Internationalization | | | | | | | | |
| No Exports | 0.474 | 0.499 | 0.475 | 0.499 | 0.470 | 0.499 | 0.446 | 0.497 |
| Export share 1-25% | 0.292 | 0.455 | 0.221 | 0.415 | 0.151 | 0.358 | 0.190 | 0.392 |
| Export share 26-50% | 0.077 | 0.267 | 0.071 | 0.257 | 0.038 | 0.191 | 0.130 | 0.336 |
| Export share 51-100% | 0.157 | 0.363 | 0.233 | 0.423 | 0.341 | 0.474 | 0.234 | 0.423 |
| Offshoring (0-100%) | 4.020 | 2.311 | 4.177 | 2.380 | 4.405 | 2.771 | 4.062 | 2.841 |
| Imports of consumption goods (0-100%) | 3.267 | 5.138 | 3.274 | 4.764 | 2.993 | 4.816 | 3.220 | 5.116 |
| Sector | | | | | | | | |
| Mining and other quarring | 0.021 | 0.143 | 0.013 | 0.111 | 0.011 | 0.103 | 0.008 | 0.087 |
| Food products, beverages, tobacco | 0.038 | 0.192 | 0.039 | 0.193 | 0.038 | 0.191 | 0.043 | 0.202 |
| Textiles | 0.015 | 0.122 | 0.010 | 0.101 | 0.009 | 0.094 | 0.008 | 0.090 |
| Wood | 0.012 | 0.109 | 0.013 | 0.111 | 0.011 | 0.105 | 0.010 | 0.100 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Paper | 0.014 | 0.119 | 0.014 | 0.119 | 0.013 | 0.113 | 0.014 | 0.118 |
| Printing | 0.020 | 0.141 | 0.023 | 0.150 | 0.020 | 0.139 | 0.011 | 0.103 |
| Coke and petroleum products | 0.003 | 0.051 | 0.003 | 0.050 | 0.002 | 0.048 | 0.002 | 0.042 |
| Chemicals | 0.045 | 0.207 | 0.042 | 0.200 | 0.038 | 0.190 | 0.039 | 0.194 |
| Rubber, plastic | 0.034 | 0.180 | 0.036 | 0.187 | 0.036 | 0.186 | 0.037 | 0.190 |
| Non-metallic products | 0.028 | 0.164 | 0.026 | 0.158 | 0.021 | 0.144 | 0.021 | 0.144 |
| Basic metals | 0.033 | 0.177 | 0.032 | 0.175 | 0.030 | 0.171 | 0.031 | 0.172 |
| Fabricated metal products | 0.062 | 0.241 | 0.070 | 0.255 | 0.069 | 0.254 | 0.069 | 0.253 |
| Computer, electronic, optical products | 0.035 | 0.185 | 0.042 | 0.201 | 0.040 | 0.197 | 0.026 | 0.159 |
| Electrical equipment | 0.030 | 0.172 | 0.036 | 0.186 | 0.038 | 0.192 | 0.035 | 0.185 |
| Machinery and equipment | 0.119 | 0.323 | 0.108 | 0.310 | 0.125 | 0.331 | 0.106 | 0.307 |
| Motor vehicles, trailers | 0.055 | 0.229 | 0.068 | 0.252 | 0.089 | 0.285 | 0.070 | 0.255 |
| Other transport equipment | 0.024 | 0.154 | 0.016 | 0.126 | 0.017 | 0.128 | 0.013 | 0.113 |
| Furniture etc | 0.021 | 0.142 | 0.021 | 0.143 | 0.017 | 0.129 | 0.048 | 0.213 |
| Electricity, water, recycling | 0.028 | 0.165 | 0.026 | 0.159 | 0.030 | 0.171 | 0.045 | 0.207 |
| Construction | 0.176 | 0.381 | 0.132 | 0.338 | 0.103 | 0.304 | 0.113 | 0.316 |
| Trade of vehicles | 0.032 | 0.176 | 0.038 | 0.192 | 0.044 | 0.206 | 0.046 | 0.210 |
| Wholesale trade | 0.076 | 0.266 | 0.086 | 0.281 | 0.098 | 0.297 | 0.113 | 0.317 |
| Retail trade | 0.040 | 0.196 | 0.045 | 0.208 | 0.046 | 0.209 | 0.039 | 0.194 |
| Finance and insurance | 0.039 | 0.193 | 0.062 | 0.241 | 0.054 | 0.227 | 0.054 | 0.226 |
| Firm | | | | | | | | |
| Firmsize 10-19 | 0.074 | 0.261 | 0.079 | 0.269 | 0.071 | 0.257 | 0.073 | 0.260 |
| Firmsize 20-49 | 0.151 | 0.358 | 0.166 | 0.372 | 0.150 | 0.357 | 0.161 | 0.368 |
| Firmsize 50-99 | 0.134 | 0.341 | 0.125 | 0.331 | 0.130 | 0.336 | 0.125 | 0.330 |
| Firmsize 100-199 | 0.125 | 0.330 | 0.133 | 0.339 | 0.134 | 0.341 | 0.131 | 0.337 |
| Firmsize 200-499 | 0.170 | 0.375 | 0.159 | 0.366 | 0.161 | 0.367 | 0.159 | 0.365 |
| Firmsize 500-999 | 0.097 | 0.296 | 0.103 | 0.304 | 0.106 | 0.308 | 0.094 | 0.292 |
| Firmsize >1000 | 0.250 | 0.433 | 0.236 | 0.424 | 0.247 | 0.432 | 0.258 | 0.437 |
| State-owned | 0.046 | 0.210 | 0.023 | 0.150 | 0.020 | 0.141 | 0.037 | 0.188 |
| Region | | | | | | | | |
| Schleswig-Holstein | 0.026 | 0.158 | 0.023 | 0.149 | 0.024 | 0.152 | 0.026 | 0.160 |
| Hamburg | 0.023 | 0.148 | 0.023 | 0.149 | 0.021 | 0.144 | 0.022 | 0.145 |
| Lower Saxony | 0.076 | 0.265 | 0.080 | 0.271 | 0.083 | 0.276 | 0.087 | 0.282 |
| Bremen | 0.011 | 0.103 | 0.010 | 0.100 | 0.009 | 0.095 | 0.010 | 0.098 |
| North Rhine-Westphalia | 0.254 | 0.436 | 0.262 | 0.440 | 0.235 | 0.424 | 0.223 | 0.416 |
| Hesse | 0.081 | 0.273 | 0.076 | 0.265 | 0.079 | 0.269 | 0.075 | 0.264 |
| Rhineland-Palatinate | 0.043 | 0.203 | 0.051 | 0.219 | 0.045 | 0.208 | 0.048 | 0.213 |
| Baden-Wuerttemberg | 0.163 | 0.370 | 0.171 | 0.376 | 0.177 | 0.381 | 0.167 | 0.373 |
| Bavaria | 0.166 | 0.372 | 0.166 | 0.372 | 0.182 | 0.386 | 0.183 | 0.387 |
| Saarland | 0.016 | 0.125 | 0.013 | 0.114 | 0.016 | 0.124 | 0.014 | 0.118 |
| Berlin | 0.027 | 0.163 | 0.019 | 0.137 | 0.017 | 0.129 | 0.021 | 0.144 |
| Brandenburg | 0.021 | 0.142 | 0.017 | 0.128 | 0.018 | 0.133 | 0.023 | 0.148 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Mecklenburg-West Pomerania | 0.013 | 0.115 | 0.012 | 0.107 | 0.012 | 0.107 | 0.012 | 0.109 |
| Saxony | 0.037 | 0.189 | 0.038 | 0.191 | 0.042 | 0.200 | 0.043 | 0.202 |
| Saxony-Anhalt | 0.023 | 0.151 | 0.020 | 0.139 | 0.020 | 0.141 | 0.022 | 0.147 |
| Thuringia | 0.019 | 0.137 | 0.021 | 0.144 | 0.022 | 0.145 | 0.025 | 0.155 |
| Tasks | | | | | | | | |
| Share of analytical tasks | 0.276 | 0.142 | 0.290 | 0.150 | 0.293 | 0.151 | 0.290 | 0.151 |
| Share of interactive tasks | 0.212 | 0.118 | 0.221 | 0.123 | 0.223 | 0.126 | 0.224 | 0.126 |
| Share of manual tasks | 0.513 | 0.232 | 0.489 | 0.245 | 0.485 | 0.247 | 0.486 | 0.247 |
| Unionization | | | | | | | | |
| No union coverage | 0.265 | 0.441 | 0.388 | 0.487 | 0.550 | 0.497 | 0.610 | 0.488 |
| Sectoral bargaining | 0.697 | 0.460 | 0.569 | 0.495 | 0.401 | 0.500 | 0.357 | 0.479 |
| Firm bargaining | 0.038 | 0.191 | 0.043 | 0.203 | 0.041 | 0.198 | 0.033 | 0.178 |
| Observations | 592.198 | | 359.495 | | 533.497 | | 438.352 | |

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations. Weighted data.

**Table A2** – Mapping of activities into task indicators

| Task | Activity |
|---|---|
| Analytical | Researching, evaluating, measuring |
| | Designing, planning, sketching |
| | Correcting texts or data |
| | Programming |
| | Executing laws or interpreting rules |
| Manual | Equipping or operating machinery |
| | Repairing, renovating, reconstructing |
| | Manufacturing, installing or constructing |
| | Nursing, serving, accomodating |
| | Transporting |
| Interactive | Selling, buying, advertising |
| | Teaching or training |
| | Negotiating |
| | Employing, managing personnel, organizing |

**Figure A1** – Development of inequality, comparison GSES vs. SIAB (85-15 log wage gap)

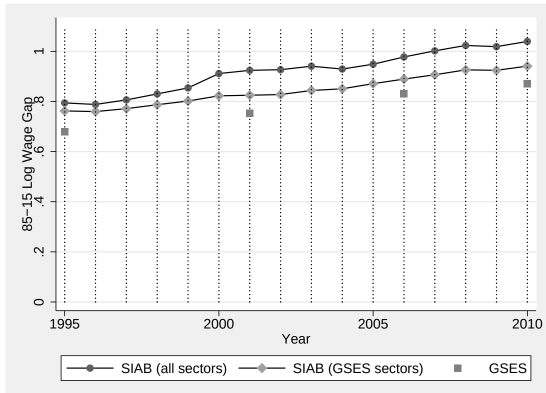



**Figure A2** – Development of inequality, comparison GSES vs. SIAB (85-50 log wage gap)

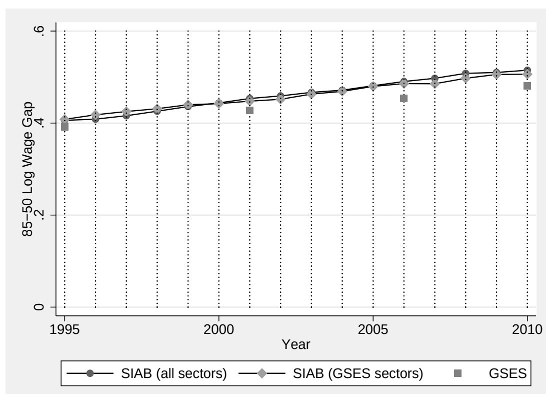
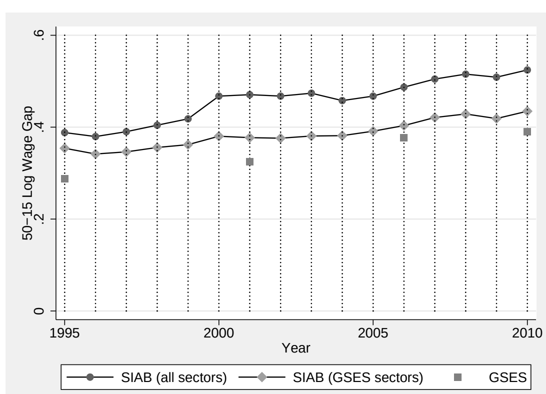
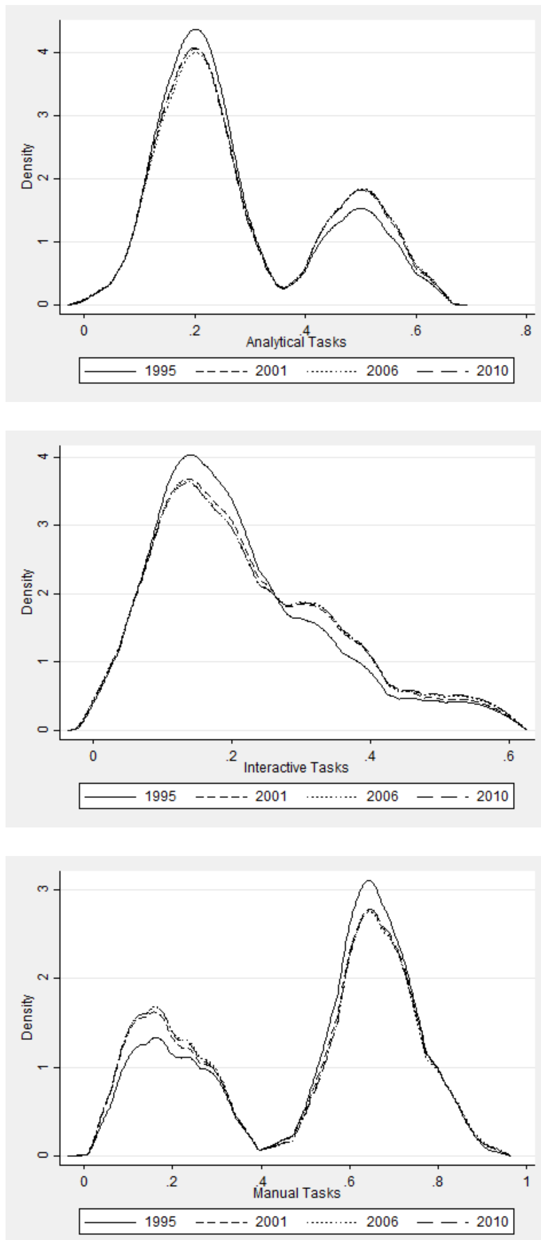



**Figure A3** – Development of inequality, comparison GSES vs. SIAB (50-15 log wage gap)




Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

**Figure A4** – Task composition 1995-2010




Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

**Table A3** – OLS regressions of log hourly wage on covariates (only individuals paid according to a union agreement)

| | 1995 | | 2001 | | 2006 | | 2010 | |
|---|---|---|---|---|---|---|---|---|
| | Coeff. | Std. | Coeff. | Std | Coeff. | Std | Coeff. | Std |
| Age 20-25 | -0.122 | 0.002 | -0.138 | 0.004 | -0.150 | 0.005 | -0.129 | 0.005 |
| Age 26-30 | -0.043 | 0.001 | -0.053 | 0.002 | -0.071 | 0.004 | -0.058 | 0.003 |
| Age 31-35 | 0.006 | 0.001 | 0.004 | 0.002 | -0.001 | 0.002 | -0.002 | 0.002 |
| Age 36-40 | 0.025 | 0.001 | 0.031 | 0.001 | 0.032 | 0.002 | 0.032 | 0.002 |
| Age 41-45 | 0.031 | 0.001 | 0.035 | 0.002 | 0.047 | 0.002 | 0.049 | 0.002 |
| Age 46-50 | 0.035 | 0.001 | 0.041 | 0.002 | 0.046 | 0.002 | 0.048 | 0.002 |
| Age 51-55 | 0.038 | 0.002 | 0.045 | 0.002 | 0.051 | 0.003 | 0.035 | 0.003 |
| Age 56-60 | 0.030 | 0.002 | 0.035 | 0.004 | 0.045 | 0.004 | 0.026 | 0.004 |
| Variance age coefficients (x100) | 0.274 | | 0.362 | | 0.469 | | 0.346 | |
| Tenure 0-5 | -0.058 | 0.002 | -0.065 | 0.003 | -0.065 | 0.004 | -0.080 | 0.004 |
| Tenure 6-10 | -0.011 | 0.001 | -0.013 | 0.004 | -0.013 | 0.003 | -0.014 | 0.003 |
| Tenure 11-15 | 0.005 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.003 |
| Tenure 16-20 | 0.008 | 0.001 | 0.019 | 0.002 | 0.016 | 0.002 | 0.015 | 0.003 |
| Tenure 21-25 | 0.020 | 0.001 | 0.024 | 0.002 | 0.027 | 0.003 | 0.029 | 0.003 |
| Tenure >25 | 0.037 | 0.002 | 0.032 | 0.003 | 0.032 | 0.003 | 0.048 | 0.004 |
| Variance tenure coefficients (x100) | 0.089 | | 0.106 | | 0.107 | | 0.166 | |
| Lower/middle secondary w/o vocational training | -0.087 | 0.002 | -0.102 | 0.004 | -0.104 | 0.005 | -0.110 | 0.005 |
| Lower/middle secondary w/ vocational training | -0.041 | 0.002 | -0.047 | 0.003 | -0.050 | 0.004 | -0.056 | 0.003 |
| Upper secondary (German high school equivalent) | -0.031 | 0.004 | -0.022 | 0.004 | -0.025 | 0.005 | -0.013 | 0.006 |
| University of Applied Science (Fachhochschule) | 0.085 | 0.003 | 0.096 | 0.004 | 0.088 | 0.005 | 0.092 | 0.005 |
| University | 0.124 | 0.005 | 0.146 | 0.006 | 0.153 | 0.008 | 0.147 | 0.006 |
| Missing information | -0.050 | 0.005 | -0.071 | 0.006 | -0.062 | 0.012 | -0.061 | 0.008 |
| Variance education coefficients (x100) | 0.589 | | 0.811 | | 0.816 | | 0.820 | |
| Non-skilled blue collar | -0.100 | 0.002 | -0.088 | 0.003 | -0.092 | 0.005 | -0.100 | 0.004 |
| Skilled blue collar and foremen | -0.008 | 0.001 | -0.013 | 0.002 | -0.016 | 0.004 | -0.005 | 0.003 |
| White collar | 0.108 | 0.002 | 0.101 | 0.004 | 0.108 | 0.005 | 0.105 | 0.005 |
| Variance occupational position coefficients | 0.724 | | 0.604 | | 0.680 | | 0.702 | |
| Offshoring (0-100%) | 0.005 | 0.001 | 0.008 | 0.001 | 0.009 | 0.001 | 0.009 | 0.001 |
| Imports of consumption goods (0-100%) | -0.002 | 0.000 | -0.002 | 0.000 | -0.002 | 0.000 | -0.001 | 0.000 |
| No Exports | -0.012 | 0.004 | -0.014 | 0.006 | -0.029 | 0.008 | -0.042 | 0.011 |
| Export share 1-25% | -0.005 | 0.002 | -0.002 | 0.004 | 0.005 | 0.006 | -0.025 | 0.006 |
| Export share 26-50% | 0.007 | 0.003 | 0.012 | 0.004 | 0.003 | 0.009 | 0.025 | 0.007 |
| Export share 51-100% | 0.010 | 0.003 | 0.004 | 0.005 | 0.021 | 0.007 | 0.041 | 0.008 |
| Variance export coefficients (x100) | 0.008 | | 0.009 | | 0.033 | | 0.117 | |
| Mining and other quarring | -0.080 | 0.012 | -0.136 | 0.045 | -0.021 | 0.022 | -0.002 | 0.036 |
| Food products, beverages, tobacco | -0.043 | 0.005 | -0.056 | 0.009 | -0.042 | 0.013 | -0.027 | 0.016 |
| Textiles | -0.102 | 0.007 | -0.085 | 0.02 | -0.110 | 0.017 | -0.161 | 0.019 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Wood | -0.011 | 0.008 | -0.017 | 0.021 | -0.049 | 0.015 | -0.130 | 0.021 |
| Paper | 0.002 | 0.008 | 0.009 | 0.008 | 0.022 | 0.016 | -0.005 | 0.011 |
| Printing | 0.153 | 0.008 | 0.164 | 0.010 | 0.164 | 0.014 | 0.155 | 0.037 |
| Coke and petroleum products | 0.090 | 0.031 | 0.138 | 0.027 | 0.156 | 0.028 | 0.227 | 0.068 |
| Chemicals | 0.038 | 0.007 | 0.027 | 0.008 | 0.031 | 0.012 | 0.024 | 0.013 |
| Rubber, plastic | -0.013 | 0.006 | -0.027 | 0.008 | -0.050 | 0.022 | -0.043 | 0.017 |
| Non-metallic products | -0.013 | 0.006 | -0.044 | 0.006 | -0.081 | 0.014 | -0.071 | 0.016 |
| Basic metals | 0.041 | 0.007 | 0.057 | 0.014 | 0.049 | 0.020 | 0.043 | 0.014 |
| Fabricated metal products | 0.024 | 0.005 | -0.004 | 0.009 | 0.006 | 0.016 | -0.010 | 0.017 |
| Computer, electronic, optical products | 0.025 | 0.006 | 0.026 | 0.008 | 0.046 | 0.01 | 0.051 | 0.013 |
| Electrical equipment | 0.020 | 0.006 | 0.024 | 0.009 | 0.016 | 0.015 | 0.032 | 0.011 |
| Machinery and equipment | 0.036 | 0.005 | 0.031 | 0.007 | 0.013 | 0.012 | 0.004 | 0.017 |
| Motor vehicles, trailers | 0.124 | 0.008 | 0.112 | 0.010 | 0.099 | 0.013 | 0.065 | 0.012 |
| Other transport equipment | -0.008 | 0.008 | 0.079 | 0.031 | 0.023 | 0.016 | 0.125 | 0.016 |
| Furniture etc | -0.019 | 0.008 | -0.027 | 0.013 | -0.051 | 0.015 | 0.010 | 0.014 |
| Electricity, water, recycling | 0.078 | 0.008 | 0.104 | 0.011 | 0.114 | 0.018 | 0.108 | 0.017 |
| Construction | 0.040 | 0.004 | -0.009 | 0.007 | -0.036 | 0.009 | -0.013 | 0.012 |
| Trade of vehicles | -0.051 | 0.010 | -0.057 | 0.009 | -0.042 | 0.013 | -0.028 | 0.019 |
| Wholesale trade | -0.110 | 0.008 | -0.091 | 0.015 | -0.053 | 0.016 | -0.088 | 0.020 |
| Retail trade | -0.157 | 0.009 | -0.162 | 0.013 | -0.173 | 0.031 | -0.233 | 0.023 |
| Finance and insurance | -0.066 | 0.007 | -0.057 | 0.010 | -0.030 | 0.010 | -0.033 | 0.014 |
| Variance sector coefficients (×100) | 0.516 | | 0.648 | | 0.609 | | 0.950 | |
| Firmsize 10-19 | -0.051 | 0.004 | -0.049 | 0.006 | -0.054 | 0.008 | -0.065 | 0.012 |
| Firmsize 20-49 | -0.033 | 0.004 | -0.039 | 0.005 | -0.040 | 0.009 | -0.043 | 0.010 |
| Firmsize 50-99 | -0.028 | 0.004 | -0.020 | 0.005 | -0.024 | 0.007 | -0.014 | 0.008 |
| Firmsize 100-199 | -0.008 | 0.003 | -0.005 | 0.006 | -0.014 | 0.007 | 0.006 | 0.009 |
| Firmsize 200-499 | 0.016 | 0.002 | 0.011 | 0.004 | 0.007 | 0.007 | 0.013 | 0.007 |
| Firmsize 500-999 | 0.038 | 0.003 | 0.036 | 0.005 | 0.048 | 0.010 | 0.038 | 0.008 |
| Firmsize >1000 | 0.066 | 0.003 | 0.065 | 0.006 | 0.077 | 0.007 | 0.065 | 0.009 |
| Variance firmsize coefficients (×100) | 0.151 | | 0.143 | | 0.194 | | 0.173 | |
| State-owned | -0.020 | 0.003 | -0.043 | 0.008 | -0.025 | 0.008 | -0.033 | 0.007 |
| Schleswig-Holstein | 0.075 | 0.006 | 0.076 | 0.012 | 0.052 | 0.012 | 0.028 | 0.013 |
| Hamburg | 0.167 | 0.008 | 0.176 | 0.035 | 0.126 | 0.017 | 0.120 | 0.018 |
| Lower Saxony | 0.069 | 0.004 | 0.064 | 0.006 | 0.030 | 0.008 | 0.029 | 0.009 |
| Bremen | 0.126 | 0.007 | 0.032 | 0.012 | 0.049 | 0.014 | 0.061 | 0.017 |
| North Rhine-Westphalia | 0.106 | 0.003 | 0.086 | 0.005 | 0.062 | 0.008 | 0.051 | 0.008 |
| Hesse | 0.076 | 0.004 | 0.072 | 0.006 | 0.059 | 0.009 | 0.054 | 0.010 |
| Rhineland-Palatinate | 0.099 | 0.006 | 0.083 | 0.006 | 0.076 | 0.015 | 0.047 | 0.010 |
| Baden-Wuerttemberg | 0.120 | 0.004 | 0.131 | 0.006 | 0.126 | 0.008 | 0.112 | 0.012 |
| Bavaria | 0.071 | 0.004 | 0.066 | 0.005 | 0.072 | 0.009 | 0.077 | 0.009 |
| Saarland | 0.068 | 0.008 | 0.068 | 0.013 | 0.051 | 0.012 | 0.056 | 0.016 |
| Berlin | 0.056 | 0.009 | 0.040 | 0.011 | 0.021 | 0.018 | 0.034 | 0.021 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Brandenburg | -0.149 | 0.011 | -0.149 | 0.014 | -0.125 | 0.012 | -0.086 | 0.031 |
| Mecklenburg-West Pomerania | -0.194 | 0.009 | -0.181 | 0.018 | -0.142 | 0.017 | -0.184 | 0.033 |
| Saxony | -0.231 | 0.008 | -0.199 | 0.015 | -0.162 | 0.019 | -0.140 | 0.018 |
| Saxony-Anhalt | -0.215 | 0.009 | -0.190 | 0.011 | -0.137 | 0.016 | -0.112 | 0.019 |
| Thuringia | -0.245 | 0.015 | -0.176 | 0.010 | -0.159 | 0.018 | -0.147 | 0.019 |
| Variance federal states coefficients (x100) | 2.048 | | 1.556 | | 1.032 | | 0.874 | |
| Share of analytical tasks | 0.165 | 0.011 | 0.187 | 0.018 | 0.188 | 0.032 | 0.192 | 0.026 |
| Share of interactive tasks | -0.014 | 0.013 | -0.046 | 0.022 | -0.059 | 0.036 | -0.056 | 0.031 |
| Share of manual tasks | -0.151 | 0.005 | -0.140 | 0.009 | -0.128 | 0.009 | -0.136 | 0.011 |
| Variance task coefficients (x100) | 1.674 | | 1.889 | | 1.840 | | 1.950 | |
| Constant | 2.853 | 0.005 | 2.882 | 0.008 | 2.906 | 0.010 | 2.927 | 0.011 |
| Root MSE | 0.186 | | 0.199 | | 0.209 | | 0.205 | |
| $R^2$ | 0.564 | | 0.542 | | 0.539 | | 0.548 | |

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

Standard errors clustered at establishment level. Coefficients within groups of categorial regressors are centered around zero.

**Table A4** – OLS regressions of log hourly wage on covariates (only individuals *not* paid according to a union agreement)

| | 1995 | | 2001 | | 2006 | | 2010 | |
|---|---|---|---|---|---|---|---|---|
| | Coeff. | Std. | Coeff. | Std | Coeff. | Std | Coeff. | Std |
| Age 20-25 | -0.162 | 0.005 | -0.189 | 0.005 | -0.201 | 0.004 | -0.193 | 0.005 |
| Age 26-30 | -0.083 | 0.003 | -0.086 | 0.003 | -0.113 | 0.003 | -0.106 | 0.003 |
| Age 31-35 | -0.023 | 0.003 | -0.008 | 0.003 | -0.020 | 0.003 | -0.028 | 0.003 |
| Age 36-40 | 0.016 | 0.003 | 0.033 | 0.002 | 0.047 | 0.002 | 0.034 | 0.003 |
| Age 41-45 | 0.041 | 0.003 | 0.050 | 0.003 | 0.077 | 0.002 | 0.079 | 0.003 |
| Age 46-50 | 0.061 | 0.003 | 0.054 | 0.003 | 0.075 | 0.003 | 0.083 | 0.003 |
| Age 51-55 | 0.081 | 0.004 | 0.066 | 0.004 | 0.071 | 0.003 | 0.073 | 0.003 |
| Age 56-60 | 0.070 | 0.005 | 0.081 | 0.005 | 0.065 | 0.004 | 0.060 | 0.004 |
| Variance age coefficients (x100) | 0.635 | | 0.758 | | 0.957 | | 0.906 | |
| Tenure 0-5 | -0.087 | 0.004 | -0.098 | 0.004 | -0.103 | 0.004 | -0.098 | 0.003 |
| Tenure 6-10 | -0.007 | 0.004 | -0.040 | 0.003 | -0.042 | 0.003 | -0.032 | 0.003 |
| Tenure 11-15 | 0.013 | 0.004 | 0.000 | 0.004 | 0.004 | 0.003 | 0.004 | 0.003 |
| Tenure 16-20 | 0.013 | 0.004 | 0.042 | 0.004 | 0.029 | 0.003 | 0.022 | 0.003 |
| Tenure 21-25 | 0.031 | 0.004 | 0.041 | 0.006 | 0.057 | 0.004 | 0.043 | 0.004 |
| Tenure $>$25 | 0.037 | 0.005 | 0.054 | 0.006 | 0.055 | 0.004 | 0.060 | 0.004 |
| Variance tenure coefficients (x100) | 0.171 | | 0.293 | | 0.325 | | 0.276 | |
| Lower/middle secondary w/o vocational training | -0.120 | 0.006 | -0.138 | 0.006 | -0.137 | 0.006 | -0.147 | 0.006 |
| Lower/middle secondary w/ vocational training | -0.065 | 0.004 | -0.084 | 0.004 | -0.091 | 0.004 | -0.101 | 0.004 |
| Upper secondary (German high school equivalent) | 0.024 | 0.009 | 0.012 | 0.010 | -0.001 | 0.008 | 0.010 | 0.008 |
| University of Applied Science (Fachhochschule) | 0.081 | 0.006 | 0.095 | 0.007 | 0.104 | 0.006 | 0.108 | 0.006 |
| University | 0.155 | 0.007 | 0.192 | 0.007 | 0.204 | 0.007 | 0.216 | 0.006 |

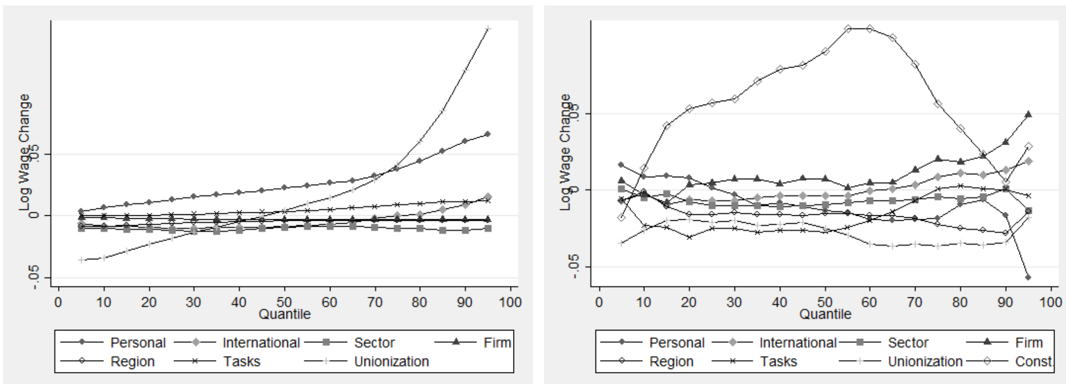
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Missing information | -0.075 | 0.008 | -0.078 | 0.010 | -0.079 | 0.009 | -0.087 | 0.006 |
| Variance education coefficients (x100) | 0.924 | | 1.304 | | 1.429 | | 1.630 | |
| Non-skilled blue collar | -0.122 | 0.004 | -0.124 | 0.005 | -0.119 | 0.004 | -0.125 | 0.004 |
| Skilled blue collar and foremen | -0.011 | 0.003 | -0.024 | 0.003 | -0.015 | 0.003 | -0.003 | 0.004 |
| White collar | 0.134 | 0.006 | 0.148 | 0.006 | 0.133 | 0.005 | 0.128 | 0.006 |
| Variance occupational position coefficients | 1.099 | | 1.262 | | 1.069 | | 1.065 | |
| Offshoring (0-100%) | -0.005 | 0.002 | 0.002 | 0.001 | 0.005 | 0.001 | 0.011 | 0.001 |
| Imports of consumption goods (0-100%) | 0.000 | 0.001 | -0.002 | 0.000 | -0.001 | 0.000 | -0.002 | 0.000 |
| No Exports | -0.036 | 0.008 | -0.022 | 0.009 | -0.010 | 0.007 | -0.037 | 0.009 |
| Export share 1-25% | -0.013 | 0.005 | 0.005 | 0.005 | -0.005 | 0.006 | -0.013 | 0.006 |
| Export share 26-50% | 0.014 | 0.007 | 0.013 | 0.008 | 0.002 | 0.007 | 0.015 | 0.007 |
| Export share 51-100% | 0.035 | 0.008 | 0.004 | 0.010 | 0.014 | 0.007 | 0.035 | 0.009 |
| Variance export coefficients (x100) | 0.072 | | 0.017 | | 0.008 | | 0.075 | |
| Mining and other quarring | 0.037 | 0.016 | 0.023 | 0.022 | -0.021 | 0.030 | 0.096 | 0.017 |
| Food products, beverages, tobacco | -0.064 | 0.011 | -0.093 | 0.011 | -0.103 | 0.010 | -0.062 | 0.012 |
| Textiles | -0.056 | 0.017 | -0.067 | 0.019 | -0.072 | 0.014 | -0.107 | 0.021 |
| Wood | -0.032 | 0.012 | -0.033 | 0.011 | -0.069 | 0.013 | -0.068 | 0.010 |
| Paper | -0.008 | 0.016 | 0.007 | 0.014 | 0.029 | 0.013 | 0.050 | 0.011 |
| Printing | 0.087 | 0.013 | 0.058 | 0.014 | 0.070 | 0.010 | 0.061 | 0.012 |
| Coke and petroleum products | 0.132 | 0.029 | 0.137 | 0.022 | 0.228 | 0.022 | 0.176 | 0.031 |
| Chemicals | 0.030 | 0.015 | 0.040 | 0.013 | 0.063 | 0.012 | 0.044 | 0.013 |
| Rubber, plastic | -0.026 | 0.012 | -0.028 | 0.010 | -0.036 | 0.012 | -0.021 | 0.012 |
| Non-metallic products | 0.018 | 0.010 | 0.011 | 0.010 | -0.015 | 0.012 | 0.006 | 0.015 |
| Basic metals | 0.034 | 0.014 | 0.048 | 0.016 | 0.029 | 0.014 | 0.025 | 0.015 |
| Fabricated metal products | 0.022 | 0.009 | 0.004 | 0.008 | -0.009 | 0.009 | -0.029 | 0.012 |
| Computer, electronic, optical products | -0.026 | 0.009 | -0.028 | 0.011 | -0.004 | 0.009 | -0.011 | 0.011 |
| Electrical equipment | -0.035 | 0.010 | -0.013 | 0.013 | -0.012 | 0.011 | -0.034 | 0.012 |
| Machinery and equipment | -0.011 | 0.009 | 0.002 | 0.010 | 0.009 | 0.011 | -0.006 | 0.010 |
| Motor vehicles, trailers | 0.057 | 0.018 | 0.046 | 0.017 | 0.078 | 0.012 | 0.079 | 0.012 |
| Other transport equipment | 0.010 | 0.023 | 0.048 | 0.025 | 0.047 | 0.016 | 0.041 | 0.020 |
| Furniture etc | -0.031 | 0.018 | -0.070 | 0.012 | -0.093 | 0.014 | -0.039 | 0.008 |
| Electricity, water, recycling | 0.057 | 0.023 | 0.072 | 0.019 | 0.097 | 0.015 | 0.025 | 0.016 |
| Construction | 0.063 | 0.008 | 0.028 | 0.007 | 0.008 | 0.007 | 0.012 | 0.009 |
| Trade of vehicles | -0.054 | 0.029 | -0.034 | 0.011 | -0.048 | 0.010 | -0.081 | 0.017 |
| Wholesale trade | -0.055 | 0.010 | -0.059 | 0.009 | -0.037 | 0.010 | -0.029 | 0.011 |
| Retail trade | -0.150 | 0.015 | -0.111 | 0.016 | -0.200 | 0.023 | -0.251 | 0.017 |
| Finance and insurance | 0.000 | 0.019 | 0.011 | 0.026 | 0.063 | 0.018 | 0.124 | 0.016 |
| Variance sector coefficients (x100) | 0.334 | | 0.313 | | 0.664 | | 0.698 | |
| Firmsize 10-19 | -0.086 | 0.007 | -0.090 | 0.007 | -0.101 | 0.007 | -0.063 | 0.008 |
| Firmsize 20-49 | -0.063 | 0.007 | -0.077 | 0.008 | -0.077 | 0.007 | -0.057 | 0.007 |
| Firmsize 50-99 | -0.053 | 0.007 | -0.052 | 0.007 | -0.052 | 0.006 | -0.036 | 0.008 |
| Firmsize 100-199 | -0.021 | 0.007 | -0.017 | 0.009 | -0.021 | 0.006 | -0.023 | 0.008 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Firmsize 200-499 | 0.042 | 0.009 | 0.030 | 0.008 | 0.026 | 0.008 | 0.005 | 0.009 |
| Firmsize 500-999 | 0.087 | 0.009 | 0.098 | 0.013 | 0.098 | 0.016 | 0.038 | 0.011 |
| Firmsize >1000 | 0.095 | 0.010 | 0.107 | 0.014 | 0.127 | 0.009 | 0.136 | 0.010 |
| Variance firmsize coefficients (x100) | 0.471 | | 0.557 | | 0.653 | | 0.414 | |
| State-owned | 0.023 | 0.015 | 0.019 | 0.016 | -0.005 | 0.011 | -0.011 | 0.009 |
| Schleswig-Holstein | 0.123 | 0.014 | 0.069 | 0.012 | 0.053 | 0.011 | 0.078 | 0.014 |
| Hamburg | 0.176 | 0.015 | 0.185 | 0.020 | 0.146 | 0.012 | 0.168 | 0.018 |
| Lower Saxony | 0.093 | 0.010 | 0.086 | 0.009 | 0.061 | 0.009 | 0.033 | 0.009 |
| Bremen | 0.103 | 0.028 | 0.078 | 0.015 | 0.088 | 0.017 | 0.085 | 0.023 |
| North Rhine-Westphalia | 0.153 | 0.008 | 0.128 | 0.008 | 0.110 | 0.007 | 0.118 | 0.008 |
| Hesse | 0.144 | 0.008 | 0.162 | 0.011 | 0.132 | 0.009 | 0.110 | 0.010 |
| Rhineland-Palatinate | 0.080 | 0.014 | 0.081 | 0.010 | 0.080 | 0.010 | 0.095 | 0.010 |
| Baden-Wuerttemberg | 0.165 | 0.007 | 0.150 | 0.006 | 0.157 | 0.007 | 0.139 | 0.008 |
| Bavaria | 0.156 | 0.007 | 0.116 | 0.008 | 0.112 | 0.009 | 0.106 | 0.009 |
| Saarland | 0.055 | 0.013 | 0.091 | 0.027 | 0.076 | 0.014 | 0.047 | 0.013 |
| Berlin | 0.034 | 0.012 | 0.000 | 0.012 | -0.019 | 0.015 | 0.014 | 0.017 |
| Brandenburg | -0.223 | 0.014 | -0.182 | 0.011 | -0.165 | 0.013 | -0.165 | 0.013 |
| Mecklenburg-West Pomerania | -0.226 | 0.015 | -0.219 | 0.012 | -0.200 | 0.011 | -0.181 | 0.013 |
| Saxony | -0.293 | 0.010 | -0.273 | 0.008 | -0.233 | 0.009 | -0.255 | 0.009 |
| Saxony-Anhalt | -0.263 | 0.014 | -0.231 | 0.012 | -0.201 | 0.011 | -0.165 | 0.013 |
| Thuringia | -0.275 | 0.010 | -0.241 | 0.009 | -0.197 | 0.011 | -0.227 | 0.010 |
| Variance federal states coefficients (x100) | 3.144 | | 2.581 | | 1.972 | | 1.964 | |
| Share of analytical tasks | 0.186 | 0.025 | 0.116 | 0.021 | 0.160 | 0.027 | 0.130 | 0.022 |
| Share of interactive tasks | 0.167 | 0.026 | 0.192 | 0.023 | 0.136 | 0.029 | 0.201 | 0.025 |
| Share of manual tasks | -0.352 | 0.012 | -0.308 | 0.012 | -0.297 | 0.009 | -0.331 | 0.011 |
| Variance task coefficients (x100) | 6.213 | | 4.839 | | 4.410 | | 5.562 | |
| Constant | 3.000 | 0.015 | 3.027 | 0.017 | 2.960 | 0.013 | 2.968 | 0.011 |
| Root MSE | 0.268 | | 0.285 | | 0.295 | | 0.299 | |
| $R^2$ | 0.703 | | 0.656 | | 0.620 | | 0.625 | |

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

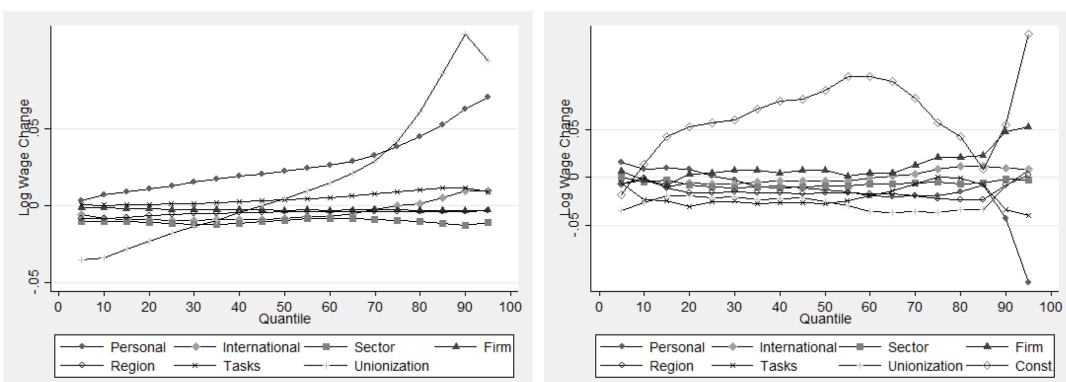
Standard errors clustered at establishment level. Coefficients within groups of categorial regressors are centered around zero.

**Figure A5** – Composition and wage structure effects: daily earnings



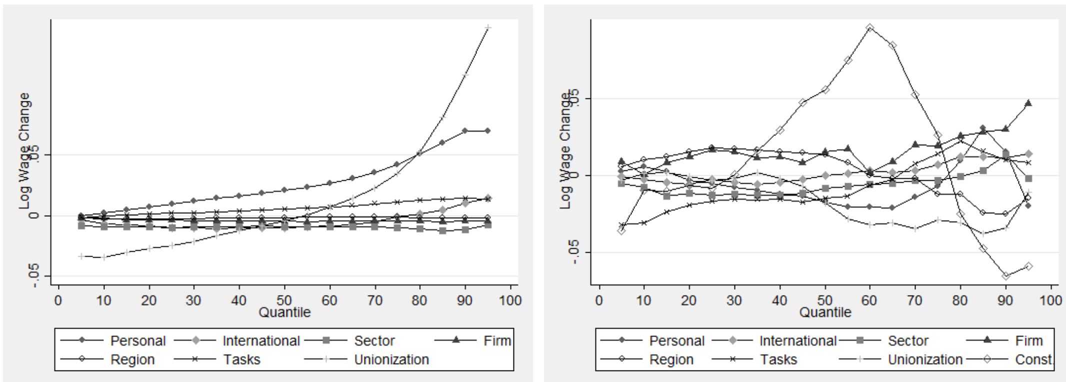

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

**Figure A6** – Composition and wage structure effects: daily earnings censored



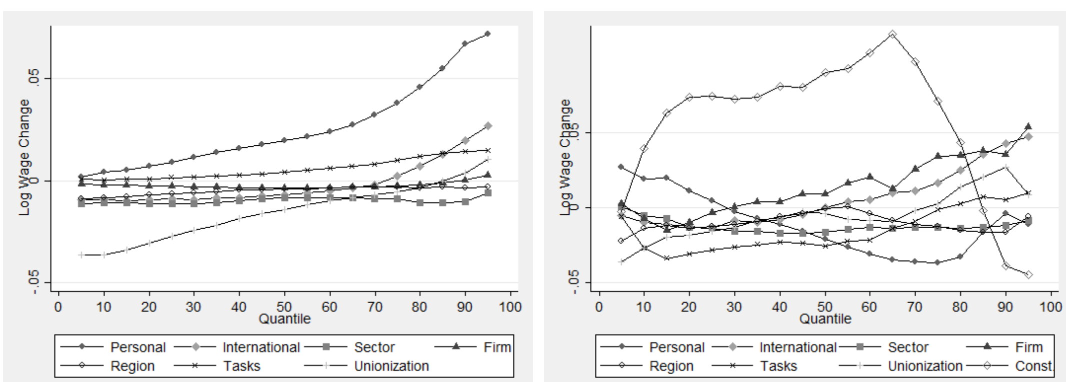

Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

**Figure A7** – Composition and wage structure effects: West Germany




Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

**Figure A8** – Composition and wage structure effects: firm-level unionization




Source: Structure of Earnings Surveys 1995, 2001, 2006, 2010 and own calculations.

# Chapter 3

# Increasing Inequality in Lifetime Earnings: A Tale of Educational Upgrading and Changing Employment Patterns*

## 3.1 Introduction

Growing wage and earnings inequality around the world has caused an increasing interest in the topic among both policymakers and academics. The latter have so far mainly focused on the increase in cross-sectional inequality over time as documented in a vast literature (see Acemoglu and Autor, 2011 for a general overview, and Dustmann et al., 2009, for

*This chapter is based on:
Seckler, M. (2019): Increasing Inequality in Lifetime Earnings: A Tale of Educational Upgrading and Changing Employment Patterns. *University of Tübingen Working Papers in Economics and Finance No. 119*.

the German case). Surprisingly, relatively little is known about how this increasing cross-sectional earnings inequality has affected the evolution of individual long-term and lifetime earnings across different birth cohorts. From a purely cross-sectional perspective, which usually compares earnings distributions at different points in time, cohort differences are usually non-distinguishable from life-cycle trends. For example, when comparing the German earnings distribution of the early 1990s with the one two decades later, it remains unclear to what extend the standard of living of later cohorts differs from their predecessors. This is a consequence of the fact that observable differences in cross-sectional earnings are the result of individuals being observed at different points of their career. Moreover, studying lifetime earnings from a cohort perspective is likely to be more informative with regards to an individual's or cohort's standard of living, which is determined by lifetime earnings rather than by earnings at a certain point in time (see, e.g., Corneo 2015).

Recent studies by Bönke et al. (2015a) and Guevenen et al. (2017) document a dramatic increase in lifetime earnings inequality for both Germany and the U.S. among men in later birth cohorts. Though being an ongoing debate, the previous literature has identified different channels underlying the increase in cross-sectional inequality, most prominently skill-biased technological change (*SBTC*), demographical and institutional factors, as well as internationalization and changes in individual employment biographies.[1] It is unclear to what extend these factors are also responsible for the increasing inequality in lifetime earnings. This paper intends to shed light on this blind spot by disentangling the increasing inequality in lifetime earnings using high-quality administrative employment data for Germany. Methodologically, the paper uses state-of-the-art RIF decomposition techniques as introduced by Firpo et al. (2009, 2018).

The paper makes the following contributions to the literature. First, the present study reveals a lower labor market participation (both in terms of longer periods of part-time employment and non-employment) to be the most important factor for the rise in inequality

[1] For a more comprehensive discussion, see section 3.2

in the lower half of the distribution. Contrary to that, much of the rising inequality at the top is associated with educational upgrading. To the best of my knowledge, this is the first study providing a decomposition analysis aimed at explaining the rising inequality in lifetime earnings. Second, the results confirm previous findings by Bönke et al. (2015a) who documented a sharp rise in lifetime earnings inequality based on a different database. Going a step further, the present paper also shows that German men born between the years 1955 and 1974 did not only face a higher level of inequality, but equally suffered from a stagnation in total earnings for a major part of their career. In fact, this development stems from losses within education groups which are counterbalanced by higher levels of educational attainment. The present paper also provides first evidence that these trends tend to accelerate for the youngest cohorts.

The rest of this paper is structured as follows: Section 3.2 summarizes the related literature. Sections 3.3 and 3.4 describe the data and the econometric method. Section 3.5 presents the empirical results. Section 3.6 concludes with a discussion of the major findings.

## 3.2 Related literature

This section provides an overview on the most relevant literature for the present paper. Most importantly, the study directly adds to the literature on the evolution of individual long-term and lifetime earnings inequality. Using data for the U.S., an important contribution by Bowlus and Robin (2004) finds that inequality in cross-sectional and lifetime earnings appear to follow a similar pattern over time. Moreover, they show that the level of inequality in lifetime earnings is substantially lower than inequality in cross-sectional earnings due to earnings mobility among young workers. However, changes in earnings mobility are not identified as an important factor in explaining the rising dispersion in lifetime earnings. As the study builds on a relatively short panel, the used measures of lifetime earnings are simulated based on estimates for different parameters (job destruction/re-employment rates, promotion/demotion rates). Kopczuk et al. (2010) provide evidence for increasing inequality in male long-term earnings, especially for U.S. *baby-boomers* born after 1945. This trend is found in all stages of the career, with the level of inequality being generally higher in later episodes of the working life. In a more recent contribution, Guevenen et al. (2017) document both a substantial decline in median lifetime earnings of U.S. men born between the years 1942 and 1958 (after observing gains in earlier cohorts) and a long-run trend of increasing inequality within male cohorts. The authors conclude that the observed changes are mostly due to differences in early career earnings across cohorts. Importantly, they show that later cohorts suffered from earning losses at young age that were not compensated by a higher future earnings growth.[2]

In a seminal contribution for Germany, Bönke et al. (2015a) documented a dramatic increase in lifetime earnings inequality based on the Insurance Account Sample (*Versicherungskontenstichprobe*) of the Federal Pension Register containing West German men born between the years 1935 and 1969. The authors resort to the concept of *up-to-age X earnings (UAX)* as a measure for individual long-term earnings, which is defined as the present value of all earnings before reaching a certain age.[3] By imputing earnings for periods of un- and non-employment, they show that parts of the increasing dispersion in lifetime earnings at the bottom can be explained by differential unemployment patterns. Moreover, they establish two other results that are important for the subsequent analysis. First, they show that earnings mobility, which is high at the beginning of the working life, mostly vanishes after age 40. Second, they conclude that the evolution of inequality in lifetime earnings most likely reflects the development up to age 40. Following this argument, the subsequent analysis focuses on earnings up-to age 40, which does not only offer important insights into changes in individual long-term earnings for a major part of the career, but can most likely

[2] In fact, the study finds that women realized substantial gains in lifetime earnings (starting from a very low level) across the study period. However, these gains only partly offset the losses suffered by men.

[3] Despite some methodological differences, the same terminology is also used in the present paper.

be generalized to inequality in lifetime earnings.[4] In a further contribution, Bönke et al. (2015b) provide evidence for an increase in the transitory component for younger workers in the 1970s and a related increase in short-term earnings risk. The present paper intends to directly add to these previous findings by trying to pin down the aforementioned increase in lifetime earnings across cohorts to different explanatory factors.

In this aspect, the present study connects to a vast literature trying to explain the well-documented increase in cross-sectional inequality during the last decades as described by various authors (see, for the German case, Dustmann et al., 2009, Card et al., 2013). These studies are usually concerned about the evolution of cross-sectional inequality and do not explicitly address the question of how these factors affect lifetime earnings inequality across different birth cohorts. Although not having reached a consensus yet, the respective literature identifies several factors that appear to be important for the increase in cross-sectional inequality, which therefore also constitute obvious candidates for the analysis in this paper. Most notably, many studies stress the importance of skill-biased technological change (SBTC) for wage polarization and a resulting increase in U.S. wage inequality (e.g. Autor and Dorn, 2013). However, previous evidence on this link seems to be mixed for Germany (see, e.g. Antonczyk et al., 2009, Rinawi and Backes-Gellner, 2015). Other contributions show that an increasing heterogeneity between firms, combined with a matching of *good workers* and *good firms*, can explain a large part of the recent increase in inequality (Card et al., 2013, Barth et al., 2016, Song et al., 2019). A different strand of the literature (e.g. Dustmann et al., 2009, Baumgarten et al., 2018, Biewen and Seckler, 2019) highlights the importance of institutional changes in the form of de-unionization, whereas internationalization seems to to be another potential explanation (Baumgarten, 2013). A recent study (Biewen et al., 2018) stresses the importance of an increasing heterogeneity in individual labor market histories. Based on a reweighting methodology, the authors link a substantial part of the rising earnings

[4] Also see Bönke et al. (2015a), p. 186. As already argued above, this finding is also confirmed in Guevenen et al. (2017). Another advantage of this approach is to obtain new evidence on very recent cohorts.

inequality to increasing heterogeneity in terms of past employment interruptions and part-time work, especially at the bottom of the distribution.

As the present studies identifies changing employment patterns as an important factor, it also relates to a broader literature on the evolution and earnings effects of employment breaks and part-time employment. Previous work by Tisch and Tophoven (2012) compares birth cohorts 1959 and 1965 of the German *baby boomers*. Similar to the results of the present paper, they document an increasing incidence of part-time and non-employment episodes in individual employment biographies among individuals born in later years. Taking also more recent cohorts into account, Bachmann et al. (2018) find a decline in regular employment together with a simultaneous increase in a-typical employment among west German men born between 1944 and 1986. These trends are not only found in young workers, i.e. as a results of substantially longer time spent in education, but across all age groups. Although providing new insights, both studies abstain from establishing a direct link to the evolution of earnings inequality over time. Brehmer and Seifert (2008) and Wolf (2010) show that part-time employment is associated with lower hourly wages relative to full-time employment. Finally, a number of studies (Beblo and Wolf, 2002, Görlich and Grip, 2008, Potrafke, 2012, Fernández-Kranz et al., 2015, Blundell et al., 2016, Paul, 2016) provide direct evidence that both employment interruptions and part-time episodes tend to adversely affect future earnings growth.

## 3.3 Data

The analysis in this paper is based on the *Sample of Integrated Employment Biographies (SIAB)*, which constitutes a 2 percent random sample of all employees covered by social security records between the years 1975 and 2014.[5] The data are well suited for studying

[5]The paper uses the weakly anonymous version of the data which were accessed on-site at the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and via remote data access.

changes in lifetime earnings across cohorts due to the fact that complete employment histories of approximately 1.75 million individuals are provided. The *SIAB* also includes a rich set of covariates related to individual employment biographies, complemented by additional firm-level information of the Establishment History Panel, that can potentially explain the increasing dispersion of lifetime earnings. In this regard, the data are more suitable for a detailed decomposition analysis than the one based on the Federal Pension Register that have mostly been used in previous research but include a very limited number of covariates only. On the downside, the *SIAB* does not contain any information prior to the year 1975. Hence, the study focuses on individuals born between the years 1955 and 1974 who can at least be observed between age 20 and 40. To facilitate comparability with previous studies, the analysis is restricted to male individuals working in West Germany only.

For the subsequent analysis, a sample comprised of individuals with a sufficient labor market attachment is defined. This is achieved by imposing the following restrictions:[6] First, to ensure that individuals can be observed throughout the relevant part of their career, a *maximum age* for labor market entry depending on educational attainment is imposed, i.e. 30 years (individuals with university degree), 28 (completed high school and vocational training), 25 (without completed high school but with vocational training) and 23 for all others (neither high school degree nor vocational training or missing educational information). Similarly, individuals who have their last observable employment spell more than 3 months before reaching a certain age threshold (e.g. age 40), as well as individuals with a single non-employment spell of more than five years, are omitted from the sample. Imposing similar restrictions is important to minimize the risk of including individuals who emigrated or became self-employed during their working life. Second, lower bounds on both annual and total long-term earnings are imposed. Regarding annual earnings, individuals are required to have real earnings greater than 5000 euros in at least half of the years they could potentially be working after age 25. For example, to be included in the up-to-age 40 (UA40) earnings

[6]Imposing similar restrictions is common in the literature on long-term earnings inequality. The restrictions used in this paper follow Guevenen et al. (2017) and Boll et al. (2017).

sample, individuals need to have real earnings of at least 5000 euros in eight years or more. Also, individuals are required to have total long-term earnings that correspond to an average of at least 5000 euros per year. Hence, for total UA40 covering all earnings starting with the year the individual turns 20, a lower bound of 105.000 euros is imposed (130.000 euros for UA45). Finally, individuals with observable employment spells in East Germany are equally omitted. Imposing these restrictions leaves 109,194 (81,271/49,864) respondents for which complete UA40 (UA45/UA50) employment biographies can be constructed. A more detailed overview on the number of observations by cohort is provided in table B1 in the appendix.[7]

As the earnings information in the SIAB is censored at the limit for the statutory pension fund, earnings above this threshold are imputed following the procedure described in Gartner (2005).[8] Depending on the year of observation, up to 15 percent of observations are affected by this right-censoring. Hence, as it is common practice in studies based on German administrative data, this paper focuses on the development of earnings inequality below the 85th percentile of the different UAX measures. Due to this property, the subsequent analysis might in fact underestimate the true increase in inequality given that parts of the development at the very top of the distribution will not be captured. Starting in 1984, one-time payments were counted towards annual earnings resulting in both an increase in average daily earnings as well as a spurious increase in annual earnings inequality between the years 1983 and 1984. To account for this structural break, the procedure introduced by Bönke et al. (2015a) is used, which denotes a modification of the procedure by Fitzenberger (1999) that works on panel data.[9]

[7] This paper does not consider women. This is due to lower labor force participation rates among German women, which in turn results in a significantly smaller number of women whose earnings biographies fulfill the imposed minimum criteria of labor market attachment. Moreover, changing patterns in terms of selection into employment (and ultimately into the sample) inherently complicates any long-run comparison across cohorts.

[8] Appendix B presents more details on the imputation procedure.

[9] Note that similar strategies were also used in other studies such as Dustmann et al. (2009) and Card et al. (2013). The procedure is outlined in the appendix

From a data perspective, another challenge lies in the German reunification and the fall of the Berlin Wall, allowing individuals to move freely between the formerly separated parts of Germany. As the *SIAB* does not include any information on earnings in East Germany before January 1, 1991, individuals with employment spells in the former German Democratic Republic (which remain unobservable), who migrated to West Germany in the aftermath of the fall of the Berlin Wall, potentially end up in the sample. However, an effect on the decomposition results (comparing pooled cohorts 1955-57 to 1972-74) can be ruled out due to the following reasons: For the analysis, individuals who can be observed in the SIAB before 1989 are assumed to only consist of West Germans, given the fact that the Berlin Wall did not fall before late 1989 and East-West migration was virtually impossible. Combined with the maximum labor market entry age of 30 (for individuals holding a university degree), individuals born before 1959 are assumed to only consist of West Germans. Similarly, individuals born after 1970 are not affected by relevant unobservable employment spells in East Germany, given the fact that starting in 1991, the SIAB covered both East and West Germany and only earnings starting at age 20 are included in the UAX earnings measures. Hence, the decomposition results are not diluted by individuals with unobservable employment spells in East Germany.[10]

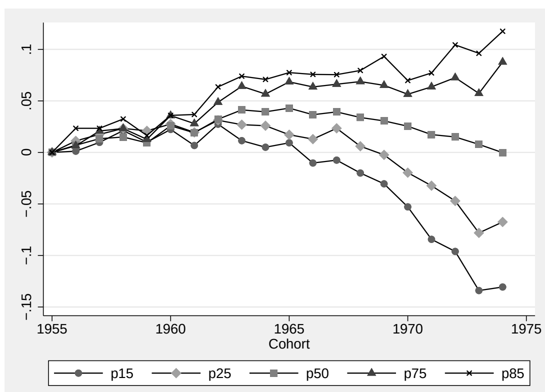
### 3.3.1 Trends in lifetime earnings

In the analysis of lifetime earnings, this paper follows the approach suggested in Bönke et al. (2015) in calculating *up-to-age X earnings (UAX)* for different ages (though with some methodological differences). The concept of up-to-age X earnings addresses and balances

[10]Parts of the descriptive analysis also use information from other cohorts, whose results might potentially be affected by East-West migration following the fall of the Berlin Wall. Note, however, that restrictions on the maximum age for labor market entry are imposed to ensure that only individuals whose (mostly) complete earnings biographies are observable are included. Also, descriptive statistics for the first observable employment spell do not detect any significant anomalies. Nevertheless, there might be rare cases of individuals who started working in East Germany and migrated to West Germany before 1991 and prior to reaching the maximum age for labor market entry.

the trade-off between the number of birth cohorts that can be included in the analysis and the time each individual can be observed in the data. In detail, the computation of *UAX* proceeds as follows. In a first step, daily earnings are aggregated to yearly earnings and inflated/deflated to the level of 2010 using the German consumer price index (CPI). In a second step, cumulative earnings are calculated for each individual between the year the person turns 20 up to and including the year the individuals is reaching a certain age threshold (e.g. age 40 for the computation of UA40). The earnings measures only include payments from employment subject to social insurance contributions before tax, i.e. social transfer-payments as well as earnings from periods of self-employment are not part of the analysis. Hence, they mirror the price of labor paid in the market.[11] Earnings from marginal part-time employment (*Minijobs*) are also not included for consistency reasons, as these episodes were unobservable in the data before April 1, 1999.

**Figure 3.1** – Indexed real growth in UA40




Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.

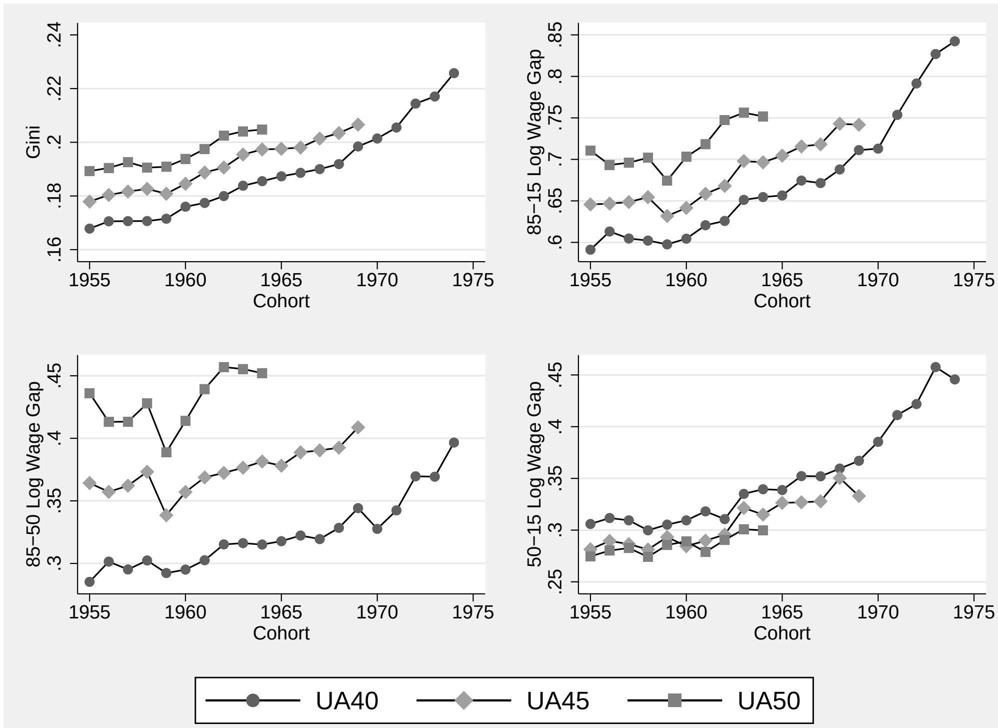
Figure 3.1 illustrates the indexed (real) growth in UA40 earnings at different percentiles of the unconditional within-cohort distribution for men born between the years 1955 and 1974. The graph reveals three important developments. First, an increasing inequality in UA40

[11]Bönke et al. (2015a) also add employers' social insurance contributions to the earnings measure as certain occupational groups, such as minors and sailors, have differing social security arrangements. As the share of these groups is negligible in the cohorts covered in the present study, a similar adjustment is not made.

earnings within cohorts which is due to a monotonic development in the sense that, when considering the overall change between cohorts 1955 and 1974, lower percentiles below the median suffered losses whereas the upper half gained. Numerically, the 85th percentile of the UA40 distribution increased by approximately 12%, whereas the 15th percentile decreased by as much as 13%. Second, over the entire period of study, the graph shows a stagnation in median UA40 earnings with the development resembling an inverse U-shape. More precisely, median earnings increased up to the birth cohort 1965 and gradually deteriorated thereafter. This finding is in contrast to previous studies considering cross-sectional distributions that have documented significant gains in median hourly/daily earnings in the cross-section 1975-2014 (e.g. Dustmann et al., 2009). Third, the graphical analysis suggests that the increase in inequality sped up dramatically among cohorts born in the early 1970s, which seems to be driven by severe real earnings losses at the bottom and some moderate gains at the top. Lastly, note that these developments are not a direct consequence of a delayed labor market entry due to longer times spent in education. As can be seen from figure B1 in the appendix, the overall picture remains mostly unchanged when only earnings starting at age 25 are taken into account.

Figure 3.2 summarizes the impact of this development on inequality in the different long-term earnings measures (UA40/UA45/UA50). Overall, the graph reveals a strong increase in all parts of the respective distributions with the aforementioned acceleration among cohorts born in the early 1970s. In terms of UA40, this is reflected in a sharp increase in the Gini coefficient from 0.168 to 0.226 (approx.+35%), which affected both the upper part (85-50 log wage gap, approx. +39%) and the lower part (50-15 log wage gap, approx. +45%) of the distribution. In this regard, the results partly differ from previous findings on Germany (see Bönke et al., 2015a) who assigned most of the increase in inequality to the bottom of the distribution. Interestingly, the increase at the top of the distribution was mostly driven by cohorts born in the early 1970s that were not included in the previous study. In line with existing evidence, inequality as captured by the different measures is increasing

**Figure 3.2** – Inequality in *up-to-age-X*



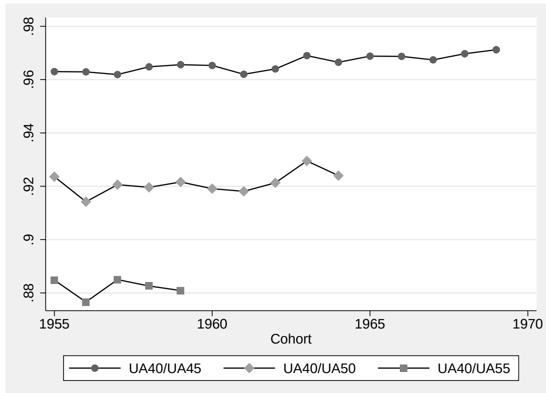

Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.

over the life-cycle.[12] Confirming the general trends previously documented in Bönke et al. (2015a), the presented graphical evidence suggests that the development in UA40 earnings appears to be closely linked to the developments in UA45/UA50 which can, however, only be observed for a limited number of cohorts.

To underpin this hypothesis, figure 3.3 contains rank correlations between UA40 and UAX at higher ages. Generally, the graph documents high and very persistent rank correlations. For example, the dark grey line documents rank correlations between 0.96 and 0.97 between UA40 and UA45. Similarly, results show rank correlations of about 0.92 (UA50) and 0.88

[12]An exception is the lower part of the distribution (50-15 log wage gap) where inequality is highest in terms of UA40.

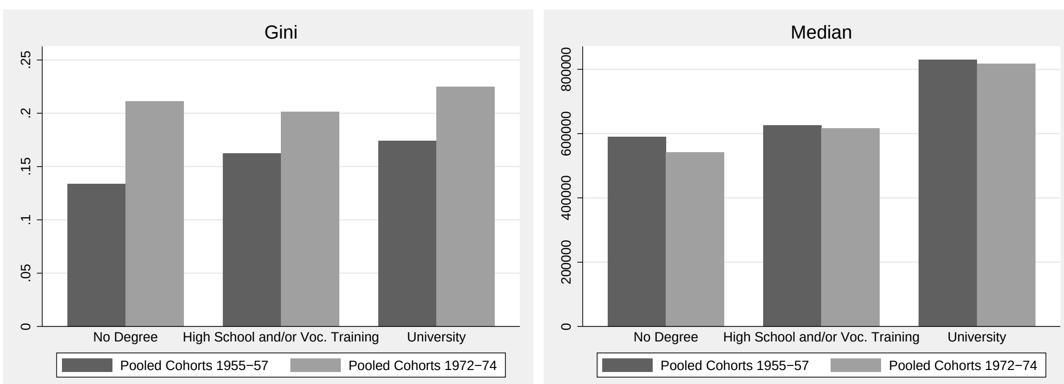
**Figure 3.3 –** Rank correlations of UA40 with selected UAX




Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.

(UA55) which can be interpreted as evidence that the evolution of lifetime earnings closely follows the development in UA40 (compare also Bönke et al., 2015a).

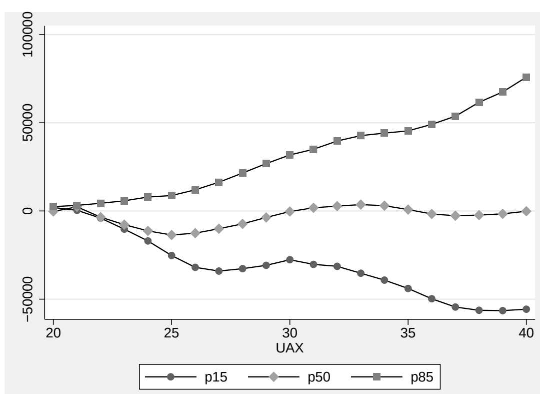
**Figure 3.4 –** Evolution of UA40 within education groups, cohorts 1955-57 vs. 1972-74




Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.

As the German workforce was subject to some major educational upgrading during the period of study (see proceeding section 3.3.3 for more details), it is important to study the development within the different education groups more carefully. Figure 3.4 summarizes the evolution of UA40 within three broad educational groups, i.e. *No Degree*, *High School* *and/or Voc. Training*, as well as *University*, for the pooled cohorts 1972-74 as opposed to 1955-57. The graph on the left includes the development of inequality in terms of the Gini, the graph on the right the change in median earnings. It becomes obvious that inequality did not only increase among all individuals of later cohorts but also within education groups. This increase was strongest within the lowest educational group (approx.+55%), followed by individuals holding a university degree (approx.+29%), and smallest among individuals with a vocational background (approx.+24%). Nevertheless, the impact of the sharp rise of inequality within the lowest educational group on overall inequality should not be overstated given the small relative group size. At the same time, the figure reveals a decrease in median earnings within all education subgroups. These losses were strongest for individuals without a degree (approx.-8%), in contrast to rather marginal losses among individuals with vocational training (approx. -1.5%) and individuals holding a university degree (approx. -1.3%). This mirrors the previous finding of losses in UA40 being mostly located at the bottom of the UA40 distribution. As overall median earnings virtually stagnated (approx. -0.2%), this results suggests that the losses within educational subgroups were neutralized by a shift towards higher average educational attainment among later birth cohorts.

**Figure 3.5** – Changes in UAX, cohorts 1955-57 vs. 1972-74




Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.

To get a better understanding of changes over the life-cycle, figure 3.5 plots the difference in up-to-age X (UAX) for different ages, once more comparing pooled cohorts 1955-57 and 1972-74. For example, the point 25 on the x-axis represents differences in up-to-age 25 (UA25) earnings between the two groups. The graph shows that some losses in the median of cumulative earnings among individuals born 1972-74 occurred until the age of 25, reflecting a delayed labor market entry as a result of the educational expansion. This was followed by a period of fast catch-up in the late 20s and early 30s, which was the likely consequence of a higher share of university graduates entering the labor market and neutralized the preceding median losses by the age of 30. Importantly, there were no further median gains between the ages 30 and 40, causing the previously described stagnation in UA40 earnings. Though being somewhat speculative, the picture suggests that this stagnation is likely to continue for UAX at higher ages that are still unobservable in the data. Simultaneously, the graph shows that the gains in cumulative earnings at the top of the within-cohort distribution increased continuously after age 25, whereas losses at the 15th percentile were already strong in terms of UA25 and (after some stabilization) sped up again in the mid 30s.[13]

### 3.3.2 Trends in employment patterns

Against the background of the trends outlined in the previous section, it is insightful to take a closer look at factors that can potentially explain this development. Hereby, it is crucial to understand whether the observed changes are caused by changes in individuals' labor market participation during the working life, or whether they are due to changes in earnings during the time individuals were actually employed (i.e. changes in lifetime hours worked

[13] Note that the percentiles always refer to differences in the within-cohort distributions for cumulative earnings at a certain age, e.g. age 30. Hence, due to high earnings mobility at young ages, individuals at the 15th percentiles of UA25 earnings are likely to be very different from those at the 15th percentile of UA40 earnings.

vs. changes in earnings conditional on employment).[14] Although the *SIAB* does not include precise information on hours worked, the data allow to consistently distinguish between episodes of full-time, part-time and non-employment in individual employment biographies using the information of the *Employee History (BeH)*, where non-employment is defined as the reference group in the further analysis. In principle, it would also be possible to distinguish episodes of unemployment from other forms of non-employment by exploiting information on unemployment benefits recorded in the *Benefit Recipient History (LeH)*, the *Unemployment Benefit II Recipient Histories (LHG and XLHG)*, as well as the *Jobseeker-Histories (ASU and XASU)* provided by the Federal Employment Agency. However, the latter data sources are not available in the early years. Furthermore, there were several reforms that affected the entitlement to unemployment benefits and hence, a consistent measure across the cohorts used in this study cannot be constructed.[15] As a consequence, the measure used for non-employment is defined as all episodes in individual employment biographies (after labor market entry) where an individual did not follow an employment subject to social insurance contributions. Besides unemployment spells, these include marginal part-time employment (*Minijobs*), self-employment as well as times spent in further education.

Figure 3.6 includes the duration spent in full-time employment (up-to-age 40) for the pooled cohorts 1955-57 and 1972-74 for different quartiles of the UA40 earnings distribution. Although full-time employment remained by far the most frequent employment form among German men, there was a considerable reduction which is found to be strongest for individuals at the bottom of the UA40 distribution. For example, the average time spent in full-time employment among individuals in the bottom quartile of UA40 decreased by approximately 16 months, or 8.9 percent, between pooled cohorts 1955-57 and 1972-74. At the same time, there was also some reduction for higher quartiles which is, however, quantitatively less pronounced and decreasing over the distribution. Numerically, the average time spent

[14] For example, Biewen and Plötze (2019) show that 10-30% of increasing inequality in monthly earnings among German men between 2001 and 2010 were due to changes in hours worked.

[15] Also see Antoni et al. (2016) for more information.

**Figure 3.6** – Full-time employment UA40 in months, cohorts 1955-57 vs. cohorts 1972-74

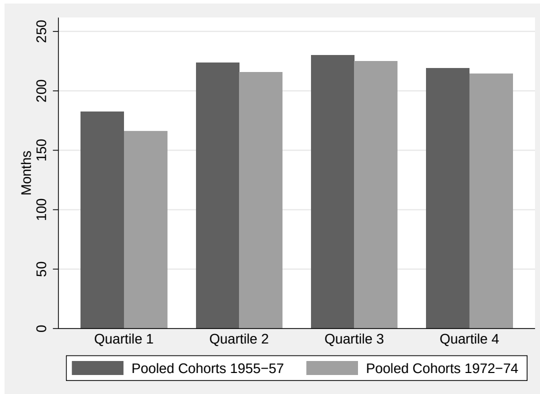



**Figure 3.7** – Non-employment UA40 in months, cohorts 1955-57 vs. cohorts 1972-74

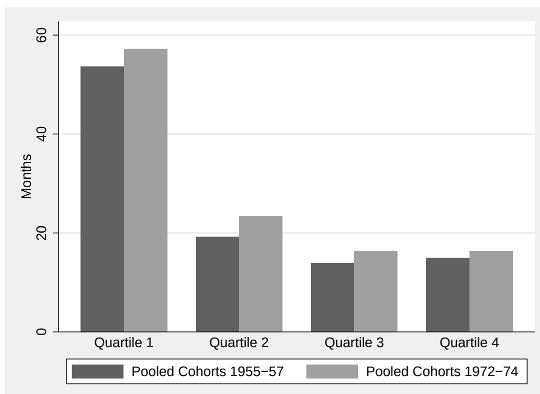
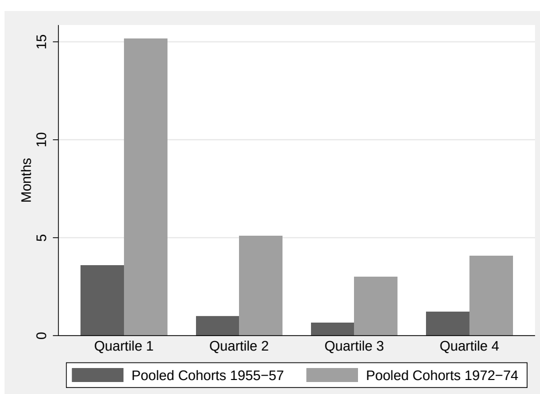



**Figure 3.8** – Part-time employment UA40 in months, cohorts 1955-57 vs. cohorts 1972-74



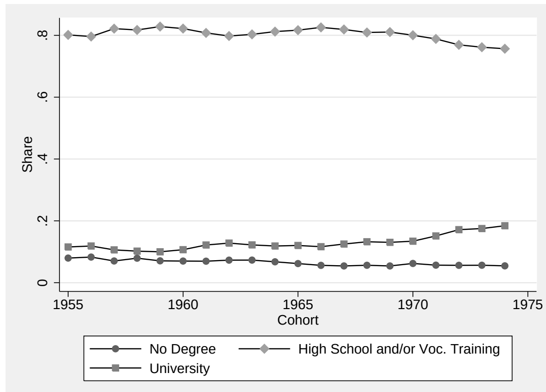

Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.

in full-time employment decreased by on average 7.8 months for quartile 2, 4.9 months for quartile 3 and 4.6 months for the highest quartile. Simultaneously, this development was accompanied by an increase in the incidence of non-employment which was strongest for the two lowest quartiles, with the average increases amounting to approximately 3.6 and 4.1 months, respectively (figure 3.7). However, these numbers also show that the increase in non-employment episodes, which has also been documented in Bönke et al. (2015a), was only partly responsible for the observed decline in full-time duration.

Figure 3.8 illustrates the evolution of part-time employment. Starting from a very low level among individuals of birth cohorts 1955-57, the graph documents a steep increase in the average duration spent in part-time employment in all parts of the UA40 distribution. The graph also shows that individuals in the bottom quartile of the UA40 distribution were by far most affected by this expansion, with the average time spent in part-time employment increasing by on average 11.6 months. This growing importance of part-time employment in recent decades applied, contrary to common perceptions, also to German men (see, e.g. Brenke, 2011, Biewen et al., 2018). Besides ongoing structural changes and a resulting demand for more flexible working arrangements, this development was also enforced by several legal changes, such as the *Teilzeit und Befristungsgesetz (TzBfG)*, which increased the relative attractiveness of part-time employment. The outlined development had a potentially twofold effect on lifetime earnings. Besides a simple reduction in lifetime labor market participation (or lifetime working hours) and the resulting earnings losses, the previous literature has also documented adverse effects of part-time employment on future earnings growth (compare section 3.2).

### 3.3.3 Trends in education

**Figure 3.9** – Share of different education groups




Source: Sample of Integrated Labour Market Biographies (SIAB) 1975 - 2014 and own calculations.

The cohorts included in the study also differ substantially in terms of their educational attainment. Figure 3.9 displays the share of individuals within cohorts in the three broad categories *No Degree*, *High School and/or Vocational Training* as well as *University*. The graph shows the educational expansion of recent decades as similarly documented in previous research. Most importantly, there was a strong increase in the share of individuals holding a university degree, which increased from 11.5% among individuals of birth cohort 1955 to 18.4% among those born in 1974. This development was accompanied by corresponding declines in both the share of medium skilled workers (i.e. individuals with a high school degree and/or vocational training) as well as the share of low skilled workers (i.e. individuals who neither completed vocational training nor hold a high school degree). Note that the later decomposition analysis will use a more fine-grained educational measure distinguishing between six categories: *Lower/middle secondary without vocational training*, *Lower/middle secondary with vocational training*, *Upper secondary (German high school equivalent) without vocational training*, *Upper secondary (German high school equivalent) with vocational training*, *University or Fachhochschule degree* as well as *Missing informa-*

*tion*. To improve on the education variable in the SIAB, which in some cases suffers from both missing and implausible information, the imputation procedure (IP2A) suggested by Fitzenberger et al. (2006) is used.

### 3.3.4 Trends in job mobility, migration and firm characteristics

Beyond the the described differences in employment patterns and educational background, further important characteristics related to individual employment biographies are considered as potential sources of increasing lifetime earnings inequality. For example, changing job mobility patterns across cohorts might constitute another source of increasing inequality in lifetime earnings. Against this background, the further analysis distinguishes two different types of job mobility in line with Gius (2014): firm changes within the same industry or occupation (job changes) on the one hand, and firm changes where both the industry and occupation change (career changes) on the other. Gius (2014) shows this to be an important distinction, given that the first type of job change is associated with a positive earnings effect, whereas the latter one is found to have an adverse effect. The underlying theoretical argument is that individuals with a high number of career changes tend to accumulate fewer industry and occupation-specific human capital and should, on average, have a slower earnings growth over their career. Contrary to that, job changes within a certain occupation or industry (or within both) could potentially be linked to positive earnings effects due to a faster accumulation of human capital. However, the net effect of this second type of job change also remains to a certain extend unclear as it potentially includes a significant share of layoffs or other types of non-voluntary job changes. The descriptive evidence presented in table B2 shows that job changes were generally more frequent than career changes and the mean of both type of firm changes moderately increased among individuals born in the years 1972-74.

To capture the potential impact of migration, a dummy variable indicating whether a person is German by birth is included. According to the definition used in this paper, a person is classified as German by birth if he or she does not have any observable employment spell with foreign nationality throughout the working life. During the observation period, there was an increase of individuals with migration background with their relative shares increasing from 11 to 22 percent between pooled cohorts 1955-57 and 1972-74. Given the previous finding that changing occupational characteristics (as a result of SBTC) potentially explain a significant share of rising cross-sectional wage inequality (see, Ehrl, 2017), a set of 32 occupation dummies is included in the analysis. Differences across industries are captured by the inclusion of sector dummies (44 categories). Both measures refer to the most frequent occupation/sector an individual worked in until the age of 40.

As the previous research on cross-sectional earnings inequality points towards an increasing importance of between firm differences (see section 3.2), the analysis includes a number of firm characteristics that can be constructed from the data. Against the background of the previous literature, the establishment size an individual worked at mostly denotes a potentially important feature for the development of individual long-run earnings. For the subsequent analysis, three firmsizes are distinguished which are small (1-50 employees), medium (51-500 employees) and large (>500 employees) establishments. To capture firm-level technological change, this paper follows a strategy similar to the most recent literature (e.g. Harrigan et al., 2016, Barth et al., 2017) by exploiting information in the Establishment History Panel on the number of engineers and natural scientists (*Techies*) working in an establishment. As these numbers potentially differ systematically across different industries, an establishment is defined as high-tech if its share of engineers and natural scientists lies above the mean of the industry. In an analogues way, regional heterogeneities are accounted for by the inclusion of federal state dummies for the establishment's location (10 categories). Once again, these firm-level measures are aggregated over an individual's biography and hence, refer to the type of firm an individual worked at mostly.

## 3.4 Econometric methods

The subsequent analysis builds on Recentered-Influence-Function (RIF) decomposition to disentangle the increasing inequality in UA40 earnings between pooled cohorts 1955-57 and 1972-74.[16] The method represents an extension of the well-known Oaxaca-Blinder decomposition that allows to decompose changes in any distributional statistics into a part being due to changes in the distribution of covariates while fixing the corresponding returns (composition effect), and one due to changes in the returns to these covariates leaving the distribution of covariates unchanged (returns effect).[17] Contrary to other decomposition techniques, the major advantage of RIF decomposition lies in the fact that it is the only method that allows for both a path-independent and detailed decomposition of any distributional statistic of interest.[18] Hence, it allows to link changes in a number of inequality measures (85-15/85-50/50-15 log wage gaps, Gini, log variance) to the different covariates outlined in the previous chapter.

The method itself is based on unconditional quantile regression as introduced in the seminal contribution by Firpo et al. (2009). The main idea is to run regressions of the recentered influence function of some distributional statistic of interest $\nu$ on explanatory variables. The RIF is a recented version of the influence function defined as $RIF\left(y,\nu\right) = \nu + IF\left(y;\nu\right)$. It can easily be shown that the RIF has the same expectation as the original statistic of interest $\nu$ and integrates to $\nu$ as $\int RIF\left(y;\nu\right) dF(y) = \int \left(\nu + IF\left(y;\nu\right)\right) dF(y) = \nu\left(F_y\right)$, where $F_y$ is the distribution function of the dependent variable. Assuming that the conditional expectation of the RIF is a linear function of the explanatory variables, the RIF is modeled as $E\left[RIF\left(Y;\nu\right)|X\right] = X\gamma$, where $\gamma$ can be estimated by OLS.[19] Given this linear

[16]Section 3.4 in parts follows Biewen and Seckler (2017, 2019). For a more in-depth description of RIF decomposition, also see Firpo et al. (2009, 2018).

[17]The decomposition literature often uses the term *wage structure effect*. However, as this paper analyzes long-term and lifetime earnings, as opposed to wages, the suggested terminology is used.

[18]See Fortin et al. (2011) for a comprehensive overview on alternative techniques.

[19]Fig. 1B in Firpo et al. (2009) shows that modeling the RIF as a linear function of covariates yields

specification, an Oaxaca-Blinder decompositions using the RIF regression coefficients can be used to split up the overall change $\Delta_O^\nu$ in a distributional statistic of interest $\nu$ into a composition $\Delta_X^\nu$ and a returns effect $\Delta_S^\nu$

$$\Delta_O^\nu = \underbrace{\nu\left(F_{Y_0|c=1}\right) - \nu\left(F_{Y_0|c=0}\right)}_{\Delta_X^\nu} + \underbrace{\nu\left(F_{Y_1|c=1}\right) - \nu\left(F_{Y_0|c=1}\right)}_{\Delta_S^\nu}, \tag{3.1}$$

where $F_{Y_0|c=s}$, $F_{Y_1|c=s}$ denote the distributions of UA40 earnings among workers in cohort $s$ receiving the returns to characteristics of cohort 0 and cohort 1, respectively.

Due to their linear specification, the RIFs are only local approximations which potentially leads to biased results in case of large changes in the distribution of characteristics.[20] This shortcoming is addressed by a refined version of the decomposition suggested in Firpo et al. (2014, 2018), which additionally incorporates inverse probability weighting (DiNardo et al., 1996). The main idea lies in the creation of an artificial cohort 01, in which the cohort 0 distribution of characteristics $X$ is reweighted to that of the target cohort 1. Using two separate Oaxaca-Blinder decompositions, the overall change $\Delta_O^\nu$ is split up into four components

$$\Delta_O^\nu = \underbrace{\left(\bar{X}_{01} - \bar{X}_0\right)\hat{\gamma}_0^\nu}_{\Delta_{X,p}^\nu} + \underbrace{\bar{X}_{01}\left(\hat{\gamma}_{01}^\nu - \hat{\gamma}_0^\nu\right)}_{\Delta_{X,c}^\nu} + \underbrace{\bar{X}_1\left(\hat{\gamma}_1^\nu - \hat{\gamma}_{01}^\nu\right)}_{\Delta_{S,p}^\nu} + \underbrace{\left(\bar{X}_1 - \bar{X}_{01}\right)\hat{\gamma}_{01}^\nu}_{\Delta_{S,c}^\nu}. \tag{3.2}$$

where $\Delta_{X,p}^\nu$ denotes the estimate for the detailed composition effect, i.e. the effect from changing the distribution of a certain group of covariates while fixing its returns (at the level of cohort 0). For instance, the detailed composition effect linked to part-time employment would reflect the change in $\nu$ that results from changing the distribution of UA40 part-time

very similar results compared to more flexible specifications in the case of quantiles. The usage of a linear specification is also recommended in Firpo et al. (2018).

[20] As outlined in Firpo et al. (2014), this would for example be the case if the underlying true relationship between $Y$ and $X$ was in fact convex (and not linear as assumed by OLS). In such a scenario, an upward-shift of the distribution of $X$ would mechanically increase the estimated coefficients even if the true return structure remained unaltered.

spells of cohort 0 to that of cohort 1. The term $\Delta^{\nu}_{X,c}$ denotes the specification error that reflects differences in the estimated RIF coefficients between the cohorts 01 and 0. In other words, it corresponds to the difference between the linear approximation of the composition effect estimated by RIF decomposition and the estimate of the composition effect received from applying DiNardo et al (1996)-reweighting. Hence, a small value for the specification error indicates that a linear approximation of the composition effect is appropriate. The term $\Delta^{\nu}_{S,p}$ denotes the detailed returns effects which capture the effect from changes in $\gamma$ for a certain group of covariates. As $\gamma$ is estimated from unconditional (as opposed to conditional) quantile regression, it represents changes both between and within subgroups. Lastly, $\Delta^{\nu}_{S,c}$ represents the reweighting error that stems from differences in the distribution of covariates between cohort 1 and the reweighted base cohort 01 and should, in case the reweighting procedure was successful, be close to zero.

Fortin et al. (2011), among others, point out that the detailed decomposition results of the returns effect for groups of categorical variables depend arbitrarily on the choice of the omitted reference group. To address this concern, RIF regression coefficients are normalized such that they sum up to zero within a group of categorical variables $J$, i.e. $\sum_{j\in J}\gamma_j = 0$ (see, Gardezabal and Ugidos, 2004), effectively making the results independent of the chosen reference group. As another advantage, this kind of normalization facilitates the interpretation of results as information on the general level of $\nu$ are captured by the intercept, whereas the regression coefficients mirror deviations of individual categories from this general level. Accordingly, the intercept also captures changes in the relative importance of different groups of covariates as well as the contribution of unobservable factors (see Biewen and Seckler, 2017, 2019, for a more rigorous discussion).

Finally, note that the results from RIF decomposition should not be interpreted as causal effects. This is due to the fact that statistical decomposition techniques (including RIF decomposition) do not account for general equilibrium effects, as they generally assume invariance of the conditional distribution. Similarly, the method does not account for the fact that different explanatory factors might be dynamically related, i.e. changes in one group of covariates (e.g. job mobility) might be the result of changes in another group (e.g. education). Despite these limitations, RIF decomposition represents a highly useful tool to deepen the understanding of what factors are associated with the observed changes in the distribution of individual long-term and lifetime earnings.

## 3.5 Decomposition results

**Table 3.1** – Groups of covariates

| Group | Covariates |
|---|---|
| 1. Non-employment | Years of non-employment UA40 (= days of full-time employment/365) |
| 2. Part-time employment | Years of part-time employment UA40 (= days of part-time employment/365) |
| 3. Education | Highest educational degree UA40 (6 categories) |
| 4. Occupation | Most frequent occupation UA40 (32 categories) |
| 5. Nationality | German by birth (binary, no spells with foreign nationality) |
| 6. Job mobility | Number of firm changes UA40 (with change in both occupation/industry)<br>Number of firm changes UA40 (without change in both occupation/industry) |
| 7. Firm | Most frequent firmsize UA40 (3 categories)<br>Mostly in high-tech firm UA40 (binary)<br>Most frequent sector UA40 (44 categories)<br>Most frequent federal state UA40 (10 categories) |

This section presents RIF decomposition results comparing pooled cohorts 1955-57 and 1972-74. For reasons of clarity, the previously presented covariates are summarized in seven groups in line with table 3.1. For the baseline model, these are *Non-employment*, *Part-time employment*, *Education*, *Occupation*, *Job mobility*, *Nationality* and *Firm*. In the presentation of results, it is insightful to start with a graphical analysis. Results of an alternative specification restricted to German nationals are provided in table B3 in the appendix. As these are very similar, they are not discussed in more detail at this point. In order to highlight the differences due to delayed labor market entry as a consequence of the educational expansion, three sets of results are presented: one for up to age 40 earnings, one controlling in addition for age at labor market entry and one for earnings received between age 25 and 40.

### 3.5.1 Results for UA40

Figure 3.10 includes the total change in unconditional quantiles together with the aggregate composition and returns effect. The total change in unconditional quantiles was characterized by a monotonic development in the sense that unconditional quantiles below the median suffered losses in terms of UA40, whereas the upper half gained. In this regard, the development somewhat resembles previous findings on inequality in daily/hourly earnings. However, note once more the stagnation in median earnings which is in contrast to significant long-run gains in median daily/hourly earnings (see, e.g. Dustmann et al., 2009). The aggregate composition effect reveals a similarly monotonic pattern, but was negative for most of the distribution and only had a weakly positive effect above the 70th percentile. In a similar way, the aggregate returns effect is found to be positive above the 40th percentile and negative in the lower part of the distribution.

Figure 3.11 further disentangles the overall composition effect by displaying the detailed effects linked to the groups of covariates. The graph shows strong composition effects linked to changing employment patterns as well as education. The increasing incidence of both part-time and non-employment spells played an important role at the bottom of the UA40 distribution. Interestingly and in line with the descriptive evidence, the effect linked to the expansion of part-time employment was even slightly stronger than the effect associated with the increasing incidence of non-employment. Note also that both effects, despite being strongest at the bottom, had a negative effect on most parts of the distribution. Being the most important individual composition effect (but smaller than the joint effect from changing employment patterns), compositional changes in education led to an upward shift of the

**Figure 3.10** – Aggregate decomposition, cohorts 1955-57 vs. 1972-74

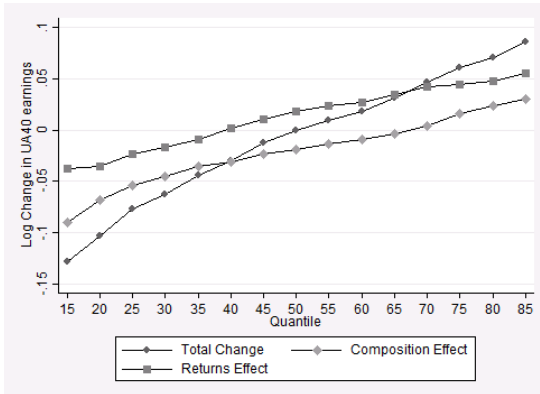



**Figure 3.11** – Detailed composition effect, cohorts 1955-57 vs. 1972-74

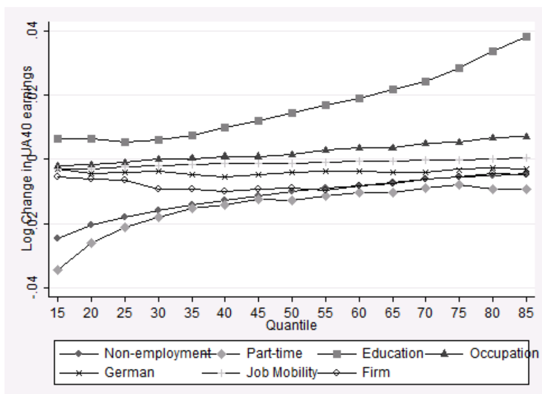
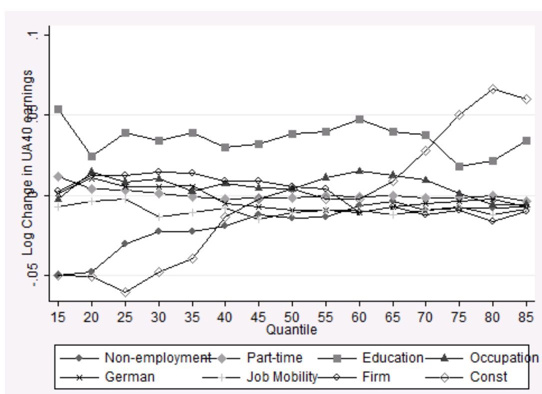



**Figure 3.12** – Detailed returns effect, cohorts 1955-57 vs. 1972-74




Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.

UA40 distribution across all quantiles. It is found to be the most important single factor for inequality at the top. In this regard, both the results on changing employment patterns and educational upgrading are in line with the general trends described in the preceding chapters. The analysis also reveals a moderate composition effect linked to changes in the occupational background as well as a minor effect of job mobility, with the other factors being rather negligible.

Figure 3.12 provides detailed results for the returns effect, which seems to be similarly important when compared to the overall composition effect. Besides the constant, the graph indicates an important contribution from changes in the returns to non-employment that led to a downward shift of the lower half of the distribution. In other words, besides being more likely to be affected by non-employment episodes, later birth cohorts equally faced greater losses in terms of long-term earnings following an episode of non-employment. A possible interpretation would be that these cohorts found it increasingly difficult to re-integrate into the labor market after an episode of non-employment, which might potentially reflect the difficult labor market conditions in the late 1990s (which the individuals in the cohort 1972-74 had to face at an early stage of their career), but might as well reflect factors such as a faster human capital depreciation or a lower job match quality upon re-entry. Note that this finding should be interpreted with some caution due to the relatively large standard errors shown in table 3.2. The picture also suggests a positive returns effect linked to education, which shifted up the entire within-cohort distribution. As its effect is very homogenous, no significant effect on the different inequality measures is found (see, table 3.2).

The analysis also reveals an important contribution of a general returns effect as captured by the constant, which had a very negative impact on the bottom of the distribution and was very favorable for the top. As argued in section 3.4, the constant captures that part of the returns effect that cannot be attributed to the characteristics included in the decomposition, but might as well reflect changes in the relative importance of different groups of covariates.

**Table 3.2** – RIF decomposition results, UA40

| Inequality measure | 85-15 | 85-50 | 50-15 | Gini | Logvar |
|---|---|---|---|---|---|
| Total change | 21.35*** | 8.58*** | 12.77*** | 4.91*** | 7.25*** |
| | (1.13) | (0.58) | (0.98) | (0.18) | (0.31) |
| Total composition | 9.05*** | 4.53*** | 4.52*** | 2.24*** | 3.19*** |
| | (0.68) | (0.40) | (0.47) | (0.17) | (0.33) |
| Non-employment | 2.03*** | 0.55*** | 1.48*** | 0.55*** | 0.95*** |
| | (0.31) | (0.09) | (0.23) | (0.09) | (0.15) |
| Part-time | 2.51*** | 0.34** | 2.18*** | 0.69*** | 1.33*** |
| | (0.34) | (0.16) | (0.26) | (0.09) | (0.23) |
| Education | 3.17*** | 2.35*** | 0.82*** | 0.62*** | 0.46*** |
| | (0.33) | (0.26) | (0.19) | (0.07) | (0.09) |
| Occupation | 0.92*** | 0.57*** | 0.35** | 0.19*** | 0.26*** |
| | (0.21) | (0.15) | (0.16) | (0.04) | (0.06) |
| Nationality | 0.00 | 0.10 | -0.10 | 0.08** | 0.15* |
| | (0.19) | (0.13) | (0.19) | (0.04) | (0.08) |
| Job Mobility | 0.38*** | 0.19*** | 0.19*** | 0.08*** | 0.08** |
| | (0.10) | (0.05) | (0.07) | (0.02) | (0.04) |
| Firm | 0.04 | 0.42* | -0.38 | 0.03 | -0.04 |
| | (0.28) | (0.23) | (0.22) | (0.06) | (0.09) |
| Total effect returns | 9.09*** | 3.61*** | 5.47*** | 2.70*** | 4.04*** |
| | (1.40) | (0.73) | (1.23) | (0.19) | (0.40) |
| Non-employment | 4.22* | 0.69 | 3.53* | 0.57** | 3.13*** |
| | (2.24) | (0.81) | (1.90) | (0.29) | (0.74) |
| Part-time | -1.51** | -0.19 | -1.32** | -0.24** | -0.09 |
| | (0.69) | (0.25) | (0.59) | (0.12) | (0.29) |
| Education | -1.97 | -0.46 | -1.51 | -1.06 | -4.27 |
| | (4.95) | (2.77) | (4.03) | (0.92) | (2.23) |
| Occupation | -0.33 | -0.95 | 0.62 | -0.17 | -0.11 |
| | (1.67) | (1.06) | (1.46) | (0.31) | (0.89) |
| Nationality | -0.84 | 0.20 | -1.04 | -0.08 | 0.20 |
| | (1.15) | (0.82) | (1.11) | (0.22) | (0.56) |
| Job Mobility | -0.16 | 0.24 | -0.40 | 0.00 | -0.15 |
| | (1.95) | (0.91) | (1.51) | (0.34) | (0.83) |
| Firm | -1.29 | -1.54 | 0.25 | -0.38 | -1.05 |
| | (1.99) | (1.41) | (1.51) | (0.36) | (0.67) |
| Constant | 10.96* | 5.61 | 5.35 | 4.05*** | 6.37** |
| | (5.93) | (3.73) | (4.80) | (1.10) | (2.73) |
| Specification Error | 2.95*** | 0.33 | 2.62*** | -0.07 | -0.03 |
| | (0.86) | (0.39) | (0.87) | (0.05) | (0.11) |
| Reweighting Error | 0.27 | 0.11 | 0.15 | 0.04 | 0.06 |
| | (0.42) | (0.14) | (0.33) | (0.09) | (0.16) |

Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.
Log wage differentials×100. *** / ** / * statistically significant at 1%/5%/10%-level
Bootstrapped standard errors (100 replications) in parentheses

For example, the constant might represent factors such as systematic differences in earnings dynamics within firms as well as differing idiosyncratic shocks that remain unobservable in administrative data.

Table 3.2 presents the corresponding numerical results for the decomposition of UA40 earnings, which underpin the findings of the preceding graphical analysis. Numerically, both the total composition (9.05) and the total returns effect (9.09) contributed equally to the overall 21.35 log percentage points increase in the 85-15 wage gap, with the specification and reweighting error amounting to 3.22 points. The strongest composition effects were due to changes in educational attainment (3.17 points) as well as changes in part-time (2.51 points) and non-employment (2.03 points) patterns. Further, there seemed to be moderate composition effects linked to changes in the occupational structure (0.92 points) as well as job mobility (0.38 points). The bottom half of the table, displaying detailed results for the returns effect, shows that the estimated effects are generally less precise. Besides the previously described returns effect linked to non-employment, it equally reveals a moderate inequality-reducing effect from changing returns to part-time employment at the bottom of the distribution.

### 3.5.2 Results for UA40 controlling for age at labor market entry

Regarding the question of whether the increasing inequality in lifetime earnings was driven by changes in labor market participation as opposed to changes in earnings received during times of employment, the evidence presented in the previous section suggests that some 21 percent of the overall increase was linked to a lower labor market participation among individuals of later cohorts. However, this baseline decomposition did not control for changes in the age at labor market entry due to its presumable very close relationship with educational upgrading. Hence, the estimate of the effect linked to a lower lifetime labor market participation did not capture the delayed labor market entry of later cohorts as a result of educational upgrading.

**Table 3.3** – RIF decomposition results, UA40, including age at labor market entry

| Inequality measure | 85-15 | 85-50 | 50-15 | Gini | Logvar |
|---|---|---|---|---|---|
| Total change | 21.35*** | 8.58*** | 12.77*** | 4.91*** | 7.25*** |
| | (1.13) | (0.58) | (0.98) | (0.18) | (0.31) |
| Total composition | 9.63*** | 4.83*** | 4.80*** | 2.36*** | 3.37*** |
| | (0.70) | (0.41) | (0.47) | (0.17) | (0.34) |
| Non-employment | 2.05*** | 0.57*** | 1.49*** | 0.55*** | 0.96*** |
| | (0.33) | (0.10) | (0.24) | (0.09) | (0.16) |
| Part-time | 2.38*** | 0.28* | 2.11*** | 0.67*** | 1.28*** |
| | (0.33) | (0.16) | (0.26) | (0.09) | (0.23) |
| Age at labor market entry | 2.27*** | 0.97*** | 1.30*** | 0.47*** | 0.85*** |
| | (0.29) | (0.19) | (0.20) | (0.06) | (0.11) |
| Education | 1.91*** | 1.87*** | 0.04 | 0.36*** | -0.05 |
| | (0.35) | (0.28) | (0.22) | (0.08) | (0.11) |
| Occupation | 0.84*** | 0.54*** | 0.30* | 0.17*** | 0.22*** |
| | (0.22) | (0.16) | (0.16) | (0.05) | (0.06) |
| Nationality | -0.15 | 0.04 | -0.19 | 0.05 | 0.09 |
| | (0.18) | (0.13) | (0.18) | (0.04) | (0.08) |
| Job Mobility | 0.37*** | 0.19*** | 0.19*** | 0.08*** | 0.08** |
| | (0.10) | (0.05) | (0.07) | (0.02) | (0.04) |
| Firm | -0.05 | 0.38 | -0.43* | 0.02 | -0.07 |
| | (0.29) | (0.24) | (0.22) | (0.06) | (0.09) |
| Total effect returns | 8.63*** | 3.66*** | 4.97*** | 2.68*** | 3.98*** |
| | (1.36) | (0.75) | (1.20) | (0.19) | (0.41) |
| Non-employment | 3.70 | 0.82 | 2.87 | 0.57** | 3.09*** |
| | (2.31) | (0.78) | (1.97) | (0.29) | (0.74) |
| Part-time | -1.43** | -0.16 | -1.27** | -0.22* | -0.05 |
| | (0.68) | (0.25) | (0.59) | (0.12) | (0.29) |
| Age at labor market entry | 23.39 | 30.01* | -6.62 | 6.47* | 9.30 |
| | (19.98) | (12.19) | (16.11) | (3.71) | (7.30) |
| Education | -1.00 | 0.41 | -1.41 | -0.77 | -3.83* |
| | (4.87) | (2.70) | (4.05) | (0.89) | (2.19) |
| Occupation | 0.61 | -0.86 | 1.46 | -0.11 | -0.07 |
| | (1.68) | (1.09) | (1.43) | (0.31) | (0.88) |
| Nationality | -1.83 | 0.23 | -2.06* | -0.42* | -1.01* |
| | (1.18) | (0.82) | (1.11) | (0.22) | (0.58) |
| Job Mobility | -1.56 | -2.41*** | 0.85 | -0.22 | -0.24 |
| | (1.92) | (0.92) | (1.48) | (0.33) | (0.82) |
| Firm | 1.04 | 1.04 | -0.00 | 0.25 | 0.35 |
| | (2.02) | (1.41) | (1.55) | (0.36) | (0.66) |
| Constant | -14.28 | -25.43 | 11.15 | -2.87 | -3.57 |
| | (21.76) | (13.42) | (17.56) | (4.02) | (8.17) |
| Specification error | 3.30*** | 0.13 | 3.16*** | -0.09* | -0.02 |
| | (0.86) | (0.40) | (0.86) | (0.05) | (0.12) |
| Reweighting error | -0.21 | -0.04 | -0.16 | -0.04 | -0.08 |
| | (0.44) | (0.15) | (0.35) | (0.09) | (0.16) |

Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.
Log wage differentials×100. *** / ** / * statistically significant at 1%/5%/10%-level
Bootstrapped standard errors (100 replications) in parentheses

An alternative model specification that equally controls for the age at labor market entry is provided in table 3.3. The results of this alternative specification suggest that up to 31 percent of the increase in 85-15 might in fact be due to the overall lower lifetime labor market participation (i.e. due to the joint effect from changes in non-employment/part-time/age at labor market entry). However, the composition effect linked to education equally shrinks significantly in this specification, confirming the a-priori expectation of both effects being closely related.

### 3.5.3 Results for earnings between ages 25 and 40

Looking more closely at earnings between ages 25 and 40, i.e. only considering earnings from an age where most individuals already entered the labor market, reveals further valuable insights. The corresponding decomposition results are presented in table 3.4. This earnings measure is better comparable with the literature which typically considered earnings starting at age 25 (e.g. Guevenen et al., 2017). Overall, the results point towards a greater importance of composition effects (11.71 points), which are found to be more important than the overall returns effect (6.49 points) in explaining the overall increase of 20.01 points in terms of the 85-15 log wage differential. Most importantly, there seems to be a much stronger composition effect linked to education explaining up to 32 percent (or 6.38 points) in terms of 85-15 and up to 87 percent at the top of the distribution. In fact, this finding does not come as a surprise given that this specification does not account for forgone earnings during times of education. Hence, the educational expansion has a mechanically stronger effect on inequality between age 25 and 40 (as opposed to UA40). This is accompanied by a reduction in the composition effect linked to non-employment, which likely reflects both a generally higher incidence of unemployment among very young individuals as well as the fact that parts of the increase in non-employment at young ages might be due to the additional time spent in education (though only non-employment spells after labor market entry are

**Table 3.4** – RIF decomposition results, earnings age 25-40

| Inequality measure | 85-15 | 85-50 | 50-15 | Gini | Logvar |
|---|---|---|---|---|---|
| Total change | 20.01*** | 7.84*** | 12.17*** | 4.82*** | 7.31*** |
| | (1.21) | (0.90) | (0.97) | (0.20) | (0.36) |
| Total composition | 11.71*** | 7.88*** | 3.83*** | 2.08*** | 2.66*** |
| | (0.99) | (0.66) | (0.63) | (0.17) | (0.36) |
| Non-employment | 1.40*** | 0.02 | 1.38*** | 0.34*** | 0.71*** |
| | (0.39) | (0.03) | (0.39) | (0.10) | (0.20) |
| Part-time | 2.17*** | -0.47* | 2.64*** | 0.49*** | 1.07*** |
| | (0.45) | (0.26) | (0.34) | (0.09) | (0.21) |
| Education | 6.38*** | 6.79*** | -0.41** | 0.92*** | 0.62*** |
| | (0.62) | (0.60) | (0.19) | (0.09) | (0.11) |
| Occupation | 1.40*** | 0.92*** | 0.47*** | 0.18*** | 0.26*** |
| | (0.34) | (0.25) | (0.17) | (0.05) | (0.07) |
| Nationality | -0.52** | -0.23 | -0.29 | -0.06 | -0.11 |
| | (0.22) | (0.17) | (0.19) | (0.04) | (0.09) |
| Job Mobility | 1.04*** | 0.54*** | 0.49*** | 0.20*** | 0.23*** |
| | (0.18) | (0.10) | (0.13) | (0.03) | (0.08) |
| Firm | -0.15 | 0.30 | -0.46** | 0.00 | -0.13 |
| | (0.42) | (0.34) | (0.20) | (0.07) | (0.11) |
| Total effect returns | 6.49*** | -0.24 | 6.73*** | 2.79*** | 4.42*** |
| | (1.42) | (0.94) | (1.07) | (0.21) | (0.45) |
| Non-employment | 5.39** | 2.29*** | 3.10 | 0.38* | 2.34*** |
| | (2.55) | (0.85) | (2.30) | (0.23) | (0.82) |
| Part-time | -2.31*** | -0.17 | -2.13*** | -0.31*** | -0.26 |
| | (0.71) | (0.31) | (0.65) | (0.10) | (0.28) |
| Education | 2.47 | -1.72 | 4.19 | -0.29 | -2.91 |
| | (5.87) | (3.75) | (4.31) | (0.86) | (1.95) |
| Occupation | -1.62 | -3.34** | 1.73 | -0.48 | -0.41 |
| | (2.29) | (1.44) | (1.62) | (0.31) | (0.81) |
| Nationality | -0.55 | 0.42 | -0.97 | -0.11 | 0.30 |
| | (1.46) | (1.11) | (1.14) | (0.22) | (0.68) |
| Job Mobility | -2.15 | -0.89 | -1.26 | -0.33 | -1.32 |
| | (2.31) | (1.24) | (1.44) | (0.36) | (1.16) |
| Firm | -3.01 | -0.45 | -2.56 | -0.24 | -1.03 |
| | (2.51) | (1.75) | (1.82) | (0.38) | (0.67) |
| Constant | 8.27 | 3.63 | 4.64 | 4.17*** | 7.71*** |
| | (6.97) | (4.70) | (5.68) | (1.04) | (2.50) |
| Specification error | 1.27* | -0.08 | 1.35*** | -0.12** | 0.07 |
| | (0.68) | (0.49) | (0.44) | (0.05) | (0.13) |
| Reweighting error | 0.55 | 0.28 | 0.27 | 0.07 | 0.16 |
| | (0.52) | (0.23) | (0.44) | (0.09) | (0.18) |

Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.
Log wage differentials$\times$100. *** / ** / * statistically significant at 1%/5%/10%-level
Bootstrapped standard errors (100 replications) in parentheses

counted as non-employment). Note also that the overall share explained by the increasing incidence of part-time employment remains virtually unchanged. At the same time, there is a moderate increases in the relative importance of compositional effects linked to occupations and job mobility when compared to the decomposition of UA40.

As to the returns effect, the overall picture remains mostly unchanged with a persistently strong effect linked to non-employment (5.39 points in terms of 85-15). At the same time, the previously found returns effect linked to part-time employment becomes more pronounced, suggesting that it had an inequality-reducing effect of -2.13 points (or -18 percent in terms of 50-15) at the bottom of the distribution, but hardly any effect in the upper half of the distribution.

## 3.6 Summary and Discussion

This study has investigated potential determinants of increasing lifetime earnings inequality using detailed employment biographies of West German men born between the years 1955 and 1974. Adopting a perspective based on cohorts, the paper contributes to a comparatively small but growing literature documenting an increasing inequality in individual long-term and lifetime earnings (Bönke et al., 2015a, Guevenen et al., 2017). The paper confirms major trends previously documented for Germany using an alternative data base. It goes beyond previous contributions by formally disentangling these changes by means of a detailed decomposition analysis based on RIF regression.

The empirical results suggest that a lower labor market participation of younger cohorts explains some 20-30 percent of the overall increase in lifetime earnings inequality, with the effect being mostly limited to the lower half of the distribution. Compared to the findings in Bönke et al. (2015a), the analysis assigns a smaller part of this effect to non-employment periods and instead highlights the growing importance of part-time employment. Nevertheless, this results is not at odds with the results in Bönke et al. (2015a) due to the additional cohorts included in the present study. The findings presented in the preceding chapters also complement Biewen et al. (2018) by showing that the increasing incidence of part-time employment among German men does not only explain increasing inequality in cross-sectional earnings, but also adds substantially to the increasing inequality in lifetime earnings. At the same time, changing employment patterns can only partly explain the losses in UA40 earnings at the bottom of the distribution. Hence, this points towards some similarities with the development in the U.S. for which Guevenen et al. (2017) showed that losses in lifetime earnings among later cohorts are mostly due to a decline in earnings conditional on employment.

Beside changing employment patterns, composition effects linked to the educational upgrading of younger cohorts explain another 15-30 percent of the increasing dispersion in lifetime earnings. Importantly, these changes were favorable for all parts of the distribution but more favorable for individuals at the top, thereby increasing inequality in the upper half of the distribution. Beyond educational upgrading, the analysis finds only limited evidence of skill-biased technological change (SBTC). As such, only a moderate impact from changes in the composition of occupations in the range of 4-7 percent and mostly insignificant results regarding their returns are found.

The analysis also points towards a potentially important returns effect linked to episodes of non-employment, which had an adverse effect on long-term earnings for individuals in the lower half of the distribution. A natural interpretation of this result is that individuals in later cohorts found it increasingly difficult to re-integrate into the labor market after being non-employed, resulting in stronger long-term earnings losses. Possible mechanisms behind this finding may be a faster depreciation of human capital as well as a poorer job match quality following a period of non-employment.

The present study also provides evidence for a stagnation in UA40, i.e. during a major part of the career, among German men born between the years 1955 and 1974. In this regard, the development in Germany resembles the one in the U.S., though somewhat delayed and less pronounced, for which previous research by Guevenen et al. (2017) documented significant losses in lifetime earnings among men starting already with birth cohort 1942. Importantly, the results of the present paper point towards moderate earnings losses within all educational subgroups (though being strongest for the lowest education group), which were counterbalanced by the educational expansion. This is an interesting finding given significant gains in hourly/daily earnings found in cross-sectional data during study period 1975-2014 (see, e.g. Dustmann et al., 2014). It suggests that the cross-sectional earnings gains were only beneficial to individuals of older cohorts, whereas younger cohorts suffered both a stagnation and increasing inequality in lifetime earnings.

# Appendix B

## Correction for structural break 1983/1984

As outlined in the main part, the information on daily earnings in the *SIAB* is subject to a structural break between the years 1983 and 1984. More precisely, one-time payments such as annual bonuses or Christmas/holiday allowances were not included before 1984, which results in a spurious increase in both the level and dispersion of earnings between both years. The literature suggests different correction methods (see, e.g. Dustmann et al., 2009, Bönke et al., 2015a) which usually build on the technique by Fitzenberger (1999). Being most closely related to the present study, daily earnings are corrected following the procedure suggested in Bönke et al. (2015a). Accordingly, log earnings growth in year t is estimated by a random effects (RE) model of the following form

$$\Delta w_t = \alpha_0 + \alpha_1 D_{1984} + \alpha_2 age_t + \alpha_3 age_t^2 + \alpha_4 age_t^3 + \alpha_5 D_{1984} age_t$$
$$+\alpha_6 D_{1984} age_t^2 + \alpha_7 D_{1984} age_t^3 + \boldsymbol{D_q'}\boldsymbol{\beta} + \boldsymbol{D_q'}\boldsymbol{\gamma} D_{1984} + \boldsymbol{D_q'}\boldsymbol{\delta} age_t + \epsilon$$

where $\Delta w_t$ denotes the growth in log earnings between time periods t and t+1 and $D_{1984}$ a dummy variable indicating the structural break. The model also includes a set of dummy variables $\boldsymbol{D_q}$ for an individual's average rank in the earnings distribution between age 35 and 40, which intends to approximate an individual's permanent position in the earnings distribution and accounts for the previous finding by Fitzenberger (1999) that the effect of one-time payments is more important for the upper part of the earnings distribution. Moreover, three polynomials of age as well as their interactions with the structural break dummy $D_{1984}$ are included. Finally, the model includes interactions between the rank dummies $\boldsymbol{D_q}$ and both the structural break dummy $D_{1984}$ and age. These are used to estimate an age and quantile specific spurious growth factor to correct observations before the year 1984.

## Imputation of earnings above the contribution limit

The imputation of daily earnings above the contribution limit is done following the procedure suggested in Gartner (2005). Hence, wages above the censoring point are estimated by a series of tobit models which are computed separately for each year. The regressions include two polynomials of age, six education categories as wells as interactions between age and education. Instead of solely using the expected values from the tobit model, which suffer from a too high correlation with the covariates and downward-biased standard errors in later estimations, daily earnings above the threshold are drawn from a truncated normal distribution. The lower limit of this distribution is given by the censoring threshold, its standard deviation is estimated from the tobit model.

**Table B1** – Observations per cohort

| Cohort | (1) UA40 | (2) UA45 | (3) UA50 |
|---|---|---|---|
| 1955 | 4602 | 4192 | 3751 |
| 1956 | 4894 | 4480 | 4023 |
| 1957 | 4961 | 4525 | 4091 |
| 1958 | 5003 | 4588 | 4158 |
| 1959 | 5283 | 4775 | 4374 |
| 1960 | 5253 | 4801 | 4394 |
| 1961 | 5516 | 5053 | 4592 |
| 1962 | 5736 | 5258 | 4864 |
| 1963 | 5845 | 5368 | 4911 |
| 1964 | 5869 | 5397 | 4918 |
| 1965 | 5795 | 5340 | - |
| 1966 | 5948 | 5484 | - |
| 1967 | 5634 | 5256 | - |
| 1968 | 5351 | 5007 | - |
| 1969 | 5252 | 4798 | - |
| 1970 | 4863 | - | - |
| 1971 | 4555 | - | - |
| 1972 | 4086 | - | - |
| 1973 | 3624 | - | - |
| 1974 | 3517 | - | - |
| Total | 109,194 | 81,271 | 49,864 |

Source: SIAB 1975-2014 and own calculations.

**Table B2** – Descriptive statistics UA40

| | 1955-57 | | 1972-74 | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Non-employment (= days of non-employment/365) | 2.120 | 2.435 | 2.361 | 2.505 |
| Part-time employment (= days of part-time employment/365) | 0.135 | 0.857 | 0.569 | 1.784 |
| Age at labor market entry | 21.069 | 1.950 | 21.557 | 2.045 |
| Lower/middle secondary without vocational training | 0.078 | 0.268 | 0.056 | 0.230 |
| Lower/middle secondary with vocational training | 0.765 | 0.424 | 0.644 | 0.479 |
| Upper secondary without vocational training | 0.002 | 0.047 | 0.006 | 0.078 |
| Upper secondary with vocational training | 0.040 | 0.195 | 0.112 | 0.315 |
| University/Fachhochschule | 0.113 | 0.317 | 0.177 | 0.381 |
| Missing information | 0.002 | 0.049 | 0.005 | 0.068 |
| Number of firm changes (with change in both occupation/industry) | 1.731 | 2.621 | 1.929 | 2.441 |
| Number of firm changes (without change in both occupation/industry) | 2.183 | 2.712 | 2.475 | 3.427 |
| German nationality | 0.890 | 0.314 | 0.779 | 0.415 |
| Firmsize 1-50 | 0.338 | 0.473 | 0.352 | 0.477 |
| Firmsize 51-500 | 0.330 | 0.470 | 0.377 | 0.485 |
| Firmsize 500+ | 0.331 | 0.470 | 0.271 | 0.444 |
| Mostly in high-tech firm | 0.304 | 0.460 | 0.294 | 0.455 |
| Most frequent federal state: | | | | |
| Schlewsig-Holstein | 0.032 | 0.177 | 0.029 | 0.168 |
| Hamburg | 0.028 | 0.165 | 0.030 | 0.169 |
| Lower Saxony | 0.103 | 0.304 | 0.104 | 0.306 |
| Bremen | 0.015 | 0.121 | 0.012 | 0.110 |
| North Rhine-Westphalia | 0.290 | 0.454 | 0.269 | 0.444 |
| Hesse | 0.093 | 0.291 | 0.098 | 0.298 |
| Rhineland-Palatinate | 0.060 | 0.238 | 0.052 | 0.228 |
| Baden-Wuerttemberg | 0.169 | 0.374 | 0.179 | 0.383 |
| Bavaria | 0.187 | 0.390 | 0.209 | 0.407 |
| Saarland | 0.022 | 0.148 | 0.017 | 0.129 |
| Most frequent sector: | | | | |
| Agriculte and Forestry | 0.006 | 0.076 | 0.010 | 0.097 |
| Mining | 0.022 | 0.146 | 0.004 | 0.067 |
| Food products, beverages and tobacco products | 0.026 | 0.158 | 0.025 | 0.156 |
| Textiles | 0.010 | 0.010 | 0.005 | 0.073 |
| Wood and wood products | 0.007 | 0.085 | 0.010 | 0.097 |
| Pulp, paper, paper product | 0.009 | 0.094 | 0.010 | 0.098 |
| Publishing, printing and reproduction of recorded media | 0.019 | 0.135 | 0.012 | 0.110 |
| Coke, refined petroleum products and nuclear fuel | 0.003 | 0.051 | 0.002 | 0.043 |
| Chemicals and chemical products | 0.032 | 0.175 | 0.024 | 0.152 |
| Rubber and plastic products | 0.022 | 0.147 | 0.021 | 0.145 |

| | | | | |
|---|---|---|---|---|
| Other non-metallic mineral products | 0.014 | 0.117 | 0.011 | 0.105 |
| Basic metals | 0.028 | 0.164 | 0.021 | 0.145 |
| Fabricated metal products, except machinery and equipment | 0.043 | 0.202 | 0.039 | 0.193 |
| Machinery and equipment n.e.c. | 0.081 | 0.273 | 0.065 | 0.247 |
| Office machinery and computers | 0.006 | 0.080 | 0.003 | 0.057 |
| Electrical machinery and apparatus | 0.026 | 0.159 | 0.024 | 0.153 |
| Radio, television and communication equipment and apparatus | 0.009 | 0.097 | 0.012 | 0.107 |
| Medical, precision and optical instruments, watches and clocks | 0.024 | 0.154 | 0.020 | 0.140 |
| Motor vehicles, trailers and semi-trailers | 0.055 | 0.229 | 0.060 | 0.238 |
| Other transport equipment | 0.009 | 0.093 | 0.008 | 0.088 |
| Furniture; manufacturing n.e.c. | 0.013 | 0.115 | 0.012 | 0.107 |
| Electricity, Water, Recycling | 0.015 | 0.122 | 0.011 | 0.105 |
| Construction | 0.107 | 0.309 | 0.094 | 0.292 |
| Sale, maintenance, repair of motor vehicles | 0.029 | 0.169 | 0.045 | 0.208 |
| Wholesale trade | 0.064 | 0.246 | 0.063 | 0.243 |
| Retail trade | 0.044 | 0.204 | 0.047 | 0.211 |
| Hotels and restaurants | 0.012 | 0.108 | 0.015 | 0.123 |
| Transportation | 0.027 | 0.161 | 0.024 | 0.152 |
| Supporting and auxiliary transport activities | 0.026 | 0.158 | 0.032 | 0.175 |
| Post and telecommunications | 0.011 | 0.104 | 0.011 | 0.106 |
| Financial intermediation | 0.031 | 0.172 | 0.031 | 0.175 |
| Insurance and pension funding | 0.008 | 0.091 | 0.008 | 0.089 |
| Activities auxiliary to financial intermediation | 0.002 | 0.048 | 0.003 | 0.059 |
| Real estate activities, Renting of machinery and equipment | 0.005 | 0.067 | 0.008 | 0.088 |
| Computer and related activities | 0.005 | 0.072 | 0.028 | 0.150 |
| Research and development | 0.004 | 0.059 | 0.006 | 0.074 |
| Other business activities | 0.034 | 0.180 | 0.078 | 0.269 |
| Public administration and defence; compulsory social security | 0.046 | 0.210 | 0.028 | 0.166 |
| Education | 0.009 | 0.094 | 0.010 | 0.100 |
| Health and social work | 0.031 | 0.174 | 0.041 | 0.199 |
| Sewage and refuse disposal, sanitation and similar activities | 0.006 | 0.076 | 0.006 | 0.075 |
| Activities of membership organizations n.e.c. | 0.008 | 0.089 | 0.006 | 0.076 |
| Recreational, cultural and sporting activities | 0.006 | 0.077 | 0.007 | 0.083 |
| Other service activities | 0.008 | 0.089 | 0.004 | 0.064 |
| Most frequent occupation: | | | | |
| Occ. in agriculture, forestry, and farming | 0.005 | 0.073 | 0.004 | 0.066 |
| Occ. in gardening and floristry | 0.006 | 0.080 | 0.008 | 0.090 |
| Occ. in production and processing of raw materials, glass- and ceramic-making etc. | 0.016 | 0.127 | 0.008 | 0.087 |
| Occ. in plastic-making and -processing, and wood-working and -processing | 0.035 | 0.184 | 0.041 | 0.199 |
| Occ. in paper-making and -processing, printing, and in technical media design | 0.024 | 0.152 | 0.018 | 0.132 |
| Occ. in metal-making and -working, and in metal construction | 0.103 | 0.304 | 0.090 | 0.286 |
| Technical occ. in machine-building and automotive industry | 0.147 | 0.354 | 0.144 | 0.351 |

| | | | | |
|---|---|---|---|---|
| Occ. in mechatronics, energy electronics and electrical engineering | 0.055 | 0.227 | 0.037 | 0.188 |
| Occ. in technical research and development, construction, production planning etc. | 0.045 | 0.207 | 0.050 | 0.218 |
| Occ. in textile- and leather-making and -processing | 0.007 | 0.080 | 0.005 | 0.072 |
| Occ. in beverage production | 0.022 | 0.147 | 0.028 | 0.164 |
| Occ. in construction scheduling, architecture and surveying | 0.011 | 0.106 | 0.009 | 0.093 |
| Occ. in building construction above and below ground | 0.044 | 0.205 | 0.032 | 0.176 |
| Occ. in interior construction | 0.021 | 0.144 | 0.020 | 0.142 |
| Occ. in building services engineering and technical building services | 0.026 | 0.160 | 0.029 | 0.167 |
| Occ. in mathematics, biology, chemistry, physics, geography, geology etc | 0.031 | 0.174 | 0.026 | 0.158 |
| Occ. in computer science, information and communication technology | 0.018 | 0.133 | 0.035 | 0.184 |
| Occ. in traffic and logistics (without vehicle driving) | 0.064 | 0.245 | 0.072 | 0.259 |
| Drivers and operators of vehicles and transport equipment | 0.071 | 0.257 | 0.049 | 0.216 |
| Occ. in safety and health protection, security and surveillance | 0.009 | 0.092 | 0.010 | 0.098 |
| Occ. in cleaning services | 0.004 | 0.064 | 0.008 | 0.089 |
| Occ. in purchasing, sales and trading | 0.028 | 0.164 | 0.028 | 0.164 |
| Sales occ. in retail trade | 0.018 | 0.133 | 0.028 | 0.164 |
| Occ. in tourism, hotels and restaurants | 0.006 | 0.074 | 0.007 | 0.083 |
| Occ. in business management and organisation | 0.086 | 0.280 | 0.103 | 0.303 |
| Occ. in financial services, accounting and tax consultancy | 0.046 | 0.209 | 0.048 | 0.213 |
| Occ. in law and public administration | 0.003 | 0.053 | 0.008 | 0.081 |
| Medical and health care occupations | 0.015 | 0.120 | 0.024 | 0.155 |
| Occ. in non-medical healthcare, body care, wellness and medical technicians | 0.006 | 0.077 | 0.003 | 0.058 |
| Occ. in education and social work, housekeeping, and theology | 0.012 | 0.109 | 0.012 | 0.110 |
| Occ. in teaching and training | 0.005 | 0.073 | 0.005 | 0.068 |
| Occ. in humanities, social sciences, economics, media etc. | 0.012 | 0.109 | 0.014 | 0.116 |

Source: Sample of Integrated Labour Market Biographies (SIAB) 1975 - 2014 and own calculations.

Numbers refer to individuals with valid UA40 biography

**Table B3** – RIF decomposition results, UA40, German nationals

| Inequality measure | 85-15 | 85-50 | 50-15 | Gini | Logvar |
|---|---|---|---|---|---|
| Total change | 19.07*** | 8.33*** | 10.74*** | 4.53*** | 6.49*** |
| | (1.15) | (0.67) | (0.91) | (0.19) | (0.33) |
| Total composition | 9.22*** | 4.76*** | 4.46*** | 2.14*** | 2.82*** |
| | (0.69) | (0.39) | (0.47) | (0.16) | (0.27) |
| Non-employment | 1.94*** | 0.52*** | 1.42*** | 0.54*** | 0.94*** |
| | (0.31) | (0.09) | (0.23) | (0.09) | (0.15) |
| Part-time | 2.48*** | 0.35** | 2.13*** | 0.68*** | 1.18*** |
| | (0.34) | (0.16) | (0.26) | (0.09) | (0.18) |
| Education | 3.42*** | 2.60*** | 0.82*** | 0.64*** | 0.44*** |
| | (0.35) | (0.28) | (0.22) | (0.07) | (0.10) |
| Occupation | 1.12*** | 0.67*** | 0.45*** | 0.20*** | 0.29*** |
| | (0.22) | (0.16) | (0.17) | (0.05) | (0.06) |
| Job Mobility | 0.30*** | 0.16*** | 0.14** | 0.06** | 0.06 |
| | (0.11) | (0.05) | (0.07) | (0.03) | (0.05) |
| Firm | -0.04 | 0.45* | -0.49** | 0.02 | -0.09 |
| | (0.30) | (0.25) | (0.24) | (0.07) | (0.09) |
| Total effect returns | 7.75*** | 3.22*** | 4.53*** | 2.61*** | 3.89*** |
| | (1.28) | (0.76) | (0.96) | (0.20) | (0.39) |
| Non-employment | 5.06** | 0.67 | 4.40** | 0.69** | 3.23*** |
| | (2.09) | (0.83) | (1.75) | (0.31) | (0.73) |
| Part-time | -1.27* | -0.26 | -1.01* | -0.21* | 0.09 |
| | (0.65) | (0.25) | (0.56) | (0.12) | (0.25) |
| Education | 0.81 | -1.12 | 1.93 | -1.09 | -4.38* |
| | (6.30) | (4.05) | (4.90) | (1.05) | (2.65) |
| Occupation | -1.88 | -2.18** | 0.30 | -0.41 | -0.50 |
| | (1.87) | (1.06) | (1.59) | (0.32) | (0.73) |
| Job Mobility | -1.66 | 0.42 | -2.08 | -0.11 | -0.32 |
| | (2.00) | (0.94) | (1.46) | (0.38) | (0.83) |
| Firm | -1.43 | -2.07 | 0.64 | -0.34 | -0.58 |
| | (1.92) | (1.59) | (1.39) | (0.35) | (0.64) |
| Constant | 8.11 | 7.77 | 0.34 | 4.08*** | 6.35** |
| | (6.99) | (4.73) | (5.60) | (1.21) | (3.10) |
| Specification error | 2.20*** | 0.30 | 1.90*** | -0.16*** | -0.12 |
| | (0.67) | (0.40) | (0.62) | (0.05) | (0.09) |
| Reweighting error | -0.10 | 0.05 | -0.14 | -0.06 | -0.11 |
| | (0.38) | (0.13) | (0.29) | (0.08) | (0.13) |

Source: Sample of Integrated Labour Market Biographies (SIAB) 1975-2014 and own calculations.
Log wage differentials × 100. Bootstrapped standard errors (100 replications) in parentheses
*** / ** / * statistically significant at 1%/5%/10%-level

**Figure B1** – Indexed real growth in earnings age 25-40

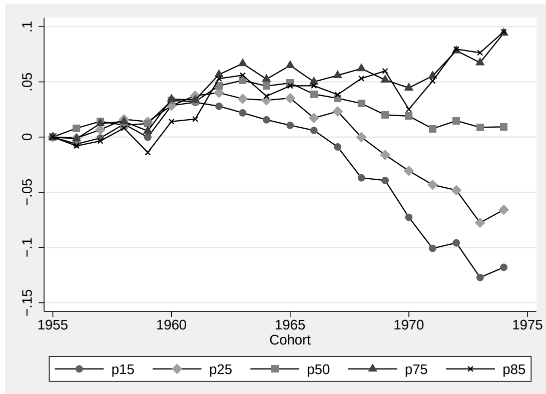
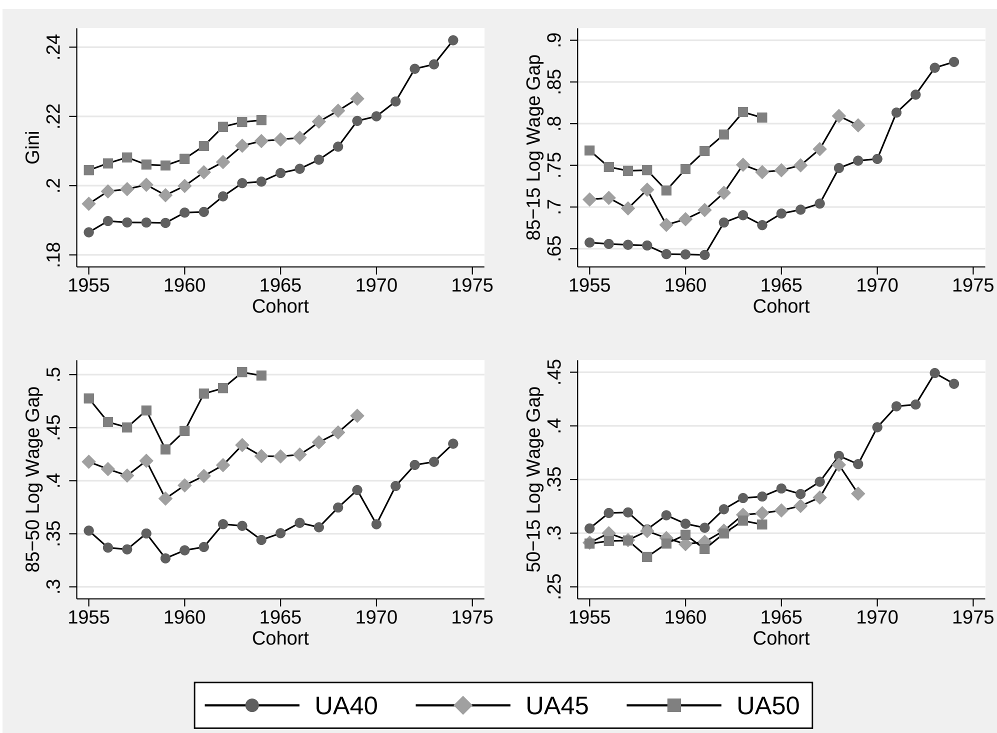



**Figure B2** – Inequality in earnings age 25-40



# Chapter 4

# Counterfactual Quantile Decompositions with Selection Correction Taking into Account Huber/Melly (2015): An Application to the German Gender Wage Gap*

## 4.1 Introduction

Decomposition analyses beyond pure comparison of means have become increasingly popular in applied economic research (for an overview, see Fortin et al., 2011). One particularly popular approach proposed by Machado and Mata (2005) and Melly (2005) uses conditional

*This chapter is based on:
Biewen, M., Fitzenberger, B., and M. Seckler (2019): Counterfactual Quantile Decompositions with Selection Correction Taking into Account Huber/Melly (2015): An Application to the German Gender Wage Gap. Unpublished manuscript, University of Tübingen and Humboldt-University Berlin.

quantile regression to estimate marginal wage distributions for different counterfactual scenarios. Building on the control function approach of Buchinsky (1998, 2001), this technique was extended by Albrecht et al. (2009) to account for selection on unobservables. An important application of the Albrecht et al. (2009) method is the gender wage gap for which a number of authors have assessed the impact of female selection into the labor market (e.g., Albrecht et al., 2009, Chzhen and Mumford, 2011).[1]

However, the use of this method has been called into question by Huber and Melly (2015) who pointed out that Buchinsky's (1998, 2001) control function approach is only valid if regressors and error terms are independent conditional on selection probabilities. This rules out heterogenous slope coefficients, which are often the motivation for using quantile regression approaches. Based on this idea, Huber and Melly (2015) developed a test of the conditional independence assumption which should be carried out before applying Buchinsky (1998, 2001) selection corrections. If the Huber and Melly (2015) test passes, then the results of an Albrecht et al. (2009) analysis can be trusted. If the test rejects, however, users of this method are in the uncomfortable position of applying a method whose assumptions are likely to be violated.

The aim of this paper is to propose a transformation of the original quantile regression model aimed at eliminating violations of conditional independence.[2] For this transformation, we draw on Chen and Khan (2003) who used a similar transformation for the analysis of a sample selection model for the mean in the presence of heteroscedasticity. We demonstrate

[1] Bollinger et al. (2011) also apply the control function approach of Buchinsky to estimate counterfactual wage distributions based on selection corrected quantile regressions. More recently, a number of alternative methods to account for unobserved selectivity in distributional analysis have been proposed, see the literature review below. Some of these methods are more general but also much less practical than the Albrecht et al. (2009) method.

[2] The idea to transform the dependent variable is similar to the approach applied in the companion paper Fitzenberger and de Lazzer (2019), which estimates the selection bias in employment for male workers to account for unemployment. However, the actual implementation of the transformation approach differs between the two papers. Fitzenberger and de Lazzer (2019) estimate selection corrected slope coefficients based on an identification-at-infinity assumption. Further, the empirical application and the dataset used also differ strongly between the two papers.

theoretically and empirically that such a transformation may be successful in arriving at a model in which conditional independence holds and the Huber and Melly (2015) test passes. Our application of this approach to the gender wage gap in Germany provides a real data case in which the Huber/Melly test passes in the transformed model but overwhelmingly rejects in the untransformed model.

This paper makes two contributions. First, we offer users of the Albrecht et al. (2009) method a way to circumvent the Huber and Melly (2015) critique. Second, we present new results on the gender wage gap in Germany taking into account selection on unobservables. As Albrecht et al. (2009) for the Netherlands and Chzhen and Mumford (2011) for the UK, we find positive selection on unobservables leading to an underestimation of the gender gap. However, our results also suggest that this underestimation is more severe compared to that in an untransformed model that is susceptible to the Huber and Melly (2015) critique.

The rest of the paper is structured as follows. Section 4.2 discusses the related literature. Section 4.3 describes the data used in our empirical analysis, which were taken from the German Socio-Economic Panel. In section 4.4, we propose the transformation aimed at eliminating violations from conditional independence. Section 4.5 presents our empirical findings. Section 4.6 concludes.

## 4.2 Related literature

The following literature review specifically deals with contributions that aim at accounting for selection on unobservables when measuring differences in distributions between population groups such as the gender wage gap. For an overview of research on the gender gap in general, see the recent reviews by Blau and Kahn (2017) and Kunze (2017).

A common approach in the literature to accounting for potential selectivity in female labor supply is the imputation of wages of non-working women according to different imputation rules, e.g. Blau and Kahn, 2006, and Olivetti and Petrongolo, 2008. Using this approach, Olivetti and Petrongolo (2008) find an increase in median wage gaps after taking into account selection. Mulligan and Rubinstein (2008) study the gender wage gap in the U.S. using selectivity corrections based on Heckman's two-step estimator combined with an identification-at-infinity approach. Contrary to some of the more recent literature, these studies concentrate on the effect of selection on mean or median wage gaps.

Regarding the effect of selection along the whole distribution, a first strand of the literature builds on the seminal contributions by Buchinsky (1998, 2001) who suggests a selection correction technique for conditional quantile regression based on a control function approach. Buchinsky (1998, 2001) applies this method to study the impact of selectivity on female wages in the U.S. In an important contribution, Albrecht et al. (2009) combine the Buchinsky correction with the popular Machado and Mata (2005) decomposition method to estimate counterfactual wage distributions for the Netherlands (see also Bollinger et al. (2011) for a similar methodological approach addressing a different economic question). Chzhen and Mumford (2011) present a related analysis for the U.K. Both studies find sizeable positive selection effects across the distribution but Chzhen and Mumford (2011) assign a larger proportion (i.e. about one half) of the total selection effect to unobservables. As described above, selectivity corrections based on Buchinsky (1998) have been called into question by Huber and Melly (2015).

Picchio and Mussida (2011) propose a method based on estimating hazard functions to model the impact of covariates on conditional distributions. Applying methods developed to account for dynamic selection, the study corrects for selection on unobservables in the gender gap using panel data but without the need for exclusion restrictions. They apply their method to studying the gender wage gap in Italy, also finding evidence for positive unobserved selection. It remains an open question to what extent modelling assumptions typically used in hazard rate analysis are plausible for modelling wage distributions. Popli

(2013) extends previous contributions by allowing for multiple outcomes in the selection equation. In order to correct for selection into different types of employment, she includes Heckman-type correction terms into her analysis. Her empirical analysis examines the gender wage gap in Mexico accounting for selection into both the formal and informal sector based on a multinomial logit model.

In a recent important paper, Arellano and Bonhomme (2017) suggest a copula based method to provide consistent estimates of quantile regressions with selection correction. They estimate quantile regressions while assuming a fixed copula between the conditional rank in the wage distribution and the rank in the error term of the selection equation. The approach amounts to estimating rotated quantile regressions, which relate the $\theta^{th}$ quantile regression in the nonselected sample to a rotated value, which represents the rank of the $\theta^{th}$ unselected quantile in the selected sample, thus linking the two for estimation purposes. This is an alternative to Buchinsky's selection correction approach which estimates the difference between the $\theta^{th}$ quantiles in the two samples. The approach of Arellano and Bonhomme has two disadvantages. First, the authors estimate the copula while assuming a specific functional form, and they allow only for the covariates to have a limited impact on the joint distribution of ranks. Second, the estimation of the copula is computationally very involved. We view the two approaches to model the relationship between the ranks or the quantiles, respectively, in the unselected and selected sample as complementary, both having to address the dependence of this relationship upon covariates. Maasoumi and Wang (2019) use a variant of this method in a decision-theoretic analysis of the evolution of the U.S. gender wage gap across time. Another very general but practically demanding method that accounts for unobserved selection in distributional analysis has recently been proposed by Fernandez-Val et al. (2018).

Following a different approach, D'Haultfoeuille et al. (2018) suggest a method to correct for selection on unobservables based on the observation of extreme values that also does not require exclusion restrictions. Specifically aimed at selection on unobservables in the context of the gender gap, Machado (2017) develops a technique for correcting for unobserved selection into employment. Her setup slightly differs from other setups considered so far by focussing on the selection on unobservables of the subpopulation of the 'always employed' women who are similar to the full-time male labor force. Finally, Töpfer (2017) suggests to extend the popular RIF decomposition technique (Firpo et al., 2018) to include terms for unobserved sample selection and applies this method to study the gender wage gap in Italy.

The previous literature on the effects of unobserved selection on the gender wage gap in Germany is relatively sparse. Beblo et al. (2003) applied basic correction techniques including Heckman-type corrections. They found partly contradictory results depending on the chosen specification. In an early contribution, Fitzenberger and Wunderlich (2004) also employed the Buchinsky (1998, 2001) control function approach but obtained inconclusive results on the impact of unobserved selection.

## 4.3 Data and descriptive statistics

Our empirical analysis builds on the German Socio-Economic Panel (GSOEP), a representative panel survey of German households that has been carried out on annual basis since the year 1984.[3] Despite some limitations such as a moderate sample size, the GSOEP is ideal for our purpose as it comprises a large number of individual and household characteristics including potential instruments for modeling female labor market participation. For our analysis, we pool two years 2012 and 2013 in order to make our results less year-specific and in order to increase statistical precision.[4] We include all subsamples of the GSOEP, except the so-called high-income sample oversampling individuals with very high incomes, which

[3]This paper uses the German Socio-Economic Panel (GSOEP), years 1984– 2013 (DOI: 10.5684/soep.v30).

[4]See, e.g., Hyslop and Mare (2005) or Blundell et al. (2007). In what follows, we fully take into account our longitudinal data structure for all procedures of statistical inference by re-sampling from the set of individuals rather than from the set of yearly observations.

we exclude for robustness reasons. For our final sample, we apply the following sample selection criteria. First, we only consider individuals aged between 20 and 60 years. Second, we exclude the self-employed, individuals in vocational training, civil servants and retirees.

The dependent variable of our analysis is the hourly wage computed as the current monthly gross labor income divided by the monthly working time (including overtime). For reasons of plausibility, we omit individuals with an hourly wage below 4 Euros and respondents who reported a monthly working time of more than 350 hours. We also exclude observations with missing values in any of our covariates as listed below. Our sample consists of 5,697 women and 4,769 men who provided a total of 9,596 female observations (3,632 in full-time employment and 5,964 not in full-time employment) and 8,060 male observations (6,627 in full-time and 1,433 not in full-time employment). Table 4.1 presents our variable list along with descriptive statistics for our estimation sample. In our quantile wage regressions, we include a set of standard covariates including dummies for educational qualifications (3 categories: (i) no degree, (ii) high school degree (*Abitur* and/or vocational training, (iii) university or university of applied science (*Fachhochschule*), dummies for German nationality, urban areas and marital status, as well as a quadratic polynomial in work experience (full-time or part-time work experience in years, where one year of part-time experience counts as .5 years of full-time experience).

**Table 4.1** – Descriptive statistics

| | *Men* | | *Women* | | | |
|---|---|---|---|---|---|---|
| | | | Full-time | | Not full-time | |
| Variable | Mean | Sd | Mean | Sd | Mean | Sd |
| Hourly Wage (Euros) | 17.69 | 9.62 | 15.12 | 7.42 | 13.29 | 6.27 |
| Abitur or voc. training | 0.669 | 0.471 | 0.628 | 0.483 | 0.655 | 0.475 |
| University or Fachhochschule | 0.247 | 0.431 | 0.302 | 0.459 | 0.173 | 0.378 |
| German | 0.928 | 0.259 | 0.945 | 0.228 | 0.873 | 0.333 |
| Urban | 0.676 | 0.468 | 0.688 | 0.464 | 0.648 | 0.478 |
| Married | 0.578 | 0.494 | 0.369 | 0.483 | 0.665 | 0.472 |

| | | | | | | |
|---|---|---|---|---|---|---|
| East | 0.169 | 0.375 | 0.188 | 0.391 | 0.163 | 0.369 |
| Work experience | 19.68 | 10.94 | 16.47 | 10.91 | 11.39 | 8.47 |
| Work experience (squared) | 507.27 | 437.84 | 390.44 | 420.45 | 201.55 | 262.49 |
| Age | 42.91 | 10.36 | 41.28 | 11.13 | 43.08 | 10.62 |
| Age (squared) | 1948.74 | 863.32 | 1828.12 | 915.47 | 1968.99 | 890.57 |
| Number of childen 0-2 years | 0.067 | 0.260 | 0.016 | 0.124 | 0.108 | 0.328 |
| Number of childen 3-5 years | 0.089 | 0.310 | 0.028 | 0.169 | 0.140 | 0.373 |
| Number of childen 6-11 years | 0.183 | 0.479 | 0.070 | 0.288 | 0.293 | 0.578 |
| Number of childen 12-18 years | 0.230 | 0.555 | 0.135 | 0.402 | 0.327 | 0.629 |
| Religious | 0.117 | 0.321 | 0.098 | 0.297 | 0.200 | 0.400 |
| Unemployment rate | 7.61 | 2.78 | 7.78 | 2.88 | 7.61 | 2.71 |
| Female labor force participation rate | 72.20 | 3.17 | 72.40 | 3.15 | 72.07 | 3.16 |
| Income of partner (thousand Euros) | 0.848 | 1.270 | 1.202 | 2.009 | 1.835 | 2.419 |
| Household asset income (thousand Euros) | 1.138 | 4.115 | 1.184 | 6.858 | 1.078 | 4.686 |
| Care | 0.012 | 0.110 | 0.012 | 0.109 | 0.027 | 0.162 |
| Observations | 6,627 | | 3,632 | | 5,964 | |
| Individuals | 3,842 | | 2,149 | | 3,548 | |

Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.

In our employment selection model, we include all of these variables plus a set of additional variables assumed to influence participation probabilities but not wage outcomes. Our choice of these variables is similar to that in the literature, see, e.g., Albrecht et al. (2009) or Chzhen and Mumford (2011). In particular, we include a quadratic polynomial in age, dummies for the number of children in different age groups (0-2, 3-5, 5-11, 12-18), a dummy for being religious, the income of the partner, the household's asset income as well as a dummy indicating whether there is a person living in the same household requiring care. In addition, we include two variables representing aggregate labor market information at the level of the federal states which reflect general macroeconomic trends and business cycle effects (the regional unemployment rate and the regional female labor force participation rate, see Bundesagentur für Arbeit, 2015, table 2.1, and Statistisches Bundesamt, 2014, table 5.4). As table 4.1 shows, women working full-time are more likely to hold a university degree and have significantly more work experience. They also have fewer children, are less likely to be married, less religious, have partners with a lower income and are less likely to have family members requiring care.

## 4.4 Econometric model

### 4.4.1 Quantile regressions with selection correction and counterfactual wage distributions (Albrecht et al., 2009)

We start with a description of the approach using the Machado-Mata (2005) decomposition approach based on selection corrected quantile regression as suggested by Albrecht et al. (2009). The method augments the original decomposition technique by correcting for selection on unobservables using the Buchinsky (1998, 2001) quantile selection correction. For our application, consider two groups (A) and (FT). The first group (A) represents all women regardless of whether they actually work full-time or not. By contrast, the second group (FT) includes only the women who in fact work full-time and therefore provide observations of full-time wages. The description of conditional wage quantiles in the total population (A) is given by

$$Q_\theta\left(y_A|x_A\right) = x_A'\beta^A\left(\theta\right) \tag{4.1}$$

where $Q_\theta\left(y_A|x_A\right), \theta \in (0,1)$ denotes the $\theta$-quantile of the distribution of potential (log) wages $y_A$ for women with characteristics $x_A$. This is the distribution of wages a randomly drawn woman with characteristics $x_A$ *would* face if she decided to work full-time. The vector $\beta^A\left(\theta\right)$ is the vector of true, quantile-specific coefficients for this distribution of potential wages.

However, as not all women eventually choose to work full-time, estimation of $\beta^A\left(\theta\right)$ can only proceed using the wages of women who are actually observed to work-full time. Conditional quantiles in this non-randomly selected subpopulation may be higher or lower depending on which women selected themselves into full-time work. As shown by Buchinsky (1998), this implies the existence of sample selection correction terms $h_\theta\left(z_{FT}'\gamma\right)$ at each quantile $\theta$

so that the conditional quantiles in the population of women who selected themselves into full-time work can be written as

$$Q_\theta\left(y_{FT}|z_{FT}, D=1\right) = x'_{FT}\beta^A\left(\theta\right) + h_\theta\left(z'_{FT}\gamma\right), \tag{4.2}$$

where $z$ is a vector of individual characteristics explaining participation in full-time work according to the model

$$Pr\left(D=1|z\right) = Pr\left(z'\gamma + \epsilon > 0|z\right). \tag{4.3}$$

The vector $z$ is assumed to contain all the characteristics $x$ that determine wages plus additional characteristics that influence selection without influencing wages. In an analogy to the Heckman (1979) selection model, Buchinsky (1998) approximates the selection correction term $h_\theta\left(z'_{FT}\gamma\right)$ as a function of the inverse Mills ratio using a power series of the form

$$\hat{h}_\theta\left(z'_{FT}\gamma\right) = \delta_0\left(\theta\right) + \delta_1\left(\theta\right)\lambda\left(z'_{FT}\gamma\right) + \delta_2\left(\theta\right)\lambda\left(z'_{FT}\gamma\right)^2 + \delta_3\left(\theta\right)\lambda\left(z'_{FT}\gamma\right)^3 ... \tag{4.4}$$

with $\lambda(\cdot)$ denoting the inverse Mills ratio. The parameters $\gamma$ can be estimated using a binomial choice model such as probit or more general models.[5]

To solve the problem that the intercepts in the wage and the selection model (i.e. $\beta_0^A\left(\theta\right)$ and $\delta_0\left(\theta\right)$) are not separately identified, Buchinsky (1998) and Albrecht et al. (2009) propose to use identification-at- infinity arguments, implying that $\beta_0^A(\theta)$ can be consistently estimated in a subsample of women whose probability of working full-time is close to one without correcting for selection.[6]

[5]In our empirical application, we initially experimented with the semi-parametric estimators developed by Gallant and Nycka (1987) and Klein and Spady (1993). A disadvantage of these estimators is that they exhibit occasional convergence problems leading to difficulties when computing bootstrap standard errors. We found only minor differences between the values for $\lambda\left(z'_{FT}\hat{\gamma}\right)$ resulting from a probit model and those from the more general models so that we eventually decided to work with the probit model.

[6]In our empirical application, we treat women with a participation probability of over 75 percent as being close enough to sure participation. Our results are robust to increasing this threshold, but at the cost of higher standard errors.

Following Machado and Mata (2005), a number of different counterfactual wage distributions can be estimated by repeatedly sampling values $\theta$ from the interval $(0,1)$ and values $x$ from an empirical distribution of characteristics. For example, if one repeatedly samples $\theta$ from the interval $(0,1)$ and $x_A$ from the population of all women, one can simulate values

$$\hat{y}_A = x'_A \hat{\beta}^A(\theta) \tag{4.5}$$

whose distribution corresponds to the distribution of womens' wages in the counterfactual scenario where all women work full-time.

Comparing this resulting counterfactual distribution, denoted by $F(\hat{y}_A|x_A, \hat{\beta}^A)$, where $\hat{\beta}^A$ involves the family of quantile regression coefficients for $\theta \in (0,1)$, with the observed wage distribution of full-time women $F(y_{FT})$ yields the total effect of selection into full-time employment

$$F(y_{FT}) - F(\hat{y}_A|x_A, \hat{\beta}^A). \tag{4.6}$$

According to Albrecht et al. (2009), this effect can be split up into a part due to differences in observable characteristics and a part due to differences in unobservable characteristics. The first part is given by the difference between the observed wages of full-time women $F(y_{FT})$ and the counterfactual that would prevail if all women worked full-time and received the returns of full-time women without accounting for selection $F(\hat{y}_A|x_A, \hat{\beta}^{FT})$ , i.e.

$$F(y_{FT}) - F(\hat{y}_A|x_A, \hat{\beta}^{FT}), \tag{4.7}$$

where $\hat{\beta}^{FT}$ are the coefficients for full-time working women without selection correction. The effect of selection due to unobservables is given by comparing the counterfactual scenarios where all women work full-time and receive the returns of full-time women without accounting for selection $F(\hat{y}_A|x'_A\hat{\beta}^{FT})$ with the counterfactual scenario used in the calculation of the total selection effect, i.e. assuming all women work full-time and receive the selection adjusted returns of full-time women $F(\hat{y}_A|x_A, \hat{\beta}^A)$. This yields the difference

$$F(\hat{y}_A|x_A, \hat{\beta}^{FT}) - F(\hat{y}_A|x_A, \hat{\beta}^A). \tag{4.8}$$

### 4.4.2 Huber and Melly (2015) conditional independence test

As described in Huber and Melly (2015), the Buchinsky (1998) selection correction approach assumes a potential outcome model of the form

$$y^* = x'\beta + u, \tag{4.9}$$

where $y^*$ denotes the log wage individuals with $x$ and $u$ *would* earn if they decided to work full-time. In practice, one only observes log wages $y = y^*$ of individuals who in fact chose to work full-time. Selection into full-time status is determined by

$$D = \mathbf{1}\left[z^{'}\gamma + \epsilon \geq 0\right] \tag{4.10}$$

with $z$ denoting a strict superset of $x$ as outlined in the previous section. Huber and Melly (2015) pointed out that Buchinsky's (1998) control function approach assumes conditional independence in the sense that the joint density $f_{u,\epsilon}$ of the error terms of the potential outcome model $u$ and the selection model $\epsilon$ is independent of $z$, conditional on the probability of selection $Pr(D = 1|z)$, i.e.

$$f_{u,\epsilon}(.|z) = f_{u,\epsilon}(.|Pr\,(D = 1|z)). \tag{4.11}$$

This follows from assumptions C and E in Buchinsky (1998) and is equivalent to $f_{u,\epsilon}(.|z) = f_{u,\epsilon}(.|z^{'}\gamma)$ as $\epsilon$ is assumed to be independent of $z$.

As a consequence,

$$
\begin{aligned}
Q_\theta(y|z, D=1) &= x'\beta + Q_\theta(u|z, D=1) \\
&= x'\beta + Q_\theta(u|z, z'\gamma \geq -\epsilon) \\
&= x'\beta + Q_\theta(u|z'\gamma, z'\gamma \geq -\epsilon) \\
&= x'\beta + Q_\theta(u|z'\gamma, D=1) \\
&= x'\beta + h_\theta(z'\gamma),
\end{aligned} \tag{4.12}
$$

where the third line follows from conditional independence. Huber and Melly (2015) point out that this holds for any quantile $\theta$, implying that, net of the selection terms $h_\theta(z'\gamma)$, all quantile regression models in the selected sample have to be parallel under the conditional independence assumption. In substantive terms, the conditional independence assumption (4.11) rules out that the distribution of $u$ directly depends on $z$ (i.e. is heteroscedastic in $z$), although it may depend on the selection probability $P(D=1|z)$ (i.e. it may be heteroscedastic in the index $z'\gamma$).

To test whether this sort of conditional independence is fulfilled in our application, we implement the Huber/Melly test, which tests the equality of quantile regression coefficients in the selected sample. The null hypothesis of the test is

$$H_0 : \beta(\theta) = \beta(0.5) \quad \text{for all} \quad \theta \in \Theta \tag{4.13}$$

against

$$H_1 : \beta(\theta) \neq \beta(0.5) \quad \text{for some} \quad \theta \in \Theta. \tag{4.14}$$

Following Huber and Melly (2015), our implementation tests for departures from the $H_0$ in a fine grid with step length 0.01 in the regions $\Theta_1$=[0.05,0.95], $\Theta_2$=[0.1,0.9], $\Theta_3$=[0.15,0.85] and $\Theta_4$=[0.2,0.8]. Departures from the $H_0$ are measured using the Kolmogorov-Smirnov

(KS) and Cramér-von Mises (CM) test statistics, i.e.

$$T_n^{KS} = sup_{\theta\in\Theta}\sqrt{n}\left\|\hat{\beta}\left(\theta\right) - \hat{\beta}\left(0.5\right)\right\|_{\hat{\Lambda}_\theta} \tag{4.15}$$

$$T_n^{CM} = n\int_{\Theta}\left\|\hat{\beta}\left(\theta\right) - \hat{\beta}\left(0.5\right)\right\|_{\hat{\Lambda}_\theta}^2 d\theta \tag{4.16}$$

with $\|a\|_{\hat{\Lambda}_\theta}$ defined by $\sqrt{a'\hat{\Lambda}_\theta a}$ and $\hat{\Lambda}_\theta$ denoting a positive weighting matrix (as described in Huber and Melly, 2015). The unknown distributions of the test statistics are obtained by resampling score functions, which are linear approximations of the empirical process $\sqrt{n}\left(\hat{\beta}\left(\theta\right) - \hat{\beta}\left(0.5\right)\right)$.[7]

### 4.4.3 Model transformation

In order to avoid the assumption that the data generating process directly meets the conditional independence assumption, we suggest to transform the original model so that the transformed model meets this assumption.[8] Consider a transformation by dividing the original model (4.9) by $h(z) > 0$ (a positive, possibly nonlinear function of the covariates $z$)

$$\left[\frac{y^*}{h(z)}\right] = \left[\frac{x'}{h(z)}\right]\beta + \left[\frac{u}{h(z)}\right] \quad \Leftrightarrow \quad \widetilde{y}^* = \widetilde{x}'\beta + \widetilde{u}. \tag{4.17}$$

[7]See Huber and Melly (2015) for more details. Our Stata implementation of the test is available upon request.

[8]The idea to transform the dependent variable is similar to the companion paper Fitzenberger and de Lazzer (2019), which estimates the selection bias in employment for male workers to account for unemployment. However, there are two key methodological differences. First, Fitzenberger and de Lazzer (2019) only transform the dependent variable while leaving the covariates unchanged, while our paper transforms both the dependent variable and the covariates. Second, the approach taken by Fitzenberger and de Lazzer (2019) to determine the transformation factor relies on the identification-at-infinity approach, which is plausible in a setting where the median selection probability lies between 94% and 97%. As described below, the transformation in the present paper is motivated by a location-scale model using an arguably plausible assumption about the dispersion in the selective sample and the selection probability.

If the conditional independence assumption holds in this model (i.e. if the distribution of $\widetilde{u}$ is independent of $z$ but possibly dependent on $z'\gamma$), then one can apply the Buchinsky (1998) selection correction to this model. But the validity of the Buchinsky representation (4.12) is directly tested by the Huber/Melly test (i.e. constant quantile regression slope coefficients and correction terms that may depend on $z'\gamma$ but not on $z$). As a consequence, the Buchinsky (1998) selection correction may be applied to any transformed model that passes the Huber/Melly test. For the Albrecht et al. (2009) method, it is then necessary to produce a prediction of $Q_\theta(y^*|z)$ referring to the untransformed values $y^*$ and $z$. But such a prediction is easy to produce because

$$Q_\theta(y^*|z) = Q_\theta(h(z)\widetilde{y}^*|z) = h(z)Q_\theta(\widetilde{y}^*|\widetilde{x}), \tag{4.18}$$

i.e. one reverses the transformation after computing predictions in the transformed model.[9]

Next, we give an example of a realistic data generating process and a transformation so that the transformed model satisfies the conditional independence assumption, whereas the untransformed model does not. Note however, that the argument in the previous paragraph is completely general and independent of the upcoming example. If one succeeds in monotonically transforming the original model so that the transformed model passes the Huber/Melly test, then one may apply the Buchinsky (1998) selection correction no matter how the original data generating process looks like.

The example data generating process is a standard regression model with multiplicative heteroscedasticity (similar but more general than the one discussed in Huber and Melly, 2015), i.e.

$$y^* = x'\beta + g(x)\,v, \tag{4.19}$$

[9] This works because, given $z$, dividing by $h(z) > 0$ is a monotonic transformation so that the transformed dependent and explanatory variables contain exactly the same information, and because the quantiles of a monotonically transformed variable are identical to the transformed quantiles of the untransformed variable. In principle, one might therefore also consider more complicated monotonic transformations of the original model than dividing by a function $h(z) > 0$, as long as these only depend on $z$.

where $g(x) > 0$ and $v$ is assumed to satisfy the conditional independence assumption (i.e. its distribution does not depend on $z$ but may depend on $z'\gamma$). The model represents a realistic data generating process because it includes the two features that are most interesting to economists: location shifts depending on personal characteristics on the one hand (via $\beta$), and heteroscedasticity depending on $x$ on the other (via $g(x)$).

Using a similar idea as in Chen and Khan (2003), in this model for two quantiles $\alpha_2 > \alpha_1$

$$Q_{\alpha_2}(y|z, D=1) = x'\beta + g(x)Q_{\alpha_2}(v|z'\gamma, D=1) \tag{4.20}$$

$$Q_{\alpha_1}(y|z, D=1) = x'\beta + g(x)Q_{\alpha_1}(v|z'\gamma, D=1) \tag{4.21}$$

(the quantiles of $v$ do not depend on $z$ but only on $z'\gamma$ by the conditional independence assumption for $v$). This implies that

$$g(x) = \frac{Q_{\alpha_2}(y|z, D=1) - Q_{\alpha_1}(y|z, D=1)}{Q_{\alpha_2}(v|z'\gamma, D=1) - Q_{\alpha_1}(v|z'\gamma, D=1)} = \frac{\Delta q(z)}{\Delta v(z'\gamma)}. \tag{4.22}$$

In this expression, $\Delta q(z)$ can be easily estimated as it is an interquantile spread for individuals with characteristics $z$ in the selected population. The other quantity $\Delta v(z'\gamma)$ is in principle unobservable. However, a natural assumption would be that in the generating model (4.19), $v$ represents a standardized disturbance whose quantile spreads do not depend on $z'\gamma$, i.e. $\Delta v(z'\gamma) = \Delta v$. A natural choice would be a standard normally distributed variable $v$ for which one can compute interquantile spreads $\Delta v = \Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)$. This does not restrict the generality of the model in any important way because the variance of the conditional distributions $y^*|x$ may still dependent very generally on $x$ through the heteroscedastic scaling factor $g(x)$.

In this case, transforming the data generating process by $h(z) = \Delta q(z)/\Delta v$ yields the transformed model

$$\begin{bmatrix} \frac{y^*}{\frac{\Delta q(z)}{\Delta v}} \end{bmatrix} = \begin{bmatrix} \frac{x'}{\frac{\Delta q(z)}{\Delta v}} \end{bmatrix} \beta + g(x) \begin{bmatrix} \frac{v}{\frac{\Delta q(z)}{\Delta v}} \end{bmatrix} = \begin{bmatrix} \frac{x'}{\frac{\Delta q(z)}{\Delta v}} \end{bmatrix} \beta + v \tag{4.23}$$

in which the conditional independence assumption holds (by assumptions on $v$).

Note that it is not strictly necessary to assume that the interquantile spreads of the disturbance term $\Delta v(z'\gamma)$ are independent of $z'\gamma$ for obtaining a transformed model in which the conditional independence assumption holds. For example, assume that the interquantile spread takes the form

$$\Delta v(z'\gamma) = \Delta v^* \cdot c(\alpha_2, \alpha_1, z'\gamma), \tag{4.24}$$

i.e. the interquantile spread of $v$ may be widened or compressed depending on the selection index $z'\gamma$. In this case, one can transform the model by $h(z) = \Delta q(z)/\Delta v^*$ yielding

$$\begin{bmatrix} \frac{y^*}{\frac{\Delta q(z)}{\Delta v^*}} \end{bmatrix} = \begin{bmatrix} \frac{x'}{\frac{\Delta q(z)}{\Delta v^*}} \end{bmatrix} \beta + g(x) \begin{bmatrix} \frac{v}{\frac{\Delta q(z)}{\Delta v^*}} \end{bmatrix} = \begin{bmatrix} \frac{x'}{\frac{\Delta q(z)}{\Delta v^*}} \end{bmatrix} \beta + \begin{bmatrix} \frac{v}{c(\alpha_2, \alpha_1, z'\gamma)} \end{bmatrix}. \tag{4.25}$$

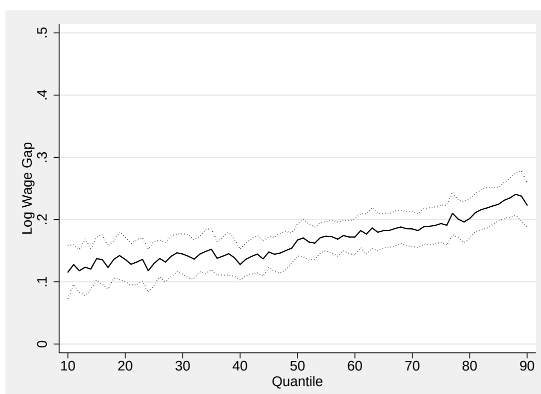
The conditional independence assumption still holds in this model because the transformed error term depends only on $z'\gamma$ and not on $z$. Assumption (4.24) is quite natural and plausible for the location-scale model in eq. (4.19), because the distribution of the scale independent error term $v$ only depends upon $z'\gamma$ and the error term of the selection equation is assumed to be independent of $z$.

In our empirical application below, we use $\Delta v^* = \Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)$ along with $\alpha_2 = .85$ and $\alpha_1 = .15$. With these choices, the Huber/Melly test passes for the transformed model, whereas it clearly rejects in the untransformed model (see below). Again, note that according to our argument above, it suffices to find one transformation for which the Huber/Melly test passes in order to be able to apply the Buchinsky (1998) selection correction and therefore the Albrecht et al. (2009) framework.

## 4.5 Application to the gender wage gap in Germany

In the following, we consider the impact of female selection into full-time employment for the gender wage gap. As in Albrecht et al. (2009), we make the simplifying assumption that there is no selection problem for men, which appears to be a reasonable assumption given the high male participation rates in full-time work. For women, we define the selected group as those who work full-time, while the non-selected group is defined as those who do not full-time (i.e. those who do not work at all or who work only part-time).

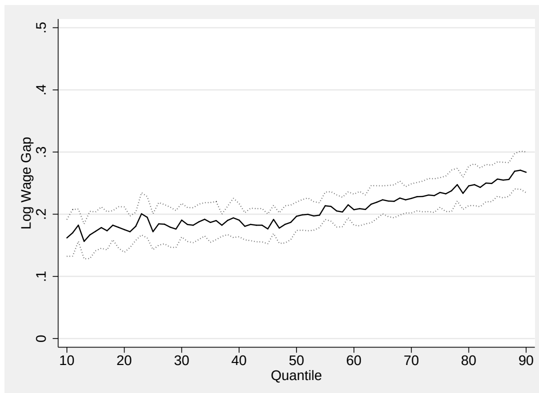
**Figure 4.1** – Raw wage gap, full-time men vs. full-time women




Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Confidence bands based on bootstrap (100 replications).

Figure 4.1 shows the raw wage gap between full-time men and full-time women. The gap is positive, ranging from around 15 percent in the lower half of the distribution, and rising to around 25 percent in the upper half. It is well-known that this gap overestimates differences in wages paid between men and women with the same characteristics if the women who work full-time have less favorable observed characteristics than men who work full-time. On the other hand, it may underestimate pay differences between men and women with the same characteristics, if the women who work full-time are a particularly positive selection of all women in terms of their unobservable characteristics.

**Figure 4.2** – Raw wage gap, full-time men vs. full-time/part-time women




Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Confidence bands based on bootstrap (100 replications).

A step into the direction of asking how the wage gap would look like if female participation in full-time work was less selective is done in figure 4.2. This figure shows the gender wage gap including not only the women who worked full-time but also those who worked part-time along with their part-time wages. As expected, this increases the observed raw wage gap. On the one hand, this may be due to the fact that the women who worked full-time were a particularly positively selected group of all working women.[10] On the other hand, it may also be due to the fact that part-time wages are generally lower than full-time wages.

In order to obtain a proper answer to the question of how the gender pay gap would look like if not only the potentially very positively selected group of full-time women are paid full-time wages, we follow the Albrecht et al. (2009) and compute a number of relevant counterfactual distributions accounting for nonrandom selection into observed full-time work.

[10]See Biewen et al. (2018) for a related analysis of the selectivity of full-time vs. part-time women.

### 4.5.1 Selection model and quantile regressions

**Table 4.2** – Female participation in full-time work

| | Probit coefficients | Marginal effects |
|---|---|---|
| Abitur or voc. training | 0.286 | 0.072 |
| | (0.054) | (0.014) |
| University or Fachhochschule | 0.999 | 0.251 |
| | (0.061) | (0.015) |
| German | 0.018 | 0.005 |
| | (0.072) | (0.018) |
| Urban | 0.064 | 0.016 |
| | (0.036) | (0.009) |
| Married | -0.270 | -0.068 |
| | (0.038) | (0.009) |
| East | 0.061 | 0.015 |
| | (0.084) | (0.021) |
| Work experience | 0.149 | 0.037 |
| | (0.009) | (0.002) |
| Work experience (squared) | -0.001 | -0.000 |
| | (0.000) | (0.000) |
| Age | -0.007 | -0.002 |
| | (0.017) | (0.004) |
| Age (squared) | -0.001 | -0.000 |
| | (0.000) | (0.000) |
| Number of childen 0-2 years | -1.397 | -0.350 |
| | (0.080) | (0.019) |
| Number of childen 3-5 years | -0.736 | -0.184 |
| | (0.060) | (0.015) |
| Number of childen 6-11 years | -0.457 | -0.115 |
| | (0.038) | (0.009) |
| Number of childen 12-18 years | -0.200 | -0.050 |
| | (0.030) | (0.008) |
| Religious | -0.199 | -0.050 |
| | (0.046) | (0.011) |
| Unemployment rate | -0.012 | -0.003 |
| | (0.008) | (0.002) |
| Female labor force participation rate | 0.008 | 0.002 |
| | (0.009) | (0.002) |
| Income of partner | -0.042 | -0.010 |
| | (0.008) | (0.002) |
| Household asset income | -0.000 | -0.000 |
| | (0.002) | (0.001) |
| Care | -0.327 | -0.082 |
| | (0.123) | (0.031) |
| Constant | -0.061 | - |
| | (0.754) | (-) |
| Observations | 9,596 | 9,596 |

Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Standard errors in parentheses.

Table 4.2 summarizes the results of our selection model for female participation in full-time work. The left hand column contains the probit coefficients, while the right hand column displays average marginal effects. All effects are in line with a-priori expectations. Participation in full-time work substantially increases with educational qualifications and work experience. With respect to our instrumental variables, we find the usual strong effects of children (the younger the children, the stronger) and the negative effect of partner income on participation in full-time work. Moreover, participation in full-time work is ceteris paribus lower if the person is religious or of older age. The rest of the instrumental variables have the expected signs but are not statistically significant.

Tables C1 and C2 in the appendix summarize some results of our conditional quantile regression models before any kind of transformation is carried out, with and without the Buchinsky (1998) correction. Again, all estimated coefficients are in line with a-priori expectations. In particular, we find strong positive effects of educational qualifications and work experience which are increasing over the distribution. There are also positive wage effects linked to German citizenship, living in a city and a wage penalty associated with living in East Germany.

### 4.5.2 Huber and Melly (2015) test and transformation

In order to test whether our untransformed model satisfies the assumptions necessary for the Buchinsky selection correction, we run the Huber/Melly test. Table 4.3 shows that the null hypothesis of conditional independence is overwhelmingly rejected (all p-values close to zero). We therefore apply the transformation described in section 4.4.3 to our model.[11] We then run the Huber/Melly test on the transformed model. Table 4.3 demonstrates that our transformation drastically reduces problems with the conditional independence assumption

[11] The corresponding estimated quantile regression coefficients of the transformed model are provided in tables C3 and C4 in the appendix.

as detected by the Huber/Melly test. After the transformation, the Huber/Melly test easily passes in almost all of the cases, the only exception being the Kolmogorov-Smirnov version for the extreme quantile range of 5-95 percent. As already pointed out in Huber and Melly (2015, p. 1157), it will be hard to avoid rejection of the null hypothesis in the tails of the distribution where the quantile regression estimator performs poorly if one does not make strong assumptions on the tail behavior of the underlying density (particularly in the case of the Kolmogorov-Smirnov statistic which picks out the supremum deviation).[12]

**Table 4.3** – Huber-Melly (2015) test (p-values based on 1,000 resamples)

| $\theta \in$ | $[0.05, 0.95]$ | $[0.10, 0.90]$ | $[0.15, 0.85]$ | $[0.20, 0.80]$ |
|---|---|---|---|---|
| | *Original model* | | | |
| Kolmogorov-Smirnov | 0.000 | 0.000 | 0.000 | 0.000 |
| Cramér-von Mises | 0.000 | 0.000 | 0.000 | 0.000 |
| | *Transformed model* | | | |
| Kolmogorov-Smirnov | 0.029 | 1.000 | 1.000 | 0.997 |
| Cramér-von Mises | 0.992 | 0.999 | 0.996 | 0.960 |

Source: Own calculations.

### 4.5.3 Selection effect and counterfactual comparisons

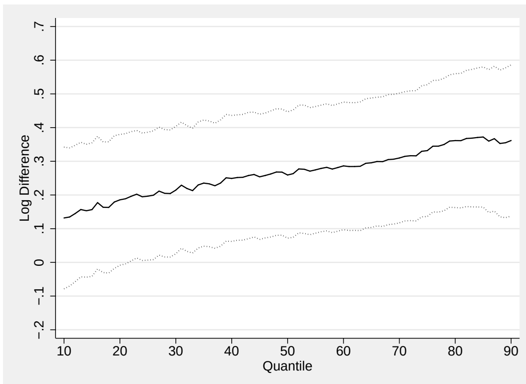
We now present an analysis analogous to Albrecht et al. (2009), but based on the transformed model. For better comparability, we follow exactly the same sequence of scenarios as in Albrecht et al. (2009). We also compare the results of our transformed model to those from our untransformed model.

Figure 4.3 plots the total selection effect across the quantiles of the female wage distribution. The left panel shows the result based on the untransformed model, while the right panel the one based on the transformed model. This total effect of selection is given by the difference

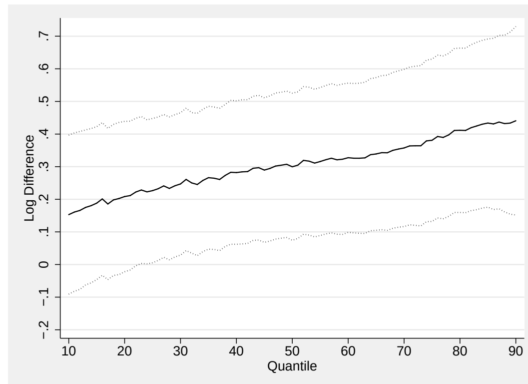
[12] We thoroughly tested our implementation of the Huber/Melly test (details are available on request). In our experience, the test tends to exhibit a 'switching' behavior in practical applications, i.e. p-values are either small or quite large.

**Figure 4.3** – Total selectivity in female population: observed full-time wage distribution minus wage distribution assuming all women work full-time and receive selectivity corrected wage returns

**(a)** Without transformation




**(b)** With transformation

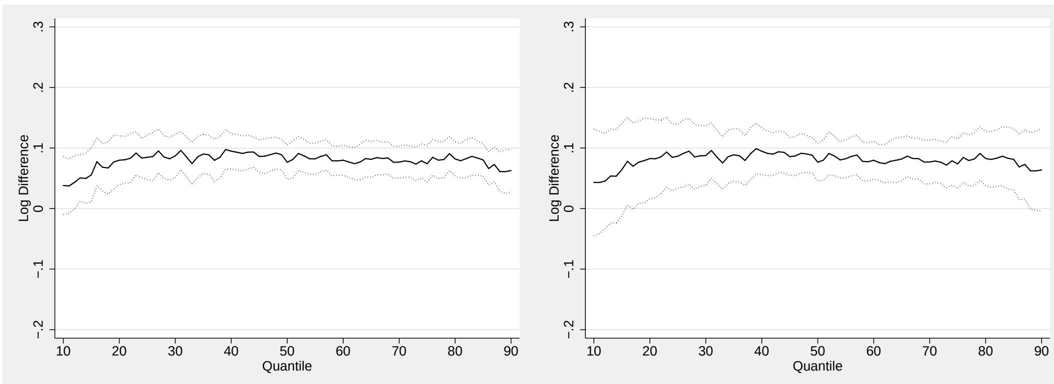



Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Confidence bands based on bootstrap (100 replications).

between the observed wage distribution of full-time women and the counterfactual scenario where all women are assumed to work full-time and receive the selection-adjusted returns of full-time women. In line with a-priori expectations, the wage effects of full-time selectivity are positive in the sense that women actually working full-time have a higher earnings potential in full-time work relative to all women. This effect is stronger in the upper part of the distribution. Comparing the untransformed with the transformed model, we find the estimated selection effect to be somewhat stronger in the transformed model. At the same time, we notice a moderate widening of the bootstrapped confidence intervals in the transformed case, which is due to the fact that the transformation contains estimated quantities. Note however, that the width of the intervals is already quite large in the untransformed case so that uncertainty added by the transformation plays only a minor role.

**Figure 4.4** – Observed selectivity in female population: observed full-time wage distribution minus wage distribution with observed characteristics of all women

**(a)** Without transformation **(b)** With transformation



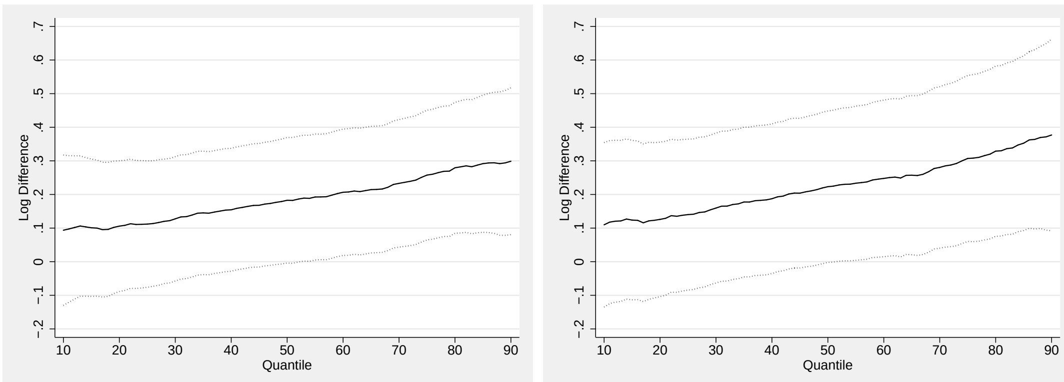

Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Confidence bands based on bootstrap (100 replications).

The next figure 4.4 shows the part of the overall selection effect that can be attributed to differences in observable characteristics between full-time women and their part-time or non-working counterparts. Recall that the observed selection effect is defined by the difference between the observed wages of full-time women and the counterfactual scenario that would prevail if all women worked full-time and received the returns without correcting for selection (equation (4.7)). This result matches the descriptive evidence presented above showing that full-time women are on average more educated and have higher work experience than the women not observed in full-time employment.

Figure 4.5 displays the effect of unobserved selectivity defined by equation (4.8). It represents the difference between the estimates uncorrected for sample selection and the ones corrected for sample selection. It shows how much full-time pay for women is overestimated if one ignores that those women observed to work full-time are positively selected. This effect is rising towards the upper part of the distribution, suggesting that participation in high-wage full-time activities is particularly selective. Still, unobserved selection is also positive at the

**Figure 4.5** – Unobserved selectivity in female population: full-time wage distribution with observed characteristics of all women minus distribution that in addition assumes selectivity corrected returns

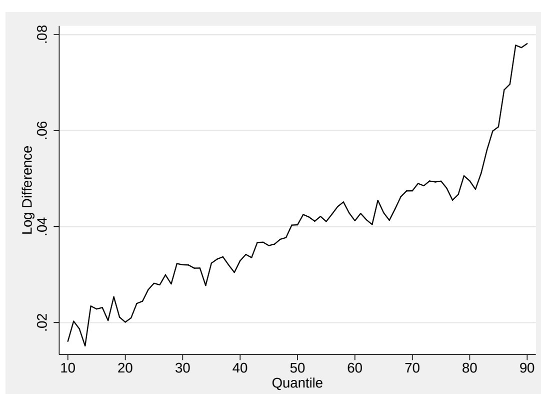
**(a)** Without transformation **(b)** With transformation




Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Confidence bands based on bootstrap (100 replications).

bottom of the distribution, because even in the case of a low earnings potential, those women with more favorable unobserved characteristics find it harder to resist participation in order to pursue child care duties or to depend on other income sources in the household.

**Figure 4.6** – Impact of transformation: difference between unobserved selectivity effect in transformed vs. in untransformed model

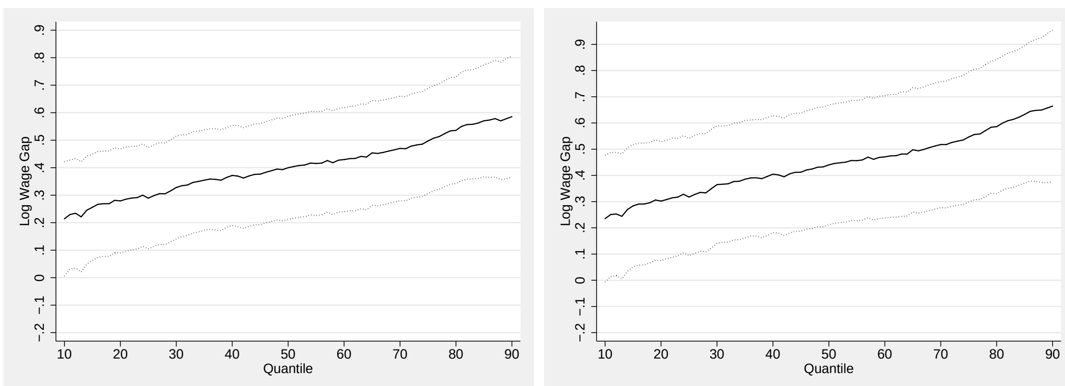



Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.

Although the patterns in figure 4.5 based on the transformed and the untransformed quantile regression model look quite similar, there is a economically significant difference between the two which depicted in a more explicit way in figure 4.6. The figure suggests that the untransformed model underestimates unobserved selectivity effects, particularly towards the upper part of the distribution.

**Figure 4.7** – Selectivity corrected gender wage gap: full-time men vs. selectivity corrected full-time wages women

**(a)** Without transformation **(b)** With transformation



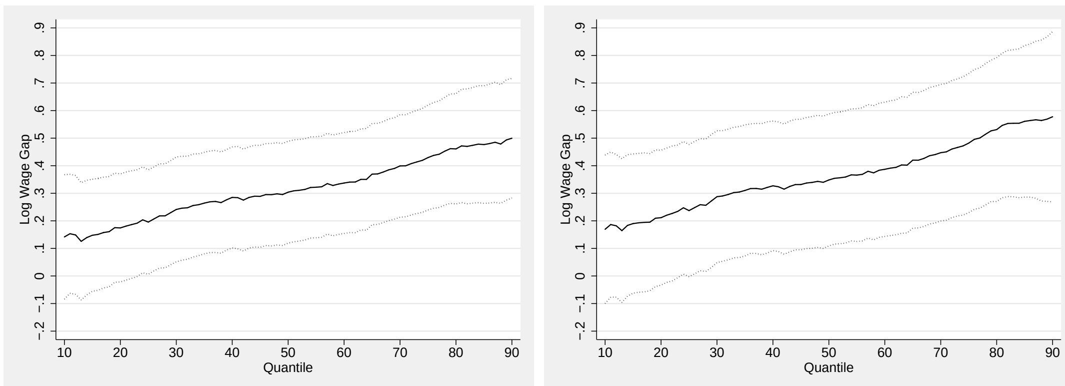

Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Confidence bands based on bootstrap (100 replications).

We now apply these corrections to compute corrected versions of the gender wage gap. Figure 4.7 displays the selection-adjusted wage gap between men and women, i.e. the wage difference that would prevail in a counterfactual scenario in which all women (with their given observed characteristics) work full-time and receive the selection-adjusted returns of full-time women. As a consequence of the positive selection effect described in the previous paragraph, this wage gap considerably widens compared to the raw gap displayed in figures 4.1 and 4.2, and it is higher in the transformed model.

Finally, figure 4.8 shows the wage gap between men and women if both selectivity in to full-time work status is corrected and the fact that women have on average less favorable observed characteristics than men. This corresponds to the counterfactual scenario in which

**Figure 4.8** – Gender wage gap if all women worked full-time and had male characteristics but womens' selectivity corrected returns

**(a)** Without transformation **(b)** With transformation




Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations. Confidence bands based on bootstrap (100 replications).

women earn the selection-adjusted returns of full-time women but have men's distribution of characteristics. Compared to the previous figure 4.7, this narrows the wage gap as women have less favorable observed characteristics than men. Still, the pay difference between men and women remains much bigger than suggested by the raw wage gaps in figure 4.1 and 4.2.

## 4.6 Summary and discussion

This paper proposes a remedy for the problem pointed out by Huber and Melly (2015) that counterfactual decomposition methods accounting for unobserved selectivity are not applicable if the underlying quantile regression model does not satisfy a conditional independence assumption. This conditional independence assumption can be investigated using the Huber and Melly (2015) test. We demonstrate theoretically and empirically for our application that applying a transformation to the original model may make the Huber and Melly (2015) test pass, while it rejects for the untransformed model. The transformation can later be reversed for the computation of quantile predictions. We illustrate this approach by an application to the gender wage gap in Germany. Our substantive results are in line with previous contributions like Albrecht et al. (2009) for the Netherlands and Chzhen and Mumford (2011) for the U.K. in that they show positive unobservable selection of women into full-time work, biasing estimates of the gender pay gap downwards. However, our results also suggest that this downward bias is underestimated if an untransformed model is used rather than a transformed one that passes the Huber and Melly (2015) test.

# Appendix C

— Intentionally left blank —

**Table C1** – Quantile regressions full-time women (model without transformation, without selectivity correction)

| Quantile | Q10 | Q20 | Q30 | Q40 | Q50 | Q60 | Q70 | Q80 | Q90 |
|---|---|---|---|---|---|---|---|---|---|
| Abitur or voc. training | 0.107 | 0.157 | 0.140 | 0.148 | 0.152 | 0.169 | 0.195 | 0.196 | 0.175 |
| | (0.046) | (0.035) | (0.036) | (0.026) | (0.029) | (0.026) | (0.030) | (0.038) | (0.050) |
| University or Fachhochschule | 0.400 | 0.481 | 0.475 | 0.486 | 0.510 | 0.525 | 0.555 | 0.557 | 0.536 |
| | (0.053) | (0.041) | (0.040) | (0.029) | (0.033) | (0.028) | (0.030) | (0.035) | (0.054) |
| German | 0.204 | 0.230 | 0.219 | 0.199 | 0.184 | 0.168 | 0.172 | 0.187 | 0.117 |
| | (0.073) | (0.058) | (0.048) | (0.043) | (0.044) | (0.042) | (0.046) | (0.058) | (0.055) |
| Urban | 0.106 | 0.103 | 0.094 | 0.070 | 0.072 | 0.079 | 0.079 | 0.070 | 0.081 |
| | (0.035) | (0.029) | (0.023) | (0.023) | (0.020) | (0.017) | (0.025) | (0.026) | (0.031) |
| Married | 0.029 | 0.016 | 0.006 | 0.009 | 0.002 | 0.002 | 0.005 | 0.005 | 0.012 |
| | (0.035) | (0.023) | (0.019 | (0.016) | (0.018) | (0.018) | (0.017) | (0.018) | (0.021) |
| East | -0.277 | -0.294 | -0.312 | -0.291 | -0.259 | -0.211 | -0.184 | -0.170 | -0.165 |
| | (0.032) | (0.029) | (0.022) | (0.028) | (0.028) | (0.023) | (0.027) | (0.027) | (0.030) |
| Work experience | 0.019 | 0.020 | 0.021 | 0.024 | 0.026 | 0.026 | 0.026 | 0.028 | 0.032 |
| | (0.004) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) |
| Work experience (squared) | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 | -0.001 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Constant | 1.563 | 1.673 | 1.803 | 1.896 | 1.956 | 2.015 | 2.050 | 2.119 | 2.322 |
| | (0.090) | (0.071) | (0.058) | (0.047) | (0.053) | (0.053) | (0.063) | (0.071) | (0.078) |
| Observations | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 |

Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Bootstrapped standard errors (100 replications) in parentheses.

**Table C2** – Quantile regressions full-time women (model without transformation, with selectivity correction)

| Quantile | Q10 | Q20 | Q30 | Q40 | Q50 | Q60 | Q70 | Q80 | Q90 |
|---|---|---|---|---|---|---|---|---|---|
| Abitur or voc. training | 0.091 | 0.113 | 0.136 | 0.145 | 0.145 | 0.171 | 0.189 | 0.186 | 0.186 |
| | (0.190) | (0.192) | (0.122) | (0.119) | (0.119) | (0.121) | (0.143) | (0.144) | (0.219) |
| University or Fachhochschule | 0.358 | 0.392 | 0.435 | 0.464 | 0.500 | 0.534 | 0.549 | 0.560 | 0.548 |
| | (0.160) | (0.156) | (0.100) | (0.096) | (0.104) | (0.109) | (0.139) | (0.147) | (0.223) |
| German | 0.176 | 0.197 | 0.201 | 0.200 | 0.179 | 0.156 | 0.137 | 0.181 | 0.114 |
| | (0.040) | (0.037) | (0.025) | (0.023) | (0.027) | (0.029) | (0.039) | (0.042) | (0.062) |
| Urban | 0.105 | 0.092 | 0.082 | 0.072 | 0.072 | 0.073 | 0.076 | 0.061 | 0.081 |
| | (0.053) | (0.040) | (0.033) | (0.030) | (0.031) | (0.025) | (0.031) | (0.041) | (0.051) |
| Married | 0.048 | 0.050 | 0.042 | 0.032 | 0.014 | 0.008 | 0.007 | 0.005 | 0.003 |
| | (0.068) | (0.049) | (0.042) | (0.036) | (0.035) | (0.029) | (0.032) | (0.036) | (0.053) |
| East | -0.285 | -0.300 | -0.313 | -0.294 | -0.262 | -0.214 | -0.184 | -0.178 | -0.172 |
| | (0.078) | (0.066) | (0.051) | (0.046) | (0.048) | (0.038) | (0.043) | (0.061) | (0.055) |
| Work experience | 0.018 | 0.021 | 0.021 | 0.023 | 0.025 | 0.026 | 0.026 | 0.028 | 0.034 |
| | (0.035) | (0.023) | (0.023) | (0.023) | (0.021) | (0.018) | (0.023) | (0.026) | (0.031) |
| Work experience (squared) | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 | -0.001 |
| | (0.036) | (0.024) | (0.022) | (0.025) | (0.021) | (0.022) | (0.024) | (0.022) | (0.026) |
| Constant | 1.578 | 1.662 | 1.777 | 1.853 | 1.800 | 1.849 | 1.816 | 1.754 | 1.909 |
| | (0.204) | (0.190) | (0.173) | (0.157) | (0.160) | (0.129) | (0.120) | (0.131) | (0.181) |
| Observations | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 |

Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Bootstrapped standard errors (100 replications) in parentheses.

**Table C3** – Quantile regressions full-time women (transformed model, without selectivity correction)

| Quantile | Q10 | Q20 | Q30 | Q40 | Q50 | Q60 | Q70 | Q80 | Q90 |
|---|---|---|---|---|---|---|---|---|---|
| Abitur or voc. training | -0.093 | -0.063 | -0.084 | -0.104 | -0.096 | -0.092 | -0.076 | -0.085 | -0.133 |
| | (0.231) | (0.233) | (0.239) | (0.238) | (0.245) | (0.248) | (0.248) | (0.242) | (0.256) |
| University or Fachhochschule | 0.034 | 0.084 | 0.060 | 0.033 | 0.051 | 0.051 | 0.068 | 0.055 | -0.015 |
| | (0.262) | (0.268) | (0.272) | (0.271) | (0.280) | (0.279) | (0.283) | (0.279) | (0.286) |
| German | 0.334 | 0.395 | 0.396 | 0.389 | 0.380 | 0.368 | 0.386 | 0.400 | 0.349 |
| | (0.287) | (0.288) | (0.292) | (0.296) | (0.304) | (0.307) | (0.308) | (0.305) | (0.325) |
| Urban | 0.232 | 0.229 | 0.235 | 0.214 | 0.223 | 0.232 | 0.235 | 0.231 | 0.260 |
| | (0.111) | (0.114) | (0.110) | (0.111) | (0.114) | (0.115) | (0.116) | (0.114) | (0.123) |
| Married | 0.039 | 0.029 | 0.025 | 0.029 | 0.026 | 0.025 | 0.026 | 0.029 | 0.041 |
| | (0.119) | (0.115) | (0.110) | (0.112) | (0.114) | (0.116) | (0.116) | (0.117) | (0.122) |
| East | -0.663 | -0.697 | -0.740 | -0.746 | -0.729 | -0.698 | -0.678 | -0.680 | -0.718 |
| | (0.153) | (0.154) | (0.154) | (0.156) | (0.159) | (0.163) | (0.158) | (0.162) | (0.172) |
| Work experience | -0.010 | -0.011 | -0.011 | -0.011 | -0.011 | -0.012 | -0.012 | -0.012 | -0.012 |
| | (0.018) | (0.017) | (0.017) | (0.017) | (0.018) | (0.018) | (0.018) | (0.018) | (0.019) |
| Work experience (squared) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Constant | 5.882 | 6.257 | 6.667 | 7.083 | 7.275 | 7.525 | 7.637 | 7.917 | 8.667 |
| | (1.442) | (1.451) | (1.476) | (1.484) | (1.544) | (1.569) | (1.586) | (1.552) | (1.664) |
| Observations | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 |

Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Bootstrapped standard errors (100 replications) in parentheses.

**Table C4** – Quantile regressions full-time women (transformed model, with selectivity correction)

| Quantile | Q10 | Q20 | Q30 | Q40 | Q50 | Q60 | Q70 | Q80 | Q90 |
|---|---|---|---|---|---|---|---|---|---|
| Abitur or voc. training | -0.117 | -0.083 | -0.104 | -0.106 | -0.111 | -0.095 | -0.076 | -0.099 | -0.128 |
| | (0.491) | (0.537) | (0.357) | (0.368) | (0.346) | (0.361) | (0.433) | (0.445) | (0.659) |
| University or Fachhochschule | -0.012 | 0.017 | -0.001 | 0.012 | 0.037 | 0.051 | 0.065 | 0.053 | 0.002 |
| | (0.443) | (0.441) | (0.298) | (0.305) | (0.300) | (0.338) | (0.428) | (0.475) | (0.700) |
| German | 0.333 | 0.370 | 0.381 | 0.391 | 0.376 | 0.362 | 0.363 | 0.395 | 0.341 |
| | (0.115) | (0.108) | (0.076) | (0.076) | (0.077) | (0.094) | (0.125) | (0.143) | (0.202) |
| Urban | 0.236 | 0.223 | 0.225 | 0.219 | 0.222 | 0.231 | 0.237 | 0.223 | 0.265 |
| | (0.236) | (0.237) | (0.244) | (0.237) | (0.246) | (0.248) | (0.244) | (0.241) | (0.259) |
| Married | 0.050 | 0.065 | 0.062 | 0.054 | 0.034 | 0.033 | 0.035 | 0.030 | 0.029 |
| | (0.274) | (0.276) | (0.282) | (0.277) | (0.283) | (0.282) | (0.282) | (0.280) | (0.292) |
| East | -0.668 | -0.719 | -0.756 | -0.754 | -0.734 | -0.704 | -0.686 | -0.691 | -0.719 |
| | (0.293) | (0.294) | (0.297) | (0.296) | (0.304) | (0.305) | (0.307) | (0.308) | (0.319) |
| Work experience | -0.010 | -0.010 | -0.012 | -0.011 | -0.011 | -0.012 | -0.013 | -0.012 | -0.010 |
| | (0.115) | (0.111) | (0.113) | (0.112) | (0.115) | (0.116) | (0.116) | (0.116) | (0.122) |
| Work experience (squared) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.118) | (0.116) | (0.114) | (0.113) | (0.114) | (0.116) | (0.114) | (0.119) | (0.120) |
| Constant | 5.642 | 6.121 | 6.825 | 6.904 | 6.733 | 6.923 | 6.726 | 6.531 | 7.121 |
| | (0.078) | (0.078) | (0.078) | (0.078) | (0.078) | (0.078) | (0.078) | (0.078) | (0.078) |
| Observations | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 | 3,590 |

Source: German Socio-Economic Panel (GSOEP), 2012/2013, and own calculations.
Bootstrapped standard errors (100 replications) in parentheses.

# Chapter 5

# Dissertation Summary and Conclusion

Over the last three decades, Germany has experienced profound changes in its wage and earnings distribution. These trends have not been unique to Germany but could be observed to varying degrees both in the U.S. as well as in many other European countries. Against this background, the present doctoral thesis is dedicated to a scientific inquiry into different dimensions of the German wage and earnings distribution. A particular focus of the analysis is on the application of modern econometric estimation methods, including the further development of a method for the estimation of counterfactual quantile decompositions with selection correction in chapter 4.

Chapter 2 investigated changes in the German wage distribution among men over the period 1995-2010, i.e. a time when Germany experienced a sharp rise in wage inequality. We are particularly interested in determining the relative importance of the most prominent factors known from the literature which are simultaneously accounted for in the RIF decomposition (Firpo et al., 2009, 2018) analysis. Another important aspect is the fact that, contrary to previous studies based on administrative data, we are able to include hourly wages that are not censored at the social security contribution threshold. These represent a more direct measure of the prices paid in the market as they are, contrary to the commonly used daily earnings, not confounded by labor supply decisions and offer differences in hours worked. Our results show that the described restrictions in administrative data do not pose a major threat to the validity of previous studies as neither the usage of daily earnings nor an artificial censoring at the social contribution threshold change our result in an economically meaningful way. We find that observed changes in the German wage distribution can largely be explained by composition effects, i.e. by distributional changes in observable characteristics of the population. The study especially highlights the importance of unions and further shows that significant parts of de-unionization occurred within establishments, an aspect that has not been considered in the related literature (see, e.g., Dustmann et al., 2009, Baumgarten et al., 2018). Hence, we conclude that previous studies likely underestimated the role played by de-unionization relative to other prominent explanations such as SBTC, internationalization or firm heterogeneity. At the same time, our findings on the significance of personal characteristics coincide with the SBTC hypothesis in the sense that the increasing demand for higher skills was met by educational upgrading as well as by a demographically-induced aging of the workforce.

While chapter 2 was devoted to an inter-temporal comparison of cross-sectional wage distributions, chapter 3 broadened this perspective by pursuing a cohort approach in the study of distributional changes in lifetime earnings among German men of birth cohorts 1955-1974. From a theoretical point of view, this type of analysis provides important additional insights due to the more direct connection with individual consumption possibilities and the possibility to uncover differences across cohorts. Thereby, I add to a relatively new but growing literature by providing a first attempt to disentangle recent changes in lifetime earnings by means of a detailed decomposition analysis. The results suggest that changing employment patterns, both in terms of an increasing incidence of part-time and non-employment, explain a substantial share of increasing inequality. At the same time, the ongoing educational expansion during the period of study was generally beneficial to individuals of later cohorts, but increased inequality in the upper half of the distribution. These findings are also confirmed in alternative specifications which equally show the composition effect of education to be partly due to resulting changes in the age at labor market entry. Furthermore, controlling for detailed occupations only yields a very moderate effect, which I take as evidence that the influence of SBTC was primarily reflected in the educational upgrading of later cohorts. In this respect, the results on lifetime earnings show clear parallels with the findings on wages, despite the partly different modeling approach (the decomposition in chapter 2 includes *BIBB-IAB* tasks, whereas the one in chapter 3 directly controls for occupations). Also similar to chapter 2, the results for the detailed returns effects show a less clear picture. However, they reveal potentially important effects in relation to employment interruptions in the sense that, at the bottom of the distribution, periods of non-employments seem to have been associated with higher long-term earnings losses for later cohort. In addition to increasing inequality, the results also show an overall stagnation of earnings up-to-age 40 which were accompanied by losses within education subgroups. This finding is of special policy relevance as it shows that the younger cohorts did not benefit from the cross-sectional wage gains usually found during the study period 1975-2014 (see, e.g. Dustmann et al., 2014).

As the comparison of female wage distributions over time is inherently complicated by a changing selection into the workforce, the analysis in chapters 2 and 3 is limited to male distributions. Chapter 4, by contrast, includes a study which, in addition to its methodological contribution, evaluates the importance of selection effects in the context of the gender wage gap. Our econometric approach enhances the Albrecht et al. (2009) method by providing a practical solution to the problem outlined in Huber and Melly (2015). Using the Huber/Melly test, we show that conditional independence is in fact rejected in a typical application which invalidates the results. We then introduce a transformation which effectively eliminates violations of conditional independence. In addition, we illustrate the functionality of this approach from both a theoretical and an empirical point of view. We believe this to be an important contribution as it points a way to shield users of the Albrecht et al. (2009) method from the critique raised in Huber and Melly (2015). Our results reveal that a positive female selection on unobservables potentially leads to an underestimation of the gender pay gap in Germany. Importantly, we show this bias to be aggravated when our transformation is not applied. Given the sparsity of the previous literature, we highlight our findings on unobservable selection as a second important result of the study.

The presented results provide many opportunities for future research. With regard to chapter 2, there remains an open question as to why wage inequality has not increased further after 2010. Due to the late publication of the *GSES* 2014 and the comparatively difficult data access, the answer to this question goes beyond the scope of the study. Given the findings from this thesis, follow-up research should place particular emphasis on the impact of unions on the most recent development. This seems all the more important as the analysis indicates that the slowdown in the increase in inequality between 2006 and 2010 was also accompanied by a reduced pace of de-unionization. Similarly, possible interactions with SBTC and internationalization should be investigated further, as it seems unlikely that the sharp decline in collective bargaining coverage occurred independently of these other factors. In chapter 3, it would be interesting to see whether de-unionization has played a similarly important role in changing the distribution of lifetime earnings. This question is not addressed in the thesis due to the lack of information on unionization in the *SIAB*. An investigation on the development of lifetime earnings on the household level could denote another promising field of study, assuming the availability of suitable data. For example, the adverse trends found for men of later cohorts could potentially be moderated by higher earnings of women. In addition to its methodological contribution, chapter 4 provides important insights into the effect of unobservable female selection on the gender wage gap in Germany. Thereby, it would be desirable for future research to equally study the impact of male selection, which is a particularly challenging task due to the difficulty of finding suitable instruments. In answering this question, the method of D'Haultfoeuille et al. (2018), which requires no additional instruments for selection, offers a possible workaround. Another promising path for future research could be to allow for heterogeneity in the selection rule itself, which potentially varies for women in different parts of the German wage distribution. In a recent contribution, Machado (2017) suggests a possible implementation which is, however, limited to the subpopulation of 'always employed' women.

In conclusion, the distribution of wages and earnings remains a highly relevant topic as well as an active area of research. My dissertation contributed to this discussion in the form of three empirical studies, each of which deals with a different dimension of wage and earnings inequality in Germany. While the thesis delivered important new insights, both economic and methodological, the preceding chapters also show that the topic still offers numerous possibilities for future research.

# Bibliography


Acemoglu, D., and D. Autor (2011): Skills, Tasks and Technologies: Implications for Employment and Earnings. In: Ashenfelter, O., D. Card (editors), *Handbook of Labor Economics*, volume 4b, chapter 12, 1044-1166. North Holland, Amsterdam.

Albrecht, J., A. van Vuuren, and S. Vroman (2009): Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands. *Labour Economics*, 16(4), 383-396.

Antonczyk, D., T. DeLeire, and B. Fitzenberger (2018): Polarization and Rising Wage Inequality: Comparing the U.S. and Germany. *Econometrics*, 6, 20.

Antonczyk, D., B. Fitzenberger, and U. Leuschner (2009): Can a Task-Based Approach Explain the Recent Changes in the German Wage Structure? *Journal of Economics and Statistics*, 229(2-3), 214-238.

Antonczyk, D., B. Fitzenberger, and K. Sommerfeld (2010): Rising Wage Inequality, the Decline of Collective Bargaining, and the Gender Wage Gap. *Labour Economics*, 17(5), 835-847.

Antoni, M., A. Ganzer, and P. vom Berge (2016): Sample of Integrated Labour Market Biographies (SIAB) 1975-2014. *FDZ data report, 04/2016 (en)*. Institute for Employment Research (IAB), Nürnberg.

Arellano, M., and S. Bonhomme (2017): Quantile Selection Models with an Application to Understanding Changes in Wage Inequality. *Econometrica*, 85(1), 1-28.

Atkinson, A.B., T. Piketty, and E. Saez (2011): Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1), 3-71.

Autor, D.H. (2014): Skills, Education, and the Rise of Earnings Inequality among the "Other 99 percent". *Science*, 344(6186), 843-851.

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

Autor, D.H., D. Dorn, G.H. Hanson, and J. Song (2014). Trade Adjustment: Worker-Level Evidence. *Quarterly Journal of Economics*, 129(4), 1799-1860.

Autor, D.H., F. Levy, and R.J. Murnane (2003): The Skill Content of Recent Technological Change: An Empirical Exploration. *Quarterly Journal of Economics*, 118(4), 1279- 1333.

Autor, D.H., L.F. Katz, and M.S. Kearney, (2008): Trends in US Wage Inequality: Revising the Revisionists. *Review of Economics and Statistics*, 90(2), 300-323.

Bachmann, R., R. Felder, and M. Tamm (2018): Labour Market Participation and Atypical Employment over the Life Cycle: A Cohort Analysis for Germany. *IZA Discussion Paper No. 12010*, IZA Institute of Labor Economics, Bonn.

Bartels, C., and K. Jenderny (2015): The Role of Capital Income for Top Income Shares in Germany. *World Top Incomes Database Working Paper 2015/1*, World Top Incomes Database, http://wid.world/.

Barth, E., A. Bryson, J.C. Davis, and R. Freeman (2016): It's Where You Work: Increases in Dispersion of Earnings across Establishments and Individuals in the United States. *Journal of Labor Economics*, 34(2), 67-97.

Barth, E., J.C. Davis, R.B. Freeman, and A.J. Wang (2017): The Effects of Scientists and Engineers on Productivity and Earnings at the Establishment Where They Work. *NBER Working Paper No. 23484*, National Bureau of Economic Research.

Baumgarten, D. (2013): Exporters and the Rise in Wage Inequality: Evidence from German Linked Employer-Employee data. *Journal of International Economics*, 90(1), 201-217.

Baumgarten, D., G. Felbermayr, and S. Lehwald (2018): Dissecting Between-Plant and Within-Plant Wage Dispersion - Evidence from Germany. *Collaborative Research Center Transregio 190*, Discussion Paper No. 125, Ludwig-Maximilians-University Munich.

Baumgarten, D., I. Geishecker, and H. Görg (2013): Offshoring, Tasks, and the Skill-wage Pattern. *European Economic Review*, 61, 132-152.

Beblo, M., D. Benninger, and F. Laisney (2003): Measuring Selectivity-corrected Gender Wage Gaps in the EU. *ZEW Discussion Paper 03-74*, Centre for European Economic Research, Mannheim.

Beblo, M., and E. Wolf (2002): How Much Does a Year Off Cost? Estimating the Wage Effect of Employment Breaks and Part-Time Periods. *Brussels Economic Review*, 45(2), 191-217.

Biewen, M., B. Fitzenberger, and J. de Lazzer (2018): The Role of Employment Interruptions and Part-Time Work for the Rise in Wage Inequality. *IZA Journal of Labor Economics, 7:10.*

Biewen, M., Fitzenberger, B. and M. Seckler (2019): Counterfactual Quantile Decompositions with Selection Correction Taking into Account Huber/Melly (2015): An Application to the German Gender Wage Gap. Unpublished manuscript, University of Tübingen and Humboldt-University Berlin.

Biewen, M., and A. Juhasz (2012): Understanding Rising Inequality in Germany, 1999/2000-2005/2006. *Review of Income and Wealth*, 58(4), 622-647.

Biewen, M., and D. Plötze (2019): The Role of Hours Changes for the Increase in German Earnings Inequality. *Journal of Economics and Statistics*, 239(2), 277-304.

Biewen, M., and M. Seckler (2017): Changes in the German Wage Structure: Unions, Internationalization, Tasks, Firms, and Worker Characteristics, *IZA Discussion Paper No. 10763*. IZA Institute of Labor Economics, Bonn.

Biewen, M., and M. Seckler (2019): Unions, Internationalization, Tasks, Firms, and Worker Characteristics: A Detailed Decomposition Analysis of Rising Wage Inequality in Germany. *The Journal of Economic Inequality*. Advance online publication. doi: https://doi.org/10.1007/s10888-019-09422-w.

Blau, F.D., and L.M. Kahn (2006): The U.S. Gender Pay Gap in the 1990s: Slowing Convergence. *Industrial and Labor Relations Review*, 60(1), 45-66.

Blau, F.D., and L.M. Kahn (2017): The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3), 789-865.

Blinder (1973): Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4), 436-455.

Blundell, R., M.C. Dias, C. Meghir, and J. Shaw (2016): Female Labor Supply, Human Capital and Welfare Reform. *Econometrica*, 84(5), 1705-1753.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): Changes in the Distribution of Male and Female Wages Accounting for the Employment Composition Using Bounds. *Econometrica*, 75(2), 323-363.

Bofinger, P. (2015): German Wage Moderation and the EZ Crisis. Voxeu.org, https://voxeu.org/article/german-wage-moderation-and-ez-crisis, last access on July 16, 2019.

Böhm, M., H.-M. von Gaudecker, and F. Schran (2016): The Polarization of Skill Prizes in Germany, 1985-2010. *Working Paper*, University of Bonn.

Boll, C., M. Jahn, and A. Lagemann (2017): The Gender Lifetime Earnings Gap: Exploring Gendered Pay from the Life Course Perspective. *Journal of Income Distribution*, 25(1), 1-53.

Bollinger, C., J.P. Ziliak, and K.R. Troske (2011): Down from the Mountain: Skill Upgrading and Wages in Appalachia. *Journal of Labor Economics*, 29(4), 819-857.

Bönke, T., G. Corneo, and H. Lüthen (2015a): Lifetime Earnings Inequality in Germany. *Journal of Labor Economics, 2015*, 33(1), 171-208.

Bönke, T., M. Giesecke, and H. Lüthen (2015b): The Dynamics of Earnings in Germany: Evidence from Social Security Records. *Ruhr Economic Papers No. 582*, RWI - Leibniz Institute for Economic Research, Essen.

Bowlus, A., and J.M. Robin (2004): Twenty Years of Rising Inequality in U.S. Lifetime Labour Income Values. *Review of Economic Studies*, 77(3), 709-742.

Boysen-Hogrefe, J., and D. Groll (2010): The German Labour Market Miracle. *National Institute Economic Review*, 214(1), R38-R50.

Brändle, T., W. D. Heinbach, and M. F. Meier (2011): Tarifliche Öffnung in Deutschland: Ausmaß, Determinanten, Auswirkungen. *Journal for Labour Market Research*, 44(1-2), 163-172.

Brehmer, W., and H. Seifert (2008): Sind atypische Beschäftigungsverhältnisse prekär? Eine empirische Analyse sozialer Risiken. *Journal for Labour Market Research*, 41(4), 501-531.

Brenke, K. (2011): Anhaltender Strukturwandel zur Teilzeitbeschäftigung. *DIW Wochenbericht*, 78(42), 2-12.

Brenke, K. (2015): Wachsende Bedeutung der Frauen auf dem Arbeitsmarkt. *DIW Wochenbericht*, 82(5), 75-86.

Buchinsky, M. (1998): The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach. *Journal of Applied Econometrics*, 13(1), 1-30.

Buchinsky, M. (2001): Quantile Regression with Sample Selection: Estimating Women's Return to Education in the U.S. *Empirical Economics*, 26(1), 87-113.

Bundesagentur für Arbeit (2015): Arbeitsmarkt in Zahlen, Jahreszahlen: Arbeitslosigkeit im Zeitverlauf 2015. Statistik der Bundesagentur für Arbeit, Nürnberg.

Burda, M.C., and J. Hunt (2011): What Explains the German Labor Market Miracle in the Great Recession? *Brookings Papers on Economic Activity*, 42(1), 273-335.

Caliendo, M., A. Fedorets, M. Preuss, C. Schröder, and L. Wittbrodt (2017): The Short-Term Distributional Effects of the German Minimum Wage Reform. *IZA Discussion Paper No. 11246*. IZA Institute of Labor Economics, Bonn.

Caliendo, M., A. Fedorets, M. Preuss, C. Schröder, and L. Wittbrodt (2018): The Short-Run Employment Effects of the German Minimum Wage Reform. *Labour Economics*, 53, 46-62.

Card, D., and J. DiNardo (2002): Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. *Journal of Labor Economics*, 20(4), 733-783.

Card, D., J. Heining, and P. Kline (2013): Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics*, 128(3), 967-1015.

Chen, S., and S. Khan (2003): Semiparametric Estimation of a Heteroskedastic Sample Selection Model. *Econometric Theory*, 19(6), 1040-1064.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013): Inference on Counterfactual Distributions. *Econometrica*, 81(6), 2205-2268.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2015): The Association Between Income and Life Expectancy in the United States, 2001-2014. *Jama*, 315(16), 1750-1766.

Chzhen, Y., and K. Mumford (2011): Gender Gaps Across the Earnings Distribution for Full-Time Employees in Britain: Allowing for Sample Selection. *Labour Economics*, 18(6), 837-844.

Corneo, G. (2015): Income Inequality from a Lifetime Perspective. *Empirica*, 42(2), 225-239.

Daly, M.C., and R.G. Valletta (2006): Inequality and Poverty in United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behaviour. *Economica*, 73(289), 75-98.

Deutsche Bundesbank (2018): Wage Growth in Germany: Assessment and Determinants of Recent Developments. *Deutsche Bundesbank Monthly Report, April 2018*, 13-27.

D'Haultfoeuille, A. Maurel, and Y. Zhang (2018): Extremal Quantile Regressions for Selection Models and the Black-White Wage Gap. *Journal of Econometrics*, 203(1), 129-142.

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

Dreher, A., and N. Gaston (2007): Has Globalisation Really Had No Effect on Unions? *Kyklos*, 60(2), 165-186.

Duca, J.V., and J.L. Saving (2016): Income Inequality and Political Polarization: Time Series Evidence over Nine Decades. *Review of Income and Wealth*, 62(3), 445-466.

Dustmann, C., B. Fitzenberger, and A. Spitz-Oener (2014): From Sick Man of Europe to Economic Superstar: Germany's Resurgent Economy. *Journal of Economic Perspectives*, 28(1), 167-188.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009): Revisiting the German Wage Structure. *Quarterly Journal of Economics*, 124(2), 843-881.

Ebenstein, A., A. Harrison, M. McMillan, and S. Phillips (2014): Estimating the Impact of Trade and Offshoring on American Workers Using the Current Population Surveys. *Review of Economics and Statistics*, 96(4), 581-595.

Ehrl, P. (2017): A Breakdown of Residual Wage Inequality in Germany: Wage Decompositions using Worker-, Plant-, Region-, and Sector-Specific Determinants. *Oxford Economic Papers*, 69(1), 75-96.

Federal Statistical Office of Germany (1999-2014): *Volkswirtschaftliche Gesamtrechnung, Input-Output-Rechnung*, Fachserie 18, Reihe 2, Wiesbaden.

Felbermayr, G., M. Battisti and S. Lehwald (2016): Einkommensungleichheit in Deutschland, Teil 1: Gibt es eine Trendumkehr? *Ifo Schnelldienst*, 69(13), 28-37.

Fernández-Kranz, D., M. Paul, and N. Rodríguez-Planas (2015): Part-Time Work, Fixed-Term Contracts, and the Returns to Experience. *Oxford Bulletin of Economics and Statistics*, 77(4), 512-541.

Fernandez-Val, I., A. van Vuuren, and F. Vella (2018): Nonseparable Sample Selection Models with Censored Selection Rules. Working Paper in Economics No. 716, Department of Economics, University of Gothenburg.

Fields, G.S. (2003): Worker Well-Being and Public Policy. *Research in Labor Economics*, 22(1), 1-38.

Firpo, S., N.M. Fortin, and T. Lemieux (2009): Unconditional Quantile Regression. *Econometrica*, 77(3), 953-973.

Firpo, S., N.M. Fortin, and T. Lemieux (2014): Occupational Tasks and Changes in the Wage Structure. *Working Paper*, University of British Columbia.

Firpo, S., N.M. Fortin, and T. Lemieux (2018): Decomposing Wage Distributions Using Recentered Influence Function Regressions. *Econometrics*, 6, 28.

Fitzenberger, B. (1999): Wages and Employment Across Skill Groups: An Analysis for West Germany. *ZEW Economic Studies*, Vol. 6, Centre for European Economic Research, Mannheim.

Fitzenberger, B., and J. DeLazzer (2019): Changing Selection into Full-Time Work and its Effects on Wage Inequality - An Application to Germany. Unpublished manuscript, Humboldt-University Berlin.

Fitzenberger, B., K. Kohn, and A. Lembcke (2013): Union Density and Varieties of Coverage: The Anatomy of Union Wage Effects in Germany. *Industrial and Labor Relations Review*, 66(1), 169-197.

Fitzenberger, B., K. Kohn, and W. Qingwei (2011): The Erosion of Union Membership in Germany: Determinants, Densities, Decompositions. *Journal of Population Economics*, 24(1), 141-165.

Fitzenberger, B., A. Osikominu, and R. Völter (2006): Imputation Rules to Improve the Education Variable in the IAB Employment Subsample. *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 126(3), 405-436.

Fitzenberger, B., and G. Wunderlich (2004): Holen die Frauen auf? *ZEW Wirtschaftsanalysen, Vol. 69*, Centre for European Economic Research, Mannheim.

Fortin, N.M., T. Lemieux, and S. Firpo (2011): Decomposition Methods in Economics. In: Ashenfelter, O., D. Card (editors), *Handbook of Labor Economics*, volume 4a, chapter 1, 1-102. North Holland, Amsterdam.

Franz,W., and V. Steiner (2000): Wages in the East German Transition Process: Facts and Explanations. *German Economic Review*, 1(3), 241-269.

Gallant, A.R., and W.N. Douglas (1987): Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica*, 55(2), 363-390.

Gardeazabal, J., and A. Ugidos (2004): More on Identification in Detailed Wage Decompositions. *Review of Economics and Statistics*, 86(4), 1034-1036.

Gartner, H. (2005): The Imputation of Wages Above the Contribution Limit with the German IAB Employment Sample. *FDZ-Methodenreport*, No. 02/2005. Institute for Employment Research (IAB), Nürnberg.

Gathmann, C., and U. Schönberg (2010): How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics*, 28(1), 1-49.

Geishecker, I., and H. Görg (2008): Winners and Losers: A Micro-Level Analysis of International Outsourcing and Wages. *Canadian Journal of Economics*, 41(1), 243-270.

Gius, M. (2014): The Impact of Job Mobility on Earnings: Using Occupational and Industrial Classifications to Identify Job Changes. *International Review of Applied Economics*, 28(2), 181-190.

Goldin, C., and L.F. Katz (2008): *The Race between Education and Technology*. Harvard University Press, Massachusetts.

Goos, M., A. Manning, and A. Salomons (2009): Job Polarization in Europe. *American Economic Review*, 99(2), 58-63.

Görlich, D., and A. De Grip (2009): Human Capital Depreciation During Hometime. *Oxford Economic Papers*, 61(suppl.1), i98-i121.

Gould, E. (2019): State of Working America Wages 2018: Wage Inequality Marches on - and is Even Threatening Data Reliability. Economic Policy Institute, Washington, DC.

Guevenen, F., G. Kaplan, J. Song, and J. Weidner (2017): Lifetime Incomes in the United States over Six Decades. *NBER Working Paper No. 23371*, National Bureau of Economic Research.

Harrigan, J., A. Reshef, and F. Toubal (2016): The March of the Techies: Technology, Trade, and Job Polarization in France, 1994-2007. *NBER Working Paper No. 22110*, National Bureau of Economic Research.

Heckman, J.J. (1979): Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161.

Holmes, G. (1977): Gregory King and the Social Structure of Pre-Industrial England. *Transactions of the Royal Historical Society*, 27, 41-68.

Huber, M., and B. Melly (2015): A Test of the Conditional Independence Assumption in Sample Selection Models. *Journal of Applied Econometrics*, 30(7), 1144-1168.

Hyslop, D. R., and D. C. Mare (2005): Understanding New Zealand's Changing Income Distribution 1983-98: A Semiparametric Analysis. *Economica*, 72(287), 469-495.

Juhn, C., K.M. Murphy, and B. Pierce (1993): Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy*, 101(3), 410-442.

Kalina, T., and C. Weinkopf (2017): Niedriglohnbeschäftigung 2015 - bislang kein Rückgang im Zug der Mindestlohneinführung. *IAQ-Report*, 2017-06, University of Duisburg-Essen.

Katz, L.F., and D.H. Autor (1999): Changes in the Wage Structure and Earnings Inequality. In: Ashenfelter, O., D. Card (editors), *Handbook of Labor Economics*, volume 4b, chapter 26, 1463-1166. North Holland, Amsterdam.

Katz, L. F., and K. M. Murphy (1992): Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *Quarterly Journal of Economics*, 107(1), 35-78.

Klein, M.W., C. Moser, and D.M. Urban (2013): Exporting, Skills and Wage Inequality. *Labour Economics*, 25, 76-85.

Klein, R.W., and R.H. Spady (1993): An Efficient Semiparametric Estimator for Binary Response Models. *Econometrica*, 61(2), 387-421.

Kopczuk, W., E. Saez, and J. Song (2010): Earnings Inequality and Mobility in the United States: Evidence from Social Security Data. *The Quarterly Journal of Economics*, 125(1), 91-128.

Krueger, D., F. Perri, L. Pistaferri, and G.L. Violante (2010): Cross-sectional Facts for Macroeconomists. *Review of Economic Dynamics*, 13(1), 1-14.

Kunze (2017): The Gender Wage Gap in Developed Countries. In: Averett, S.L., L. M. Argys, S. D. Hoffman (editors): *Handbook on Women and the Economy*, Oxford University Press.

Lemieux, T. (2006): Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? *American Economic Review*, 96(3), 1-64.

Maasoumi, E., and L. Wang (2019): The Gender Gap Between Earnings Distributions. Forthcoming in: *Journal of Political Economy*.

Machado, C. (2017): Unobserved Selection Heterogeneity and the Gender Wage Gap. *Journal of Applied Econometrics*, 32(7), 1348-1366.

Machado, J.A.F., and J. Mata (2005): Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. *Journal of Applied Econometrics*, 20(4), 445-465.

Melly, B. (2005): Decomposition of Differences in Distribution Using Quantile Regression. *Labour Economics*, 12(4), 577-590.

Möller, J. (2016): Lohnungleichheit - Gibt es eine Trendwende? *IAB Discussion Paper 9/2016*. Institute for Employment Research (IAB), Nürnberg.

Mulligan, C.B., and Y. Rubinstein (2008): Selection, Investment, and Women's Relative Wages over Time. *The Quarterly Journal of Economics*, 123(3), 1061-1110.

Oaxaca, R. (1973): Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.

OECD (2018): OECD Compendium of Productivity Indicators 2018. OECD Publishing, Paris.

Ohlert, C. (2016): Establishment Heterogeneity, Rent Sharing and the Rise of Wage Inequality in Germany. *International Journal of Manpower*, 37(2), 210-228.

Olivetti, C., and B. Petrongolo (2008): Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps. *Journal of Labour Economics*, 26(4), 621-654.

Ostry, J.D., A. Berg, and C.G. Tsangarides (2014): Redistribution, Inequality and Growth. *IMF Staff Discussion Note SDN/14/02*, International Monetary Fund, Washington, DC.

Paul (2016): Is There a Causal Effect of Working Part-Time on Current and Future Wages? *Scandinavian Journal of Economics*, 118(3), 494-523.

Picchio, M., and C. Mussida (2011): Gender Wage Gap: A Semi-Parametric Approach with Sample Selection Correction. *Labour Economics*, 18(5), 564-578.

Piketty, T., and E. Saez (2014): Inequality in the Long Run. *Science*, 344(6186), 838-843.

Popli, G.K. (2013): Gender Wage Differentials in Mexico: A Distributional Approach. *Journal of the Royal Statistical Society*, 176(2), 295-319.

Potrafke, N. (2012): Unemployment, Human Capital Depreciation and Pension Benefits: An Empirical Evaluation of German Data. *Journal of Pension Economics and Finance*, 11(2), 223-241.

Rinawi, M., and U. Backes-Gellner (2015): Skill Prices, Skill Composition, and the Structure of Wages. *Discussion Paper No. 0112*, Institute for Strategy and Business Economics (ISU), University of Zurich.

Sandmo, A. (2015): The Principal Problem in Political Economy: Income Distribution in the History of Economic Thought. In: Atkinson, A.B., F. Bourguignon (editors), *Handbook of Income Distribution*, volume 2, chapter 1, 3-65. North Holland, Amsterdam.

Schank, T., C. Schnabel, and J. Wagner (2007): Do Exporters Really Pay High Wages? First Evidence from German Linked Employer-employee Data. *Journal of International Economics*, 72(1), 52-74.

Schnabel, C. (2016): United, Yet Apart? A Note on Persistent Labour Market Differences between Western and Eastern Germany. *Journal of Economics and Statistics*, 236(2), 157-179.

Seckler, M. (2019): Increasing Inequality in Lifetime Earnings: A Tale of Educational Upgrading and Changing Employment Patterns. *University of Tübingen Working Papers in Economics and Finance No. 119*.

Shorrocks, A.F. (2013): Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value. *The Journal of Economic Inequality*, 11(1), 99-126.

Slack (2004): Measuring the National Wealth in Seventeenth-Century England. *The Economic History Review*, 57(4), 607-635.

Smolny (2009): Wage Adjustment, Competitiveness and Unemployment - East Germany after Unification. *Journal of Economics and Statistics*, 229(2-3), 130-145.

Song, J., D.J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019): Firming Up Inequality. *Quarterly Journal of Economics*, 134(1), 1-50.

Spitz-Oener, A. (2006): Technical Change, Job Tasks, and Rising Educational Demands: Looking Outside the Wage Structure. *Journal of Labor Economics*, 24(2), 235-270.

Statistisches Bundesamt (2014): Mikrozensus: Bevölkerung und Erwerbstätigkeit: Stand und Entwicklung der Erwerbstätigkeit in Deutschland. *Fachserie 1, Reihe 4.1.1*, Statistisches Bundesamt.

Tisch, A., and S. Tophoven (2012): Employment Biographies of the German Baby Boomer Generation. *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 132(2), 205-232.

Töpfer, M. (2017): Detailed RIF Decomposition with Selection: the Gender Pay Gap in Italy. *Hohenheim Discussion Papers in Business, Economics and Social Sciences*, No. 26-2017, Faculty of Business, Economics and Social Sciences, University of Hohenheim.

von Gaudecker, H.-M., and R.D. Scholz (2007): Differential Mortality by Lifetime Earnings in Germany. *Demographic Research*, 17(4), 83-108.

Winkler, H. (2019): The Effect of Income Inequality on Political Polarization: Evidence from European Regions, 2002-2014. *Economics & Politics*, 31(2), 137-162.

Wolf, E. (2010): Lohndifferenziale zwischen Vollzeit- und Teilzeitbeschäftigten in Ost- und Westdeutschland. *WSI Working Papers No. 174*, Hans-Böckler-Foundation, The Institute of Economic and Social Research (WSI), Düsseldorf.