Beyond Conventional Wage Discrimination Analysis:

Assessing Comprehensive Wage Distributions of Males and Females using Structured Additive Distributional Regression

Alexander Sohn

SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
http://www.diw.de/soeppapers

Beyond Conventional Wage Discrimination Analysis
Assessing Comprehensive Wage Distributions of Males and Females using Structured Additive Distributional Regression

Alexander Sohn[*][1]

[1]Chair of Statistics, University of Göttingen

Abstract

In this paper, we propose a new comprehensive framework for analysing wage discrimination. This framework assesses wage discrimination on the grounds of conditional wage distributions (rather than just conditional means), regards the whole population (rather than just those in work) and employs a more general definition of work based on Margaret Reid's "third party criterion" (rather than a definition based on payment).

Examining wage discrimination with respect to gender in Germany, we find that gender wage discrimination is exacerbated if considered in such a framework, as women are faced not only with a lower expected pay, but also with a more unequal distribution, lower chances of employment and a higher workload of unpaid activities.

JEL-Classification: C13, C21, D31, J24, J31

Keywords: Wage Distribution; Discrimination; Gender Inequality; Structured Additive Distributional Regression; Polarisation

1 Introduction

Economics has a vast literature to offer on the analysis of discrimination, especially on discrimination of wages (and thus income). The most common approach to inquire into the existence and the extent of wage discrimination is to employ standard mean-based regression techniques to estimate Mincer-type wage equations (see Jenkins, 1994). From these, one can contemplate the discrepancy between the average wage a member of an allegedly discriminated group and the wage a member of the reference group with an equivalent set characteristics is paid on average.

---

[*]Corresponding author: asohn@uni-goettingen.de.

Given its large popularity in applied research, this conventional approach has faced criticism from various strands of the economic literature and other social sciences. Four aspects raised are the following:

1. Conventional analyses are focussed on differences between the two groups conditioned on a set of covariates. Hence, the attention is diverted away from differences in the possibilities in attaining a given set of covariates (see Neal and Johnson, 1996). If these possibilities are skewed against the discriminated group, these differences should be included in assessing the magnitude of discrimination, which they are likely to advance.

2. Conventional analyses are based on the comparison of conditional expectations. Therefore they disregard potential differences of the underlying conditional distributions (see van Kerm, 2013). If these distributions feature aspects detrimental to the discriminated group beyond a lower mean, the conventional approach underestimates discrimination.

3. Normally, conventional analyses are based on the wages of people in (full-time) employment. Thereby, differences in employment chances due to discriminatory practices are neglected (see Hook and Pettit, 2015). This yields a downward bias in the conventional estimates on the extent of discrimination if the discriminated group is less likely to find (full-time) employment.

4. Conventional analysis implicitly defines work as those activities which are remunerated on the labour market. It thus excludes any work outside formal employment from the analysis. Such a definition is highly problematic as it effectively imposes a dichotomy which includes some activities while excluding others not on grounds of the nature of the activity but the wage paid, thus selecting on grounds of the dependent variable (see England, 2005). If this selection runs to the detriment of the discriminated group, this also causes a downward bias in the discrimination analysis.

This paper addresses these issues by broadening the theoretical framework underlying the conventional perspective on discrimination and by adapting structured additive distributional regression to the analysis of conditional wage distributions. For illustrative purposes, we will concentrate on discrimination between women and men from here on, but of course the type of analysis is also applicable to other kinds of pecuniary discrimination.

The paper is structured as follows:

In the subsequent section, we propose our theoretical framework for the consideration of wage discrimination. In the following section, we describe the data from the German Socio-Economic Panel we used in this article for our analysis. In Section 4, we introduce structured additive distributional regression for estimating wage distributions conditional on a set of covariates. We find

that the more comprehensive approach that we propose significantly increases wage discrimination. The last section concludes.

2   A Wider Conception of Wage Discrimination

2.1   Wages, choices and circumstances

In this paper we conceive discrimination along the lines of the capabilities approach (see Sen, 1973) and the literature on inequality of opportunity (see Roemer, 1998). Thereby wage discrimination[1] exists if upon birth women and men face circumstances such that they are not endowed with equivalent chances and risks with respect to the pecuniary rewards for their hardship.

Let $Y_{i,t}$ be a random variable of the wage-rate offered and/or paid to individual i at time t. This wage-rate is thought to be derived from the individual's capability to pursue a set of activities required for some specific work at a given time and the surrounding circumstances of society via some function w, i.e.

$$Y_{i,t} = w(\mathbf{A}_{i,t}, \mathbf{S}_{i,t}), \tag{1}$$

where $\mathbf{A}_{i,t}$ is a vector denoting the set of activities which is selected from a space of possible human activities $\mathcal{A}$, while vector $\mathbf{S}_{i,t}$ denotes the circumstances in a society currently facing the individual stemming from the space of conceivable circumstance-scenarios $\mathcal{S}$. The circumstances comprise aspects like demand and supply for an activity, legislation, social norms, etc. and are thought of as stochastic but possibly dependent on the gender of the individual. The ability of the individual to perform an activity is thought to be dependent on the circumstances as well as on individual choices, both past and present. Formally, we can write

$$\mathbf{A}_{i,t} = a(\mathbf{S}_{i,t}, \mathbf{C}_{i,t}), \tag{2}$$

where a is a vector valued function which determines the ability to perform a set of activities in $\mathbf{A}_{i,t}$ based on the set of societal circumstances $\mathbf{S}_{i,t}$ and the individual's choices up to and including time t, denoted $\mathbf{C}_{i,t}$.[2] The choices $\mathbf{C}_{i,t}$ are also thought of as a stochastic vectorial variable from choice-space $\mathcal{C}$, which entails all possible choice combinations, including aspects like effort, having children, etc. As a function of two random variables, $\mathbf{A}_{i,t}$ can also be seen as random unless we

---

[1]Ultimately, discrimination should be measured by gender differences in the chances and risk regarding the pursuit of happiness. Yet, in light of the lack of adequate data on the highly elusive measures of utility, we content ourselves with analysing wages. The underlying hypothesis is that utility is by and large formed from the esteem or social acknowledgement of ones activity which finds its monetary expression in its remuneration.

[2]Here we simply suppose that a is defined in such a way that $\mathbf{A}_{i,t}$ is theoretically identifiable in Equation (1).

fully condition on both $\mathbf{S}_{i,t}$ and $\mathbf{C}_{i,t}$. For a schematic representation see Figure 1.
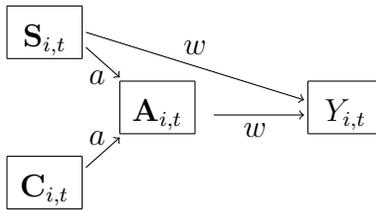

Figure 1: Relation of wages, choices and circumstances.

In this framework, wage discrimination is thought to exist, if and only if the circumstances facing men and women respectively induce structurally different resultant wage distributions. These differences can stem from Equation (1), i.e. direct discrimination, where despite the same ability to perform the necessary activities women are paid differently to men. However, in addition the differences may be the product of discrepancies of activities pursued by men and women. If these differences cannot be fully attributed to free choice but are at least in part due to differences in circumstances, this would constitute discrimination stemming from Equation (2). A concept which captures discrimination from both equations may be referred to as comprehensive wage discrimination. It is this concept that we will focus on in this article.

To facilitate the analysis, we consider the average wage rate, i.e.

$$Y_i = \frac{1}{|T_i|} \int_{t \in T_i} Y_{i,t} \, dt, \tag{3}$$

where time t is selected out of the time the individual spent on activities generally considered to be work, denoted T_i. The cardinality of T_i is denoted by $|T_i|$. Thereby, we allow for the inclusion of wage rates from various occupations over a given time-span (e.g. a month). As is standard in the literature, we will henceforth ignore all recreational activities. Note though that this does not necessarily imply that we only consider paid work (see below).

2.2 The conventional approach

The conventional economic concept of wage discrimination between men and women is focussed on the differences in the expected wages offered (and eventually paid) by employers conditional on a set of covariates, e.g. taste discrimination (see Flabbi, 2010) and statistical discrimination (see Gayle and Golan, 2012). In a typical set-up, one would estimate the expected (log-)wage[3] given

---

[3]Although the log-link allows for a nice interpretation in terms of elasticities, it also has some problems. For example, the nature of the log-link inhibits the inclusion of zero-wages. In addition, as Jenkins (1994) points out, everyday discussion are on the nature of wages and not log-wages.

a set of covariates $\mathbb{E}(\log(Y)\,|\,\mathbf{X} = \mathbf{x})$ both for men and women in the following linear set-up (see Jenkins, 1994):

$$\log(Y_i) = \tilde{\mathbf{X}}_i'\boldsymbol{\beta}^M + \varepsilon_i^M \quad \forall\ i \in M, \tag{4}$$

$$\log(Y_i) = \tilde{\mathbf{X}}_i'\boldsymbol{\beta}^W + \varepsilon_i^W \quad \forall\ i \in W, \tag{5}$$

where M and W denote the set of individuals in the sample who are male and female respectively and where $\tilde{\mathbf{X}}_i$ is the i-th row of the design matrix portraying the characteristics, $\mathbf{X}_i$. These characteristics can be thought of as proxies for the activities a person can do, and are thus called "explained" differences in the literature (see Fortin et al., 2011). For example, we tend to include education as individuals with higher education levels are thought to pursue different (supposedly more difficult) tasks that are awarded higher wages. On the other hand, the different coefficients for men ($\boldsymbol{\beta}^M$) and women ($\boldsymbol{\beta}^W$) can be thought to represent the circumstances facing an individual, which cause different remuneration depending on the gender. The wage differences resulting thereof are referred to as "unexplained" wage differences in the literature as they cannot be explained by other observable differences. Lastly, ε denotes the independent error terms centred around zero, which we will turn to in more detail below.

In such a framework, discrimination can be divided up. Direct discrimination on the labour market, or wage structure discrimination, which exists if and only if the coefficients $\boldsymbol{\beta}^M$ and $\boldsymbol{\beta}^W$ differ systematically inducing a different expected (log-)wage for a given set of covariates $\mathbf{X}_i = \mathbf{x}_i$. This discrimination perspective is first and foremost focussed on discrimination associated with the demand side of the labour market, that is discrimination in the recruitment and payment of personal which should be equally capable of performing a given activity. Such a discrimination analysis is encapsulated by Equation (1).

The second source, from which differences in the expected wage for men and women can arise, is the different composition of their characteristics. Such differences would be encapsulated in Equation (2). Although some approaches have been made in the direction of so called *detailed decomposition*, which tries to separate the impact of the individual elements on the expected wage differences, the main thrust of discriminatory analysis has been towards assessing the magnitude of the former type of discrimination analysis (see Oaxaca and Ransom, 1999).

Overall, we thus conclude that the economic literature has focussed first and foremost on discrimination stemming from Equation (1) and has by and large neglected discrimination stemming from Equation (2) and Equation (3).

2.3 Wage discrimination beyond the confinement of the labour market

If one is only interested in the discrimination caused by the "malfunctioning of the labour market" (Liu, 2015, p.2) in the sense that firms discriminate against equally qualified women, the conventional approach's focus on Equation (1) is appropriate. However, if one wants to inquire into the extent of the comprehensive discrimination as defined above, it is insufficient as circumstances embedding the labour market that inhibit or even prohibit women from attaining a position to perform a given activity are overlooked.

2.3.1 Covariates and Confounding

In their paper on discrimination, Neal and Johnson (1996) point to the high discrepancy in blacks' ability to attain the same productivity-related skills as whites. Circumstance-induced differences in productivity-related skills equally exist for the case of gender discrimination. One example, which unfortunately is not wholly unworldly, is women being forbidden any formal education in a society. If conditioning on education, as one would in a standard Mincer wage equation, much of the resultant wage discrimination would be concealed.

While in Western societies, such as Germany, women are by and large no longer facing discrimination in formal education (see Statistisches Bundesamt, 2013a,b), there is ample evidence for remaining sources of disadvantage for women inside and outside the labour market (see among others Ridgeway, 1997; Arulampalam et al., 2007; Hook and Pettit, 2015). For example, women are still much more likely to be involved in (mostly unpaid) care work but other gender specific covariate differences could equally be considered. If a mother cares for her children for a number of years she is likely to be unable to obtain work experience in full-time jobs at the same time. Thereby, women generally are disadvantaged at improving their human capital in terms of work experience with respect to men, who are less likely to do much care work. Yet, once we condition on this other key ingredient of Mincer-type wage equations without explicit contemplation of the discriminatory nature of its derivation, the assessment of the magnitude of wage discrimination is likely to suffer from confounding and underestimate the degree of wage discrimination in the comprehensive framework we propose.

A major obstacle for the assessment of the degree of discrimination is the question to which extent the differences in work experience, or any other covariate, can be ascribed to discriminatory circumstances and to which extent they are the result of genuinely free choices. Given the huge complexity of the nature of human choices and the differences in circumstances the individuals are faced with (as well as their interdependencies), any attempt to truly disentangle and identify Equations (1) and (2) must be considered futile.

In light of this, it appears that the majority of the economic literature has generally opted to ascribe observed differences in the covariate vector to individuals' choices unless there is evidence that this is not the case. Indeed, concerning the involvement in care activities some may argue along the antiquated lines of thought of Auguste Compte that women have an inherently higher motivation and capability for such activities (see Guillin, 2009). Already at that time, this view was heavily criticised by John Stuard Mill for neglecting the cultural genesis of many aspects of human behaviour (ibid.). While we are far from denying the role free choice has to play, one has to admit that any axiomatic assignment of large parts of gender wage differentials to free choice is highly problematic. In the following, we will therefore follow dialectic reasoning and take the opposite view, i.e. that most choices taken by individuals are preconditioned on the circumstances facing an individual unless there is evidence to suppose otherwise.

2.4   Including the unemployed who work

A second probable consequence of the higher women's involvement in care work is that their ability to find a job is hampered by the care work's temporal needs (see England, 2005). As a result women are more likely to stay at home and hence out of employment. Therefore, any analysis that only considers those persons in work suffers from a selection bias if inferences are to be drawn with respect to the whole population and not just those in employment. In order to mitigate this bias, conventional analysis normally constructs a counterfactual where women out of employment are offered the same payment as those women with the same attributes in employment. However, such an assumption is problematic as it negates an important factual stochastic variation. This variation may be due to choice, and/or it may be due to circumstances. In want of any credible information, we propose to treat this as a stochastic variable and to include all its realisations. Thus we include women and men who are out of work in the sample with the wage that they actually earn, which is zero if they do not work.

Additionally, one should contemplate the implicit definition of T_i in Equation (3) underlying most wage discrimination analyses, which defines work as those activities which are rewarded with monetary payment. This definition is obviously problematic as the same activity can be considered working time if it is paid and recreational time if it is unpaid. For care work this leads to frequently observable scenarios where a caretaker is paid for her care taking activity and thus considered working while the mother doing the equivalent care taking activities is not and thus not considered working. Such disregard of unpaid activities, disproportionately affects women. As a result, we would argue that conventional analysis is liable to a systematic measurement error. In order to counter this we propose to consider work along the lines of the "third party criterion" (Reid, 1934), whereby an activity can be considered work iff it is delegable to another person. Thus, we propose

7

to consider not only time spent in a remunerated job but also to include time spent on housework, errands, care work and repair works by the individual in T_i. Naturally, this significantly alters the wage rates, as substantial working time is now endowed for zero earnings. In contrast to the conventional wage rate definition, we call this the comprehensive wage rate. It is this wage rate which concentrate on in the following and which we will refer to in the following if not specifically stated otherwise. In order to facilitate comparison with the literature, we also provide the estimates from Section 5 on basis of the conventional definition of wages in Section A.6 in the appendix.

2.5 Wages and their stochastic nature

Contrary to Equation (1) and economic theory one does not observe one single equilibrium wage for worker with equivalent (observable) characteristics but rather a labour market with major deviations from the expected wage. One can certainly attribute part of this residual variation to unobservable innate abilities, and economics frequently tries to do so by using individual specific effects. Yet on the one hand, it may be possible that many economists are tempted to search for causal explanations and downplay or outright negate "the fact that luck played a role in the outcome" (Kahneman, 2011, pp.178-179). On the other hand, it is very likely that even if wages were derived in a truly deterministic fashion, that the person could not be sure of the prospective earnings a priori, given the bounded rationality of human beings (c.f. Selten and Gigerenzer, 2002). Thus, from the personal perspective, which is pivotal to the assessment of utility, there is (considerable) risk involved in the assessment of an individual's earnings. Any discrimination analysis solely focussed on differences in expected wages neglects this important aspect of risk.

Consequently, Dolton and Makepeace (1985) remark that "[in] principle, the amount of sex discrimination should be deduced from a comparison of the distribution of earnings actually paid to females and the distribution when there is no discrimination." Thus, for the comparison of males and females we should ideally compare wage distributions conditional on the choices that individuals are free to make and circumstances unrelated to discrimination.

Distributions can be compared (and possibly ranked) with regard to several of their properties or attributes. The standard approach is to compare and rank distributions with respect to their expectation as is done for example in the case of the conventional Oaxaca-Blinder decomposition. Following the reasoning from above, we aim to additionally incorporate the aspect of inequality associated with a distribution. In this paper, we generally take the view that there is an underlying aversion of inequality in the wage distribution. This inequality aversion can be based either on the valuation of an equitable society (see among others Kolm, 1969) or based on the aversion of risk at an individual level (see among others Amiel and Cowell, 1994). The view we take here essentially rests on both of these aspects, with the desirability of an equitable treatment of the two sexes first

and foremost based on the former, while the lower inequality associated with the conditional wage distribution is considered desirable on basis of the latter.

Let us consider both men and women as economic agents with bounded rationality and a set of characteristics and preferences that leads them to choose an observable attribute vector, $\mathbf{X} = \mathbf{x}$, which may entail things like education, children, etc. If we suppose that, in a discrimination free world, the underlying characteristics and preferences are independent of gender[4], this would imply that empirical wage distributions of observably equivalent men, $\mathcal{D}_{\mathbf{x}}^{M}(Y)$, and women, $\mathcal{D}_{\mathbf{x}}^{W}(Y)$, should be asymptotically equivalent in the absence of discrimination.

Based on these wage distributions and the assumption that there are no structural differences in the wage-related utility derived from a job paying a given hourly wage rate $u(Y)$, we can easily adapt the expected utility functionals in van Kerm (2013) to write down the following identity for a world free of discrimination:

$$U(\mathcal{D}_{\mathbf{x}}^{M}) = \int_{0}^{\infty} u(Y)d\mathcal{D}_{\mathbf{x}}^{M}(Y) = \int_{0}^{\infty} u(Y)d\mathcal{D}_{\mathbf{x}}^{W}(Y) = U(\mathcal{D}_{\mathbf{x}}^{W}). \tag{6}$$

Any deviation from this identity would rest on deviations in the wage distributions, which can then be used to analyse the degree of discrimination between men and women, like the ratio between expected wage-related utility for women and men with specific observable attributes $\mathbf{X} = \mathbf{x}$, i.e.

$$\Delta_{\mathbf{x}} = U(\mathcal{D}_{\mathbf{x}}^{M})/U(\mathcal{D}_{\mathbf{x}}^{W}). \tag{7}$$

These differences would record the impediments for women in their pursuit of a decent wage for a specific subpopulation.

2.6   Towards an alternative perspective on discrimination

In the following, we will analyse wage discrimination using a distributional perspective. We will condition only on variables containing choices and circumstances independent of gender discrimination and take a marginal perspective with respect to variables which we consider to be down to or influenced by discriminatory circumstances. Concerning the sampling universe, we consider not only those in employment but also those out of employment with the actual earned wage rather than a counterfactually constructed wage (e.g. van Kerm, 2013).[5]   Lastly, we consider various

---

[4]A quick glance into the history of gender struggles shows how (portrayed) females preferences and roles have changed dramatically - from the time of the Great Goddesses of the Greek Neolithic (see Gimbutas, 2007) to the struggles of feminism in the modern era (e.g. Ryan, 1992). It thus seems far from absurd to suppose that once we abstract from cultural norms that underlying characteristics and preferences are firmly attached to gender.

[5]For a discussion regarding this issue see Section A.2 in the appendix.

activities outside employment situations on the labour market as work, like care work.

Before we go on to putting such analysis to practice, it should be re-emphasised that our approach rests on the far reaching assumption that structural differences in wage-related utility can be identified via the differences in the wage distributions conditioned on a few observable characteristics. This is of course a gross simplification and one should not make the mistake of taking this view without the required pinch of scepticism. However, just as this view is ideal-typical so is the predominant view that (implicitly) most of the differences can be explained by individual preferences. Both of these views are partly right and partly wrong. Therefore, we conceive our perspective as the dialectic complement to the predominant perspective in economics.

3 The Data

In order to analyse nature and magnitude of distributional differences between male and female wages we will use the German Socio-Economic Panel (SOEP) database (see SOEP, 2014) as our primary source of data. In addition we use the German *Mikrozensus* (see Section A.1.3 in the appendix).

From the SOEP we include persons aged between 21 and 60 years of age. In contrast to most other studies (e.g. Dustmann et al., 2009; van Kerm, 2013; Card et al., 2013) we explicitly include both people who are not in (paid) work and those who are, including civil servants and the self-employed for the latter. Thus our sampling universe is the whole population in the 40 year age range during which the majority of the population is actively involved on the labour market.This yields 4,198 male observations and 4,984 females observations for 2013.

As dependent variable of interest we consider the gross hourly wage, i.e. the gross earnings divided by the time spent working, in the comprehensive sense discussed above. Following van Kerm (2013), it is computed by taking the gross monthly earning in the current job (including payments for overtime) and dividing it by the number of hours worked per week multiplied by 4.32 (for week per months). For the latter we include not only actual working hours but also hours spent on making errands, house work, child care, caring for adult persons as well as repair works - i.e. activities which could in principle and are often, at least partially, serviced by paid work. Thus, we treat this work like a secondary (or tertiary) job which is remunerated with a zero wage rate. In the appendix we also provide gross hourly wages when only considering paid work as working time.

Concerning our explanatory variables, we consider the age and the education as is standard in a Mincer type wage equation. As noted by Morduch and Sicular (2002) the coarse discretisation of

continuous variables such as age which is commonly applied in the literature (e.g. Dustmann et al., 2009) can pose problems. Hence, we consider age in a continuous manner. With respect to education, we follow van Kerm (2013) and consider 4 levels of education which are constructed as follows: the first level entails all persons who only have general elementary education or less (i.e. those who fall under the ISCED97 categories 0-2 according to the SOEP). The second level incorporates the persons with completed secondary education, (ISCED level 3). The third level incorporates those with vocational training and Abitur as well as those with a higher vocational qualification (ISCED levels 4 and 5). The last level entails all those with completed higher education (ISCED level 6).

In addition we consider the nationality of the person as a binary variable differentiating only whether the person is German or not. Concerning the household characteristics of the persons we consider a binary variable on whether the person has children who are not yet economically independent in the sense that the household still receives child support for them. Lastly, we consider the federal state of residence.

We do not include aspects like job characteristics or the industry of employment, as these are typically strongly related to gender (see among others Charles and Grusky, 2004). As Richard Anker (1998, p.3) points out: "Occupational segregation by sex is extensive and pervasive and is one of the most important and enduring aspects of labor markets around the world." Conditioning on industry of employment or type of occupation would hence capture much of the variation between males and females which we consider discriminatory. Therefore, we take a marginal perspective rather than a conditional perspective with respect to these variables, as discussed above. In addition, we do not include individual specific effects, for a similar reasoning.

Overall, we thus consider one continuous variable, two binary variables, one categorical variables with four levels and one with 16 levels. Even if we use a rather coarse discretisation of the continuous variable age into four ten-year periods this would yield 2048 combinations if we would fully interact these variables. Given the samples sizes of the SOEP, it becomes quickly apparent that regularisation is required to provide stable estimates. This regularisation is provided by the estimation strategy, which we turn to next.

4 Estimation Strategy

For the estimation we employ the fundamental idea underlying all regression techniques that individuals close to one another in the covariate space should have a similar earnings distribution. In other words, if the covariate vector $\mathbf{x}_1$ of one individual does not deviate much from the covariate vector of a second individual $\mathbf{x}_2$, then the two individuals' earnings distributions $\mathcal{D}_{\mathbf{x}_1}$ and $\mathcal{D}_{\mathbf{x}_2}$ should adhere to a similar form. In order to make an estimation feasible, we use the well established

idea of imposing an additive structure onto the covariate effects.

In order to allow for potential differences in the wage impact of the covariates on the wage distributions for males and females, we will run two separate regressions for males and females, with each regression specified as follows.

4.1 A generic representation for the predictors

As a general framework we consider structured additive distributional regression (SADR) (Klein et al., 2015) whereby the distribution of wages, $\mathcal{D}_{\mathbf{x}}$, is conditioned on a set of covariates, $\mathbf{x}$. The conditional distribution is assumed to follow a parametric form. Thus, the conditional distribution can be written in the form $\mathcal{D}(\theta_1(\mathbf{x}), \ldots, \theta_K(\mathbf{x}))$, where $\theta_k(\mathbf{x})$ is the k-th parameter in the parametric distribution and is conditioned on the covariate combination of the specific stratum. For notational brevity we will drop the suffix $(\mathbf{x})$ in the following. Additionally, we define $\boldsymbol{\theta} = \theta_1, \ldots, \theta_K$.

In this paper we use the following generic representation for every parameter of the distribution: Each parameter θ_k can be linked to a structured additive predictor η^{θ_k} via a suitably specified link function, g_k, mapping the predictor into the parameter space such that $\theta_k = g_k^{-1}(\eta^{\theta_k})$. The predictor η^{θ_k} can be specified in the following form:

$$\eta^{\theta_k} = \beta_0^{\theta_k} + f_1^{\theta_k}(\mathbf{x}) + \ldots + f_{J_k}^{\theta_k}(\mathbf{x}), \tag{8}$$

where $\beta_0^{\theta_k}$ represents the intercept of the predictor and the functions $f_j^{\theta_k}(\mathbf{x})$, $j = 1, \ldots, J_k$, can capture both linear and non-linear effects of single or multiple elements of the covariate vector $\mathbf{x}$. The latter is done by means of representing the function by a suitable linear combination of basis functions which are generally penalised such that the non-parametric estimate adheres to the required smoothness (see Fahrmeir et al., 2013).

In our application we will use the following regression set-up for all the predictors:

$$\begin{aligned}
\eta^{\theta_k} = {} & \beta_0^{\theta_k} + \beta_1^{\theta_k} kids + \beta_2^{\theta_k} nat + \beta_3^{\theta_k} educ_2 + \beta_4^{\theta_k} educ_3 + \beta_5^{\theta_k} educ_4 \\
& + f_1^{\theta_k}(age) + heduc \cdot f_2^{\theta_k}(age) + f_{spat}^{\theta_k}(region),
\end{aligned} \tag{9}$$

with *kids* and *nat* binary, effect-coded variables set to unity if the person has at least one child and has German nationality, respectively. We use education-specific intercepts by three effect-coded variables $educ_e$, where the subscript e denotes the education level, with the first education level taken as the base. As mentioned above *age* is a continuous variable expressed in years. In the standard Mincer wage equation age, as a proxy for potential experience, is incorporated by a

polynomial of degree two (see Lemieux, 2006). While this linearity in parameters has proven to perform well for the expected (log) income, a more flexible effect of age seemed desirable for the possibly non-linear nature of the effect of age on the various parameters of the whole conditional wage distribution.[6] Hence, we use a flexible smooth function, based on P-splines (see Eilers and Marx, 1996; Brezger and Lang, 2006), to model the effect of age. Thereby, f generally consists of a number of basis functions allowing for a high degree of flexibility and a penalisation term ensuring the desired degree of smoothness adhered to by the function. To account for different developments over the life-span depending whether the person has enjoyed higher education and interaction with the effect-coded variable *heduc* which is unity if the person has a degree in higher education, i.e. if the ISCED level is 6 according to the SOEP.

In order to capture differences between the economic dynamism of different federal states in Germany, we include a hierarchical spatial effect, such that:

$$f_{spat}^{\theta_k}(region) = \beta_6 east + \gamma_{region}, \tag{10}$$

where *east* is an effect-coded binary variable that is unity if the federal state is situated in the East, thus capturing the difference between the former Federal Republic of Germany and the German Democratic Republic. The state-specific random effect is denoted γ_{region} and accounts for variations across individual states. For the random effect we impose a Gaussian prior centred around zero. By incorporating this region specific term we capture the effect of contemporary region specific effects such that past-unemployment does not induce an effect via the contemporary economic background facing workers.

In our application, we condition on a handful of aspects which we believe to be identifiable and independent from gender-related circumstances. This entails two variables which we regard as choices that are by and large independent of gender discrimination in contemporary Germany: the individual's choice of whether to have children and which federal state to live in. In addition, we condition on two variables which we consider as circumstances that are equally independent of gender discrimination: whether the individual has German nationality as well as the individual's age. Lastly, we consider a hybrid of the two - education. We conceive education free of discrimination in Germany such that dependent on his/her aptitude (which we consider a non-discriminatory circumstance) the individual is free to choose the education irrespective of the gender.

Naturally, this leaves out many important variables which, be they based on choices or circumstances, will affect the individual's wages independently of the gender. Next to the issues of data availability and model stability, we do not include further variables on the grounds that the mar-

---

[6]In addition, using the DIC as a model selection criterion wo show the non-linear approach to be superior over the linear approach - see Section Section A.8 in the appendix.

ginal perspective which we take with respect to the left-out variables emulates the individual's perspective, who we conceive to see the world through the eyes of a person with bounded cognitive abilities. To such a person many influences are likely to go unnoticed and much of the variation in wages which are explained by a score of different factors will seem as random. The distributions we thus model are conditioned on variables such that they aim to represent the perceived wage distributions for men and women with limited information at their disposal.

4.2 Parameter estimation

Estimation is performed in a Bayesian framework using Markov Chain Monte Carlo (MCMC) techniques implemented in the software BayesX (Belitz et al., 2015). See Klein et al. (2014) for details on the estimation procedure. Concerning the coefficients' prior distributions, we employ non-informative flat priors for the linear effects and multivariate normal distributions for the basis function coefficients of the smooth effects which are in turn scaled by inverse gamma hyperpriors aimed at providing a data-driven degree of smoothness. Concerning the MCMC algorithm, we use 200,000 MCMC realisations for burn-in and thin out the following 800,000 MCMC realisations by a factor of 800. Thus, we use 1,000 MCMC realisations for each predictor θ_k to construct the posterior of the predictor, which is then transformed by the corresponding link function g_k into the posterior distribution of the parameter of interest. These distributions are proper under mild conditions (see Klein et al., 2015). For our inferential purposes we use the median from the posterior as point estimate for the parameters in order to provide estimates for the resultant full conditional distribution for the desired covariate combination. Using the MCMC realisations, we also provide point-wise credible intervals giving a notion of uncertainty attached to the estimators.

4.2.1 Estimating the probability of zero wages

In order to incorporate zero-wages, we follow a two-stage strategy. In the first stage, we estimate the probability mass of zero-wage and positive wages respectively. This is done by simple logistic regression. In a subsequent estimation step we estimate the conditional distribution of wages greater than zero.

First, we thus estimate the probability that a person is without any remunerated employment, π_0, either because of lack of work or the unpaid nature of the work they do. In both cases the wage rate is considered zero, yielding the probability of receiving a zero-wage $\pi_0 = P(y = 0 \mid \mathbf{X} = \mathbf{x})$:

$$g_1(\pi_0) = \eta^{\pi_0}, \tag{11}$$

where g_1 is a logit-link and η^{π_0} is the predictor as specified in Equation (9) for a given covariate combination.

4.2.2 Estimating the density of positive wages

For the estimation of the distribution of positive wage rates, we use the three-parameter Dagum distribution which has a track record of performing very well for modelling earnings at the aggregate level (e.g. Kleiber and Kotz, 2003; Chotikapanich, 2008).[7] A natural alternative which is the workhorse distribution in the literature is the log-normal distribution (see Flabbi, 2010), yet preliminary studies have found this distribution to have problems in modelling conditional earning distributions (Sohn et al., 2014). Although much more research must be done on the nature of conditional wage distributions, we consider the Dagum distribution to be a good starting point as it allows for great flexibility in the modelling of wage distributions.

The density of the Dagum distribution[8] is given by

$$p_{Dag}(y \mid a, b, c) = \frac{acy^{ac-1}}{b^{ac}(1 + (y/b)^a)^{p+1}}, \quad a \in \mathbb{R}_{>0}, \ b \in \mathbb{R}_{>0}, \ c \in \mathbb{R}_{>0}. \tag{12}$$

The estimation of the three parameters of the Dagum distribution a, b and c is done using the following generic predictor set-up as discussed above:

$$g_2(a) = \eta^a, \tag{13}$$
$$g_3(b) = \eta^b, \tag{14}$$
$$g_4(c) = \eta^c, \tag{15}$$

where all three link functions are log-link functions ensuring a positive support for the parameters and inducing a multiplicative connection between the covariates.

Overall, the estimation procedure gives us four parameters to estimate over the covariates space. Using this parametrisation, the density of the conditional wage distribution can be expressed as a

---

[7]This distribution is very similar to the Singh-Maddala distribution which has been used by Biewen and Jenkins (2005) for conditional earning distributions and by van Kerm (2013) for modelling wage rates. Yet, Kleiber and Kotz (2003) remark that the Dagum distribution generally performs slightly better than the Singh-Maddala. It should be noted that other four- and five-parameter distributions have been suggested in the literature for aggregate income distributions which outperform the Dagum distribution. However, these more complex distributions have proven to be too complex for stable estimation. A comparative study on the performance of the fit of these various parametric alternatives would be needed though to provide more profound assessment on this issue.

[8]Dagum (1997) actually includes as fourth parameter the point mass at zero, which we estimate using Equation (11). Here we use the definition of Kleiber and Kotz (2003) who refer to the Dagum distribution as a three-parameter distribution excluding the parameter for the point mass at zero.

mixture of a point-mass at zero and a continuous distribution thereafter:

$$p(y \mid \pi_0, a, b, c) = \pi_0 \mathbb{1}_{\{y=0\}} + (1 - \pi_0)p_{Dag}(y \mid a, b, c), \tag{16}$$

where p denotes the probability mass or probability density for a given wage y. For the point mass of zero wages $\mathbb{1}_{\{y=0\}}$ denotes an indicator function which is unity for a wage of zero and thus gives a probability mass of π_0. For earnings greater than zero, we obtain the density as specified by the Dagum distribution.

It should be noted that currently a lot of work is being done on various other statistical approaches allowing for the estimation of conditional distributions. For a discussion of other estimation strategies, see the Section A.3 in the appendix.

4.3 Testing for Misspecification

One quintessential assumption made for our estimation approach is the use of an adequate parametric form for the conditional wage distributions. This assumption is tested by considering the following null-hypothesis $\mathcal{H}_0$ against the alternative hypothesis $\mathcal{H}_1$:

$\mathcal{H}_0$: The conditional wage distributions can be modelled by our parametric form, $p(y \mid \boldsymbol{\theta})$, for all observed values of y and some values of $\boldsymbol{\theta}$ derived for the corresponding covariates, $\mathbf{x}$.

vs.

$\mathcal{H}_1$: The conditional wage distributions cannot be modelled by our parametric form, $p(y \mid \boldsymbol{\theta})$, for all observed values of y and any values of $\boldsymbol{\theta}$ derived for the corresponding covariates, $\mathbf{x}$.

In order to test these hypotheses we will use an adaptation of the Kolmogorov-Smirnov test, which is arguably the most renowned and most widely used test for distributional assumptions. Our adaptation is based on the work of Andrews (1997) and Rothe and Wied (2013) who proposed a frequentist framework for the testing of conditional distributions. Using the idea to transform the conditional moment restrictions imposed by the parametric specification of our structured additive distributional regression model into unconditional ones (see Rothe and Wied, 2013), we are able to specify the test statistic T_n as

$$T_n = \sqrt{n} \sup_{(y,\mathbf{x})} \mid \hat{H}_n(y, \mathbf{x}) - \hat{H}_n^0(y, \mathbf{x}) \mid, \tag{17}$$

where $\hat{H}_n(y, \mathbf{x})$ and $\hat{H}_n^0(y, \mathbf{x})$ constitute estimates of the joint cumulative distribution function of both dependent and independent variable for n observations integrated up with respect to the

marginal distribution of the conditioning variables:

$$\hat{H}_n(y, \mathbf{x}) = n^{-1} \sum \mathbb{1}_{\{Y_i \leq y\}} \mathbb{1}_{\{\mathbf{X}_i \leq \mathbf{x}\}}$$

and

$$\hat{H}_n^0(y, \mathbf{x}) = n^{-1} \sum \hat{P}_n \mathbb{1}_{\{\mathbf{X}_i \leq \mathbf{x}\}},$$

where $\hat{P}_n$ denotes the estimated cumulative density function based on n samples using structured additive regression while $\mathbb{1}$ again denotes an indicator function.

As the asymptotic distribution of T_n under the null hypothesis depends on the data-generating process in a complex fashion we use a bootstrap procedure to simulate it. In order to incorporate the uncertainty attached to the parameter estimates we use draws from the MCMC realisations and contrast it with simulated realisations of Y for a set of randomly selected covariate combinations of $\mathbf{X}$. Our bootstrap algorithm thus as follows:

Step 1 Draw a bootstrap sample of covariates $\{\mathbf{X}_{b,i}; 1 \leq i \leq n\}$ with replacement from the obtained values in the sample $\{\mathbf{X}_i; 1 \leq i \leq n\}$.

Step 2 Randomly select the $\mathbf{m}$-th MCMC draw from set $\mathcal{M}$ for the parameter estimates, yielding $\theta_{b,1}^{(\mathbf{m})}(\mathbf{x}), \ldots, \theta_{b,K}^{(\mathbf{m})}(\mathbf{x})$.

Step 3 Use $\{\theta_{b,k}(\mathbf{x}); 1 \leq k \leq K\}$ and $\{\mathbf{X}_{b,i}; 1 \leq i \leq n\}$ to simulate $\{Y_{b,i}; 1 \leq i \leq n\}$ in accordance with the parametrically specified conditional distributions.

Step 4 Use bootstrapped data $\{Y_{b,i}; 1 \leq i \leq n\}$, $\{\mathbf{X}_{b,i}; 1 \leq i \leq n\}$ and $\{\theta_{b,k}(\mathbf{x}); 1 \leq k \leq K\}$ to compute estimates $\hat{H}_{b,n}$ and $\hat{H}_{b,n}^0$ yielding the bootstrap realisation of the test statistic: $T_{b,n} = \sqrt{n} \sup_{(y,\mathbf{x})} | \hat{H}_{b,n}(y, \mathbf{x}) - \hat{H}_{b,n}^0(y, \mathbf{x}) |.$

Using the simulated distribution of T_n we can then derive the corresponding p-value and or critical values to assess the test statistic.

For further details see Section A.7 in the appendix.

5 Assessing the Conditional Earnings Distributions

In this section we consider the differences between wages for males and females in a distributional perspective as discussed in Section 2.

Before looking at the results of the estimation, we will consider the test statistics discussed in

Section 4.3. Using 1,000 bootstrap repetitions, we consider the adequacy of our estimated distributions both for males and for females and obtain the following results. For men we obtain a p-value of 0.18. For women the p-value is 0.62. As these results imply that the hypothesis that our parametric fit is adequate to describing the nature of the conditional distributions cannot be rejected, we consider our specification of the conditional earning distributions to be adequate.

5.1 Interpreting the distributional effects at the disaggregated level

As is well-known from the literature on generalised linear models, the use of link functions in Equations (11)-(15) implies that the impact of explanatory variables varies across the covariate space (see among others Nelder and Wedderburn, 1972). Following Fox (1987) we employ effect displays for three different covariate combinations - see Figure 2. Subsequently we analyse some distribution measures for these distributions.

5.1.1 Visual analysis of conditional wage distributions

First, we display the estimated wage distribution for men (D_{1M}) and women (D_{1W}) with 30 years of age, no formal education beyond primary school, no children and no German citizenship who live in Eastern Germany, i.e. persons who we would typically expect at the lower bound of the income spectrum. As can be seen both for men and women there is a considerable proportion without employment (43% and 52% for men and women respectively). Considering only those in employment, we can observe a distribution for men which interestingly does not adhere to the standard positive skew observed in most income related distributions but is negatively skewed. As we would expect we also see that men's wages are higher and slightly more dispersed than those of women.

Below that, we display the estimated wage distribution for men (D_{2M}) and women (D_{2W}) who are forty years of age, have a completed education yielding an ISCED classification 3 or 4, have at least one child, have German citizenship and live in the West of Germany, i.e. persons who would be considered the "average Joe/Jane". As we can see the wage distributions differ substantially between men and women, not only in their first two moments but also in the shape of the distribution. As expected, women are much more likely to be receiving no wages at all than men are. If they are in employment their income distribution is also shifted to the left of the men's distribution, with the female distribution also portraying a higher skewness, implying that women are much more likely to find themselves below the expectation of their distribution.

The third set of distributions displays the wages of those who are generally associated with the upper strata of society - 50-year-olds, with higher education, children and living in West Germany
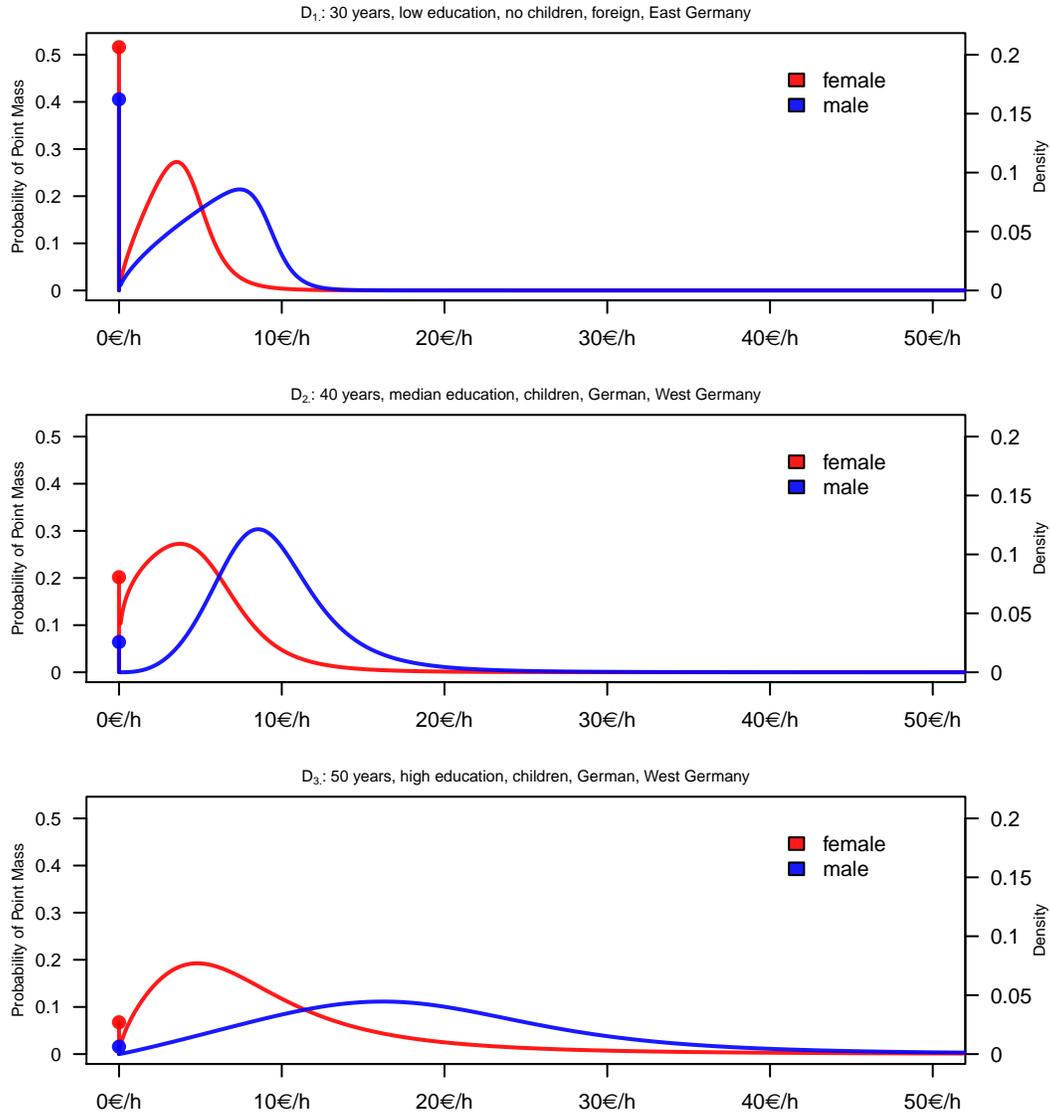
Figure 2: Conditional wage distribution estimates of males (blue) and females (red).

(D_{3M} and D_{3W}). It shows a much smaller probability mass for those without employment and a much wider distribution for those in employment. Again women generally earn a lower wage than men, with the shape differences similar to the one above, albeit a much higher expectation and dispersion both for men and women.

One of the problems with the direct interpretation of income distributions is that they are naturally very complex and cannot be as easily grasped as scalar distribution measures. In the following, we will thus consider a handful of measures which may facilitate the interpretation of the distributional differences between men and women.

5.1.2 Distributional measures of conditional wage distributions

	D_{1M}	D_{1W}	D_{2M}	D_{2W}	D_{3M}	D_{3W}
μ	3.44[2.76;4.23]	1.76[1.29;2.37]	8.91[8.07;9.87]	3.85[3.41;4.33]	19.36[17.29;22.07]	9.78[8.28;12.06]
Gini	0.25[0.19;0.31]	0.29[0.23;0.35]	0.21[0.18;0.25]	0.36[0.32;0.39]	0.31[0.25; 0.40]	0.54[0.42; 0.67]
AI(ϵ=0.5)	0.43[0.37;0.50]	0.50[0.41;0.59]	0.11[0.09;0.13]	0.28[0.23;0.33]	0.09[0.07; 0.12]	0.23[0.18; 0.29]
AI(ϵ=1.0)	0.90[0.86;0.93]	0.91[0.86;0.94]	0.41[0.35;0.47]	0.71[0.63;0.78]	0.23[0.18; 0.29]	0.54[0.46; 0.62]
AI(ϵ=1.5)	0.99[0.98;0.99]	0.98[0.98;0.98]	0.91[0.87;0.93]	0.96[0.94;0.97]	0.68[0.60; 0.75]	0.91[0.87; 0.94]
CE(ϵ=0.5)	1.96[1.42;2.56]	0.88[0.56;1.34]	7.94[7.18;8.85]	2.79[2.31;3.25]	17.67[15.59;20.15]	7.52[6.51; 8.88]
CE(ϵ=1.0)	0.35[0.22;0.55]	0.16[0.09;0.31]	5.27[4.54;6.05]	1.11[0.76;1.52]	14.91[12.90;17.24]	4.53[3.67; 5.56]
CE(ϵ=1.5)	0.05[0.04;0.07]	0.03[0.02;0.05]	0.84[0.62;1.12]	0.16[0.11;0.25]	6.24[4.82; 7.97]	0.87[0.56; 1.29]
P(α=β=1)	0.28[0.00;0.56]		1.15[0.88;1.41]		2.15[1.33;2.95]	

Table 1: Some distribution measures for 3 conditional wage distributions

In the first row of Table 1 we display the expected values (μ) for the six underlying wage distributions under consideration, which in conventional analysis is the main measure of interest[9]. In the squared brackets we display the 95% credible intervals. As would be expected, we can observe that for all three groups, the expected wage is lower for females for all three groups. The literature is full of such analyses and generally arrives at the same result, that there is wage discrimination between women and men. In terms of magnitude, the results in the literature generally find a smaller degree of discrimination, due to the fact that they neither consider those without any employment nor the work done outside employment on the labour market.

As we discussed in Section 2.5, it is important to go beyond the "one-dimensional straightjacket" (Cowell, 2011) of expected wage analysis and consider wages' stochastic nature. Here, we will focus on one additional distributional aspect in particular - the inequality associated with the distribution. To this end, we consider two types of distribution measures. The first being the Gini coefficient (G), which is generally the most widely used inequality measure (Cowell, 2000). As a second measure we consider the Atkinson inequality index (AI) proposed in Atkinson (1970), for which we consider three levels of inequality aversion ($\epsilon = \{0.5, 1.0, 1.5\}$).[10] Barring one exception (for D_{1M} and D_{1W}) we observe that the inequality according to the Gini coefficient is significantly greater among wage distributions of women than among that of men. For the Atkinson indices we observe a similar pattern, although the finding is not significant at the 5% level for a few more cases. If we regard inequality in the conditional wage distribution as undesirable, which despite some heterogeneity can generally be assumed according to the literature (e.g. Bellemare et al., 2008), this further disadvantage for women should be considered in an analysis of discrimination.

To this end, van Kerm (2013) propose the analysis of wage distributions using certainty equivalents

---

[9]This would be in line with the utility based approach put forward in Chapter 2 iff we had a linear utility function $u(Y)$.

[10]Note that the limiting form for $\epsilon \to 1$ involves the natural logarithm being applied to the wage. To obtain finite values, we replace zero-wages with very small wages (0.01€) as is standard in the literature.

which employ the Atkinson index and the underlying assumption of constant relative risk aversion to obtain a cardinal metric that is easily interpretable. In the table we thus display the certainty equivalents (CE) of the wage distributions, which show the wage level of a distribution with all its point mass centred at that wage level which is in terms of utility equivalent to the estimated distribution, i.e. the wage once we have adjusted for inequality differences. While this measure is treated in more detail below, a quick glance at the table already shows that the ratios between the selected males and females are generally greater than for the simple mean.

Lastly, we also display a measure on the polarisation (P) in Table 1, which has seen increasing attention in the literature in recent years following the publications from Esteban and Ray (1994) and Wolfson (1994). We employ the measure proposed in Esteban et al. (2007) where α is the sensitivity to polarisation between groups and β indicates the weight assigned to inequality within the wage distribution. Following Gradín (2000), we apply this measure to the wage distributions of the corresponding wage distributions of males and females using $\alpha = 1$ and $\beta = 1$. This thus gives us an indication on the degree of polarisation of wages paid to men and women respectively. As we can see, this is hardly the case for the first group, which as we observed in Figure 2 strongly overlapped, while it is more pronounced for the second and the third group, where a lot of wages earned by men are practically out of reach for women with equivalent characteristics $\mathbf{x}$.

One obvious problem with such analyses at the disaggregated level is that it only allows insights for one covariate combination at a time and does not give a comprehensive overview. In the following section, we thus aggregate the discrimination measures discussed in this section.

5.1.3 Assessing the magnitude of discrimination

As portrayed in Section 2.5, we use the ratio of both the expected wage as well as the certainty equivalents to get an impression of the magnitude of discrimination for the three groups under consideration. As Table 2 shows, the discrepancy is mostly larger once we consider the certainty equivalents with the magnitude heavily dependent on the degree of parameter aversion.

	$D_{1.}$	$D_{2.}$	$D_{3.}$
μ	1.96[1.36; 2.82]	2.32[1.98; 2.72]	1.98[1.55; 2.43]
CE(ϵ=0.5)	2.22[1.31; 3.81]	2.86[2.36; 3.51]	2.34[1.90; 2.85]
CE(ϵ=1.0)	2.19[0.99; 4.66]	4.75[3.36; 7.08]	3.29[2.57; 4.23]
CE(ϵ=1.5)	1.41[0.86; 2.22]	5.21[3.14; 8.68]	7.19[4.43;11.93]
μ_0^{conv}	1.10[0.91; 1.33]	1.29[1.16; 1.45]	1.46[1.26; 1.69]
μ_0	1.67[1.29; 2.14]	2.00[1.74; 2.31]	1.87[1.47; 2.30]

Table 2: Ratios of male and female wage distribution measures

Focussing on the group with the average men and women ($D_{2.}$), we can observe that when using an inequality aversion parameter of 1, the ratio roughly doubles. This equates to an increase in

the mark up paid to men in terms of the certainty equivalent wage from 132% to 375%. Or to look at the inverse, we find that while for the remuneration of an average women decreases from 43 cents to the Euro earned by an average man to just 21 cents. Regarding all three certainty equivalent measures, we can thus observe that for the first group, the discrimination measure actually decreases for higher inequality aversion. This shows that given the complex distributional differences between wage distributions of males and females yield complex outcomes when analysed by means of certainty equivalents.

In order to additionally give an indication of the differences in the magnitude of estimates along the lines of the conventional approach, we also provide the ratios akin to what is employed in the literature - differences based on wage rates based on a conventional wage definition and excluding the unemployed (μ_0^{conv}) as well as differences based on means using our comprehensive wage definition but excluding all individuals with a zero wage rate (μ_0). As would be expected, these estimates are much smaller than those we presented earlier. For more information on the conventional estimates, see the Section A.6 in the appendix.

5.2 An aggregate perspective on discrimination

In order to get an aggregate discrimination measure (Δ_a) we can simply integrate up the conditioning argument. Applied to a given discrimination measure ($\Delta_{\mathbf{x}}$) from Equation (7), we would integrate over the covariate space (Ξ), i.e.

$$\Delta_a = \int_{\mathbf{x} \in \Xi} \Delta_{\mathbf{x}} dF_{\mathbf{x}}(\mathbf{x}) \approx \sum_{\tilde{\mathbf{x}} \in \tilde{\Xi}} \Delta_{\tilde{\mathbf{x}}} s(\tilde{\mathbf{x}}), \tag{18}$$

where $F_{\mathbf{x}}(\mathbf{x})$ denotes the distribution function of the covariates and $s(\tilde{\mathbf{x}})$ denotes the share of individuals in the population (both of men and women) who display the covariate combination $\tilde{\mathbf{x}}$ in an adequately discretised covariate space ($\tilde{\tilde{\Xi}}$).

In our application, we use data from the German *Mikrozensus* (see Section A.1.3 in the appendix) to obtain the shares in our discretised covariate space. We consider age as a finely discretised variable on a yearly basis and the other variables in accordance with the categorisation discussed in Section 3. See Section A.1.3 for further elaboration on the covariate distribution.

The discretised covariate space yields 20,480 covariate combination for which we estimate the conditional wage distribution for both men and women. From these we can derive discrimination measures, such as the ratios presented in Table 2. The ratios can then be aggregated in accordance with Equation (18) yielding an average ratio presented below:

	μ	CE(ϵ=0.5)	CE(ϵ=1.0)	CE(ϵ=1.5)	μ_0^{conv}	μ_0
Δ_a	1.75[1.58;1.94]	2.11[1.87;2.46]	3.22[2.73;4.08]	4.15[3.40;5.10]	1.13[1.07;1.20]	1.52[1.40;1.66]

Table 3: Average ratios of male and female wage distribution measures

As can be observed, inequality in the conditional wage distributions considerably changes the magnitude of the difference between male and female wages, once we use certainty equivalents. The analysis on the basis of raw expected wages shows that males earn on average around 75% more than women per hour of work. This figure rises to over 300% once we use certainty equivalents with an inequality aversion parameter of $\epsilon = 1.5$. By contrast, if we use conventional definitions, excluding zero incomes altogether, we would get considerably lower aggregate discrimination estimates. In other words, the consideration of inequality points to a much larger discrepancy between men and women than conventional analysis would portray.

6 Conclusion

This article argued for the use of a more comprehensive conception of wage discrimination. Looking at gender related wage discrimination in Germany, we show possible inroads to such a comprehensive wage discrimination analysis. We find that, once one includes the whole population as the sampling universe (and not just those in employment), the distributional nature of wages (and not just point estimates) and an activity-based definition of work (rather than a pay-based definition), gender wage discrimination in Germany is much larger than is portrayed by conventional estimates. Our analysis thus highlights that in Germany much is left to be done to achieve a situation where a newborn faces equivalent wage prospects regardless of its gender.

References

Y. Amiel and F. A. Cowell (1994): Income Inequality and Social Welfare, in: J. Creedy (ed.), Taxation, Poverty, and Income Distribution, pp. 193–219, Edward Elgar Pub., Aldershot.

D. W. K. Andrews (1997): A Conditional Kolmogorov Test, in: Econometrica, 65(5), pp. 1097–1128.

R. Anker (1998): Gender and Jobs: Sex Segregation of Occupations in the World, International Labour Organisation, Geneva.

W. Arulampalam, A. L. Booth and M. L. Bryan (2007): Is There a Glass Ceiling over Europe?

Exploring the Gender Pay Gap across the Wage Distribution, in: Industrial and Labor Relations Review, 60(2), pp. 163–186.

A. B. Atkinson (1970): On the Measurement of Economic Inequality, in: Journal of Economic Theory, 2(3), pp. 244–263.

C. Belitz, A. Brezger, N. Klein, T. Kneib, S. Lang and N. Umlauf (2015): BayesX - Software for Bayesian Inference in Structured Additive Regression Models: Version 3.0, URL `http://www.bayesx.org`.

C. Bellemare, S. Kröger and A. van Soest (2008): Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities, in: Econometrica, 76(4), pp. 815–839.

M. Biewen and S. P. Jenkins (2005): A Framework for the Decomposition of Poverty Differences with an Application to Poverty Differences Between Countries, in: Empirical Economics, 30(2), pp. 331–358.

A. Brezger and S. Lang (2006): Generalized Structured Additive Regression Based on Bayesian P-splines, in: Computational Statistics & Data Analysis, 50(4), pp. 967–991.

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

M. Charles and D. B. Grusky (2004): Occupational Ghettos: The Worldwide Segregation of Women and Men, Stanford university press, Stanford.

V. Chernozhukov, I. Fernandez-Val and B. Melly (2013): Inference on Counterfactual Distributions, in: arXiv:, 0904.0951v6[stat.ME].

D. Chotikapanich (ed.) (2008): Modeling Income Distributions and Lorenz Curves, Springer, New York.

F. A. Cowell (2000): Measurement of Inequality, in: A. B. Atkinson and F. Bourguignon (eds.), Handbook of Income Distribution, pp. 87–166, Elsevier, Amsterdam.

— (2011): Measuring Inequality, Oxford University Press, Oxford, 3rd ed.

C. Dagum (1997): A New Approach to the Decomposition of the Gini Income Inequality Ratio, in: Empirical Economics, 22(4), pp. 515–531.

DIW Berlin (2014a): SOEP 2013 - Erhebungsinstrumente 2013 (Welle 30) des Sozio-oekonomischen Panels: Haushaltsfragebogen, in: SOEP Survey Papers, No. 181.

— (2014b): SOEP 2013 - Erhebungsinstrumente 2013 (Welle 30) des Sozio-oekonomischen Panels: Personenfragebogen, in: SOEP Survey Papers, No. 180.

P. J. Dolton and G. Makepeace (1985): The Statistical Measurement of Discrimination, in: Economic Letters, 18, pp. 391–395.

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

P. H. C. Eilers and B. D. Marx (1996): Flexible Smoothing with B-splines and Penalties, in: Statistical Science, 11(2), pp. 89–102.

P. England (2005): Emerging Theories of Care Work, in: Annual Review of Sociology, 31(1), pp. 381–399.

J. Esteban, C. Gradín and D. Ray (2007): Extensions of a Measure of Polarization, with an Application to the Income Distribution of five OECD Countries, in: Journal of Economic Inequality, 5(1), pp. 1–19.

J. Esteban and D. Ray (1994): On the Measurement of Polarization, in: Econometrica, 62(4), pp. 819–851.

L. Fahrmeir, T. Kneib, S. Lang and B. D. Marx (2013): Regression: Models, Methods and Applications, Springer, Berlin.

L. Flabbi (2010): Gender Discrimination Estimation in a Search Model with Matching and Bargaining, in: International Economic Review, 51(3), pp. 745–783.

N. M. Fortin, T. Lemieux and S. Firpo (2011): Decomposition Methods in Economics, in: O. Ashenfelter and D. E. Card (eds.), Handbook of Labor Economics, Vol. 4A, pp. 1–102, North-Holland, Amsterdam.

J. Fox (1987): Effect Displays for Generalized Linear Models, in: C. C. Clogg (ed.), Sociological Methodology, vol. 17, pp. 347–361, American Sociological Association, Washington D.C.

G.-L. Gayle and L. Golan (2012): Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968-1997, in: The Review of Economic Studies, 79(1), pp. 227–267.

M. Gimbutas (2007): The Goddesses and Gods of Old Europe: Myths and cult images, University of California Press, Berkeley.

C. Gradín (2000): Polarization by Sub-populations in Spain, 1973-91, in: Review of Income and Wealth, 46(4), pp. 457–474.

V. Guillin (2009): Auguste Comte and John Stuart Mill on Sexual Equality: Historical, Methodological and Philosophical Issues, Brill, Leiden and Boston.

J. J. Heckman (1979): Sample Selection Bias as a Specification Error, in: Econometrica, 47(1), p. 153.

J. L. Hook and B. Pettit (2015): Reproducing Occupational Inequality: Motherhood and Occupational Segregation, in: Social Politics: International Studies in Gender, State & Society.

T. Hothorn, T. Kneib and P. Bühlmann (2014): Conditional Transformation Models, in: Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(1), pp. 3–27.

S. P. Jenkins (1994): Earnings Discrimination Measurement: A Distributional Approach, in: Journal of Econometrics, 61(1), pp. 81–102.

D. Kahneman (2011): Thinking, Fast and Slow, Farrar, Straus and Giroux, New York.

C. Kleiber and S. Kotz (2003): Statistical Size Distributions in Economics and Actuarial Sciences, Wiley, Hoboken.

N. Klein, M. Denuit, S. Lang and T. Kneib (2014): Nonlife Ratemaking and Risk Management with Bayesian Generalized Additive Models for Location, Scale, and Shape, in: Insurance: Mathematics and Economics, 55, pp. 225–249.

N. Klein, T. Kneib, S. Lang and A. Sohn (2015): Bayesian Structured Additive Distributional Regression with an Application to Regional Income Inequality in Germany, in: Annals of Applied Statistics, 9(2), pp. 1024–1052.

S.-C. Kolm (1969): The Optimal Production of Social Justice, in: J. Margolis and H. Guitton (eds.), Public Economics, pp. 145–200, Macmillan, London.

T. Lemieux (2006): The "Mincer Equation". Thirty years after Schooling, Experience, and Earnings, in: S. Grossbard (ed.), Jacob Mincer. A Pioneer of Modern Labor Economics, pp. 127–145, Kluwer Academic, Boston.

K. Liu (2015): Explaining the Gender Wage Gap: Estimates from a Dynamic Model of Job Changes and Hours Changes, in: IZA Discussion paper series, No. 9255.

J. Machado and J. Mata (2005): Counterfactual Decompostion of Changes in Wage Distributions Using Quantile Regression, in: Journal in Applied Econometrics, 20(4), pp. 445–465.

J. Morduch and T. Sicular (2002): Rethinking Inequality Decomposition, with Evidence from Rural China, in: The Economic Journal, 112(476), pp. 93–106.

D. A. Neal and W. R. Johnson (1996): The Role of Premarket Factors in Black-White Wage Differentials, in: Journal of Political Economy, Vol. 104(5), pp. 869–895.

J. A. Nelder and R. W. M. Wedderburn (1972): Generalized Linear Models, in: Journal of the Royal Statistical Society: Series A (General), 135(3), pp. 370–384.

R. L. Oaxaca and M. R. Ransom (1999): Identification in Detailed Wage Decompositions, in: Review of Economics and Statistics, 81(1), pp. 154–157.

M. Reid (1934): Economics of Household Production, John Wiley, New York.

C. L. Ridgeway (1997): Interaction and the Conservation of Gender Inequality: Considering Employment, in: American Sociological Review, 62(2), pp. 218–235.

J. E. Roemer (1998): Equality of Opportunity, Harvard University Press, Cambridge.

C. Rothe and D. Wied (2013): Misspecification Testing in a Class of Conditional Distributional Models, in: Journal of the American Statistical Association, 108(501), pp. 314–324.

B. Ryan (1992): Feminism and the Women's Movement: Dynamics of Change in Social Movement, Ideology, and Activism, Perspectives on gender, Routledge, New York.

R. Selten and G. Gigerenzer (2002): Bounded Rationality: The Adaptive Toolbook, MIT Press, Cambridge (Mass.) and London.

A. Sen (1973): On Economic Inequality, Oxford University Press, Oxford.

SOEP (2014): Socio-oeconomic Panel Study (SOEP), data of the years 1984-2013, version 30, doi: 10.5684/soep.v30.

A. Sohn, N. Klein and T. Kneib (2014): A Semiparametric Analysis of Conditional Income Distributions, in: Schmollers Jahrbuch, 134, pp. 1–10.

D. M. Stasinopoulos and R. A. Rigby (2007): Generalized Additive Models for Location, Scale and Shape (GAMLSS) in R, in: Journal of Statistical Software, 23(7), pp. 1–46.

Statistisches Bundesamt (2013a): Berufsausbildung auf einen Blick, Statistisches Bundesamt, Wiesbaden.

— (2013b): Hochschulen auf einen Blick, Statistisches Bundesamt, Wiesbaden.

P. van Kerm (2013): Generalized Measures of Wage Differentials, in: Empirical Economics, 45(1), pp. 465–482.

M. C. Wolfson (1994): When Inequalities Diverge, in: American Economic Review, 84(2), pp. 353–358.

A Appendix

A.1 Data

As primary source for our data we use the SOEP database (SOEP, 2014). We use all available waves in 2013, which provides us with 10 waves (A to J). Only taking those values for which we have the full set-of variables, as described below, we obtain 9,182 observations.

A.1.1 Variables used

In order to obtain our dependent variable of the gross hourly wage we use the variable `BDP7701` from the individual questionnaire (DIW Berlin, 2014a) to obtain the monthly earnings of the individual.[11] This is divided by 4.32 times the number of hours worked per week. As discussed in Section 2.3, we do not only consider the hours of paid work but the hours spent on activities which we consider work. For the former we use the the variable `BDP66` , which is the actual number of hours worked last week. The hours spent on housework, errands, care work and repair works are taken from the answers from `BDP10$$`. These variables give the number of hours spent on each type of activity for a regular weekday, a Saturday and a Sunday, which we simply sum up to get the number of hours per week. It must be noted that these variables are likely to be rather imprecise such that our data is affected by some considerable data related uncertainty.

For the explanatory variable age we simply use the birthyear (`GEBJAHR`) from the individual questionnaire and subtract it from 2013, while the sex is determined by the variable `BDSEX`. The binary variable of whether a person has children that is still dependent on the household is based on the variable `BDH503` from the household questionnaire (DIW Berlin, 2014b). The variable *nat* is constructed on grounds of variable `BDP143` in the individual questionnaire (DIW Berlin, 2014a).

The education level is taken on grounds of the variable `ISCED13`[12] from the PGEN variables. All observations equal or lower than 2 (general elemantary and below) are put in the first education category, with all observations in category 3 (middle vocational) put in the second category. All observations with `ISCED13` values 4 and 5 (vocational with Abitur and higher vocational) are in the third group, while the highest group consists only of individuals with a value 6 (higher education).

For the spatial effect we use the variable `BDBULA` with the variable *east* set to unity for all federal

---

[11]We thus neglect aspects any transfer payments, like the *Elterngeld*. While this is obviously problematic in some cases it seems much more appropriate than the consideration of net-incomes, i.e. including all transfer payments (like *Elterngeld*), as we want to consider the remuneration of activities by a society prior to welfare considerations as is standard in the discrimination literature.

[12]It should be noted that the SOEP ISCED levels are not a 100% equivalent to the ISCED levels elsewhere.

states formerly belonging to the German Democratic Republic, including the whole of Berlin, which in the SOEP is equivalent to states 1-11. West Berlin is not accounted for separately.

All observations are weighted using the variable `BDPHRF`.

A.1.2   Observations dropped

Concerning the wage rate, we only consider those persons who for who we have a value greater or equal to zero, i.e. all observations that are not missing, not applicable or highly improbable according to the SOEP.

Concerning the age, we only consider person who are between 21 and 60 years of age and for whom we have age observations.[13]

Concerning education, we all observations who have a value greater or equal to zero, i.e. all observations that are not missing, not applicable or highly improbable according to the SOEP.

Concerning time spent, we drop observations who state more than 168 hours for activities per week.

Concerning the federal state of residence, we use all observations.

Concerning the nationality, we use all observations.

Concerning the children, we use all observations.

Concerning the weight, we use all observations with a positive weight.

A.1.3   Construction of covariate distribution

For the construction of the covariate distribution in the discretised covariate space, we use data from the German *Mikrozensus* provided upon request from the German Statistical Office. The *Mikrozensus* entails information from around 370,000 household with 830,000 persons equivalent to roughly 1% of the population. It thus allows for a relatively accurate account of the number of persons in each strata as defined by our covariates.

We use the variables in the *Mikrozensus* to categorise our population strata emulating the categorisation based on the information (and variables) from the SOEP as closely as possible. To this end, we proceed as follows:

---

[13]We explicitly do not exclude students or other persons currently in the process of enhancing their capital stock to further incomes later on. As long as such education breaks from employment are roughly equivalently distributed between males and females, which they are in the case of tertiary education (Statistisches Bundesamt, 2013a), this should not distort the discrimination estimates as we define it.

For age we have yearly age brackets at our disposal allowing us a one-to-one comparison with the variables from the SOEP. The same is possible for the categorisation into Germans and persons of foreign nationality which is also recorded in the *Mikrozensus* equivalently to the SOEP. Concerning the question of whether the individual has dependent children, we use information from the *Mikrozensus* on whether the individual has under age children. This is not 100% equivalent to the definition we use based on the SOEP, but should by and large give us similar results. Lastly for education we use information on the education level with regard to job qualification (*höchster beruflicher Ausbildungsabschluss*) available. The four education levels we use are constructed as follows:

$educ_1$ The first education level comprises all individuals who have no formal job-qualification (*ohne berufsqualifizierenden Ausbildungsabschluss*) and those on whom we have no information (*ohne Angabe zur Art des Abschlusses*).

$educ_2$ The second education level comprises all individuals who have the standard job qualification, which in Germany is generally a completed apprenticeship (*Abschluss einer Lehre oder gleichwertiger Berufsfachabschluss*).

$educ_3$ The third education level comprises all individuals who have a higher technical technical qualification (*Meister-/Techniker oder geichwertiger Fachschul-abschluss* and *Fachschulabschluss der DDR*)

$educ_4$ The fourth education level comprises all individuals with higher education (*Fachhochschulabschluss* and *Hochschulabschluss*).

Using this categorisation, we construct the covariate distribution yielding population shares for all covariate combinations in $\tilde{\tilde{\Xi}}$ as used in Equation 18.

A.2 A remark on the inclusion of those out of formal work

In this paper, we consider the whole population with the wage rate that is actually earned, including zero wage rates. This is converse to the standard approach in the literature which concentrates on those earning a positive wage and includes those with zero wages along the lines of the seminal paper by Heckman (1979) whereby labour market participation is seen as a sample selection problem.

The underlying economic concept of the Heckman approach is that individuals choose to work or not to work based on some (presumably utility-based) criteria. In order to account for this self-selection Heckman proposes a correction which would emulate the situation whereby those persons who choose not to participate are paid the wage they would have been paid had they chosen to work for the offered wage.

Yet for our assessment of the magnitude of discrimination this approach would lead a skewed assessment, as it only corrects for (some) self-selection of those who fully opt out of working for money. By contrast, those who choose (or are forced by circumstances) to work-part time (and most likely accept a lower wage) would be included in no other way than those working full-time. Equally those who choose one more lowly paid job over another for non-pecuniary reasons would also induce a problem of selection. Or to put it in the framework presented in Section 2.1, the problem is that the Heckman correction only selectively accounts for individual's choices and circumstances in Equation 2.

As mentioned before, it remains a simple (sad) fact that economic regressions and the empirical data available are far from allowing us to get anywhere near to identifying individual's choices and societal circumstances. Thus, rather than applying a counter-factual correction to a subset of the population we opt for what we consider the more consistent approach - consider the actual wages paid.

A.3 Estimation methodology

A.3.1 GAMLSS and SADR

For our estimation we use structured additive distributional regression in order to estimate conditional distributions. It should be noted that this methodology is closely related to the concept of Generalised Additive Models of Location, Scale and Shape (GAMLSS) first proposed by Stasinopoulos and Rigby (2007) in a frequentist framework and by Klein et al. (2014) in a Bayesian setting. Yet to the frequentist framework and the associated R package `gamlss` is still very limited with regard to the flexibility of the effects. Especially the ability of fitting non-linear effects in a reliable and stable manner which are well suited to modelling the age effect in a flexible manner constitutes a major advantage of structured additive distributional regression and its implementation in BayesX. Following Klein et al. (2015) we use the phrase structured additive distributional regression as the aim of our estimation methodology is not so much simply extending the number of distributional aspects, like the moments, beyond just one particular aspect, like the mean. Rather the methodology's primary focus is to model the whole conditional distribution.

Pursuing the same aim along a different path several publications have recently appeared which could in principle be used to estimate the conditional earnings distributions.

A.3.2 Direct Maximum Likelihood Estimation

In his paper, van Kerm (2013) proposes a simple maximum likelihood based estimation procedure employing a Singh-Maddala distribution with each of the three parameters allowed to vary along a coarsely discretised covariate space. Yet this approach generally becomes infeasible as the number of covariate combinations on the covariate space becomes larger. Given the large number conceivably influential covariates (and their possibly continuous nature) a regression based approach like ours seems more appropriate than a direct estimation of the distributions for many subsets of the sample.

A.3.3 Distributional Regression and Conditional Transformation Models

Another approach, which has recently received considerable attention is the proposition to estimate conditional distributions by discretising its domain into small ranges and estimating the corresponding density functions of those domains using the well established statistical machinery from survival analysis. In the econometric literature this is generally called distributional regression (Chernozhukov et al., 2013) while in the statistical literature refers to such approaches as conditional transformation models (Hothorn et al., 2014).

While this approach allows for great flexibility, this comes at the cost of high instability of the estimation. Given the rather limited number of observations with respect to the size of the covariate space this alternative distribution methodology is not applicable to our research question.

A.3.4 Quantile Regression

An obvious alternative to estimating conditional distributions is by the use of quantile regression, which has been used for the estimation of conditional income distributions by Machado and Mata (2005) among others. Again this methodology is arguably more flexible than our approach as it does not require a possibly erroneous assumption about the parametric nature of the distribution. However, again this flexibility comes at the cost of estimation stability which given the limited sample size does constitute a problem. Another particular problem with quantile regression is the fact that due to the inclusion of unemployed persons, we have a high point mass which is known to cause problems with quantile crossing, especially if the effects are simply linear (see Klein et al., 2015).

While there are alternative estimation strategies, we thus opt for structured additive distributional regression as this methodology yields a good balance between modelling flexibility and estimation stability due to its parsimonious parametric formulation. We explicitly want to stress that our

approach, while having some advantages over the other approaches presented here, naturally also has disadvantages. It is up to the specific problem at hand to choose the appropriate methodology. Further studies comparing the performances of the different approaches would greatly aide the choice but are to our knowledge not yet available.

A.4 Effects on the parameters

In Table 4 we display the estimates for the parameters of the three conditional wage distribution considered in Section 5.1.2.

	D_{1m}	D_{1f}	D_{2m}	D_{2f}	D_{3m}	D_{3f}
a	8.31[5.68;12.40]	7.66[5.24;11.32]	6.19[5.16; 7.49]	4.80[4.08; 5.67]	4.57[3.76; 5.58]	4.02[3.31; 4.94]
b	11.27[9.44;13.63]	9.06[7.40;11.14]	16.35[14.57;18.26]	13.69[12.22;15.27]	30.11[26.33;34.22]	21.88[18.81;25.21]
c	0.21[0.11; 0.40]	0.31[0.17; 0.59]	0.78[0.60; 1.05]	0.59[0.45; 0.78]	0.68[0.50; 0.92]	0.57[0.40; 0.82]
π_0	0.43[0.36; 0.51]	0.52[0.41; 0.62]	0.08[0.06; 0.10]	0.20[0.15; 0.26]	0.01[0.01; 0.02]	0.07[0.05; 0.09]

Table 4: Distribution parameters for 3 conditional wage distributions

As one can observe there is substantial variation between the parameters beyond the parameter b, which is a scale parameter and thus most comparable to the parameter μ in case of the log-normal distribution. This shows that there is considerable variation of additional aspects of the conditional distributions in the covariate space which conventional analysis is liable to neglect.

A.5 Effects on the wage for other covariates

In Figure 3, we display the resultant estimates from our model for different combinations of age and two education levels (the second and the fourth). The other covariates are fixed as follows: German Nationality, with at least one child and living in West Germany.

The upper graph showing the expected wage, shows the usual pattern of concavely increasing wages with age, while higher education also increases the expected wage with women generally paid less, c.p. The same is true for the certainty equivalent wages (with $\alpha = 1$) displayed in the graph below, although this shows that the differences between men and women and education levels are more pronounced. Additionally, it must be noted that the estimates for the inequality are slightly unstable which can be ascribed to the relatively small sample size and the fact that we only apply the smoothing over the parameters individually rather than jointly. Further work on multidimensional smoothing is this type of application would improve the estimation.
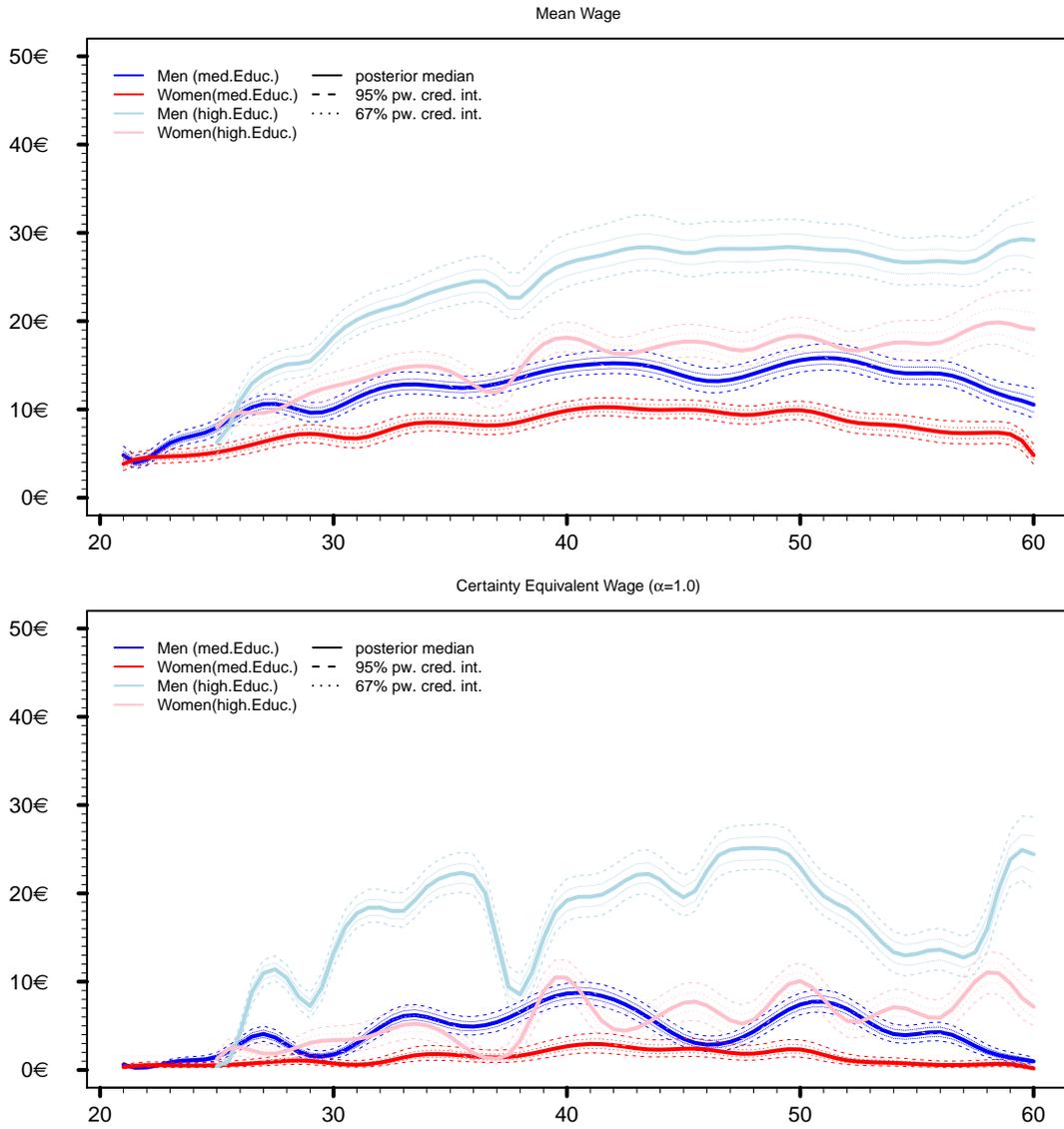
Figure 3: Expected Wage (upper) and Certainty Equivalent Wage (lower) w.r.t. age and education.

A.6 Estimates for the conventional wage concept

A.6.1 Visualising the two working hours concepts

Before we consider the estimates for the conventional wage concept, let us consider the differences in terms of working hours.

Figure 4 displays the working hours under the conventional concept only considering work under formal employment and the comprehensive concept following the concept of the "third party criterion". As would be expected working hours increase dramatically. More importantly one can also observe though that the perceived higher workload for men that one may deduce from the

Conventional weekly working hours
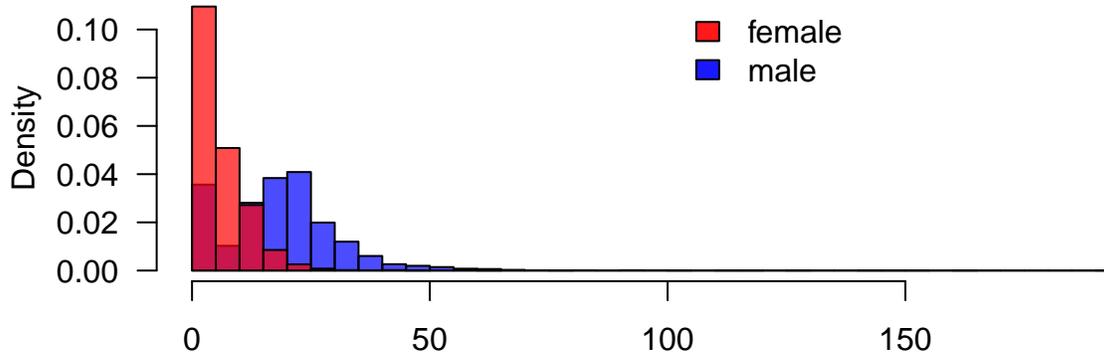
Comprehensive weekly working hours
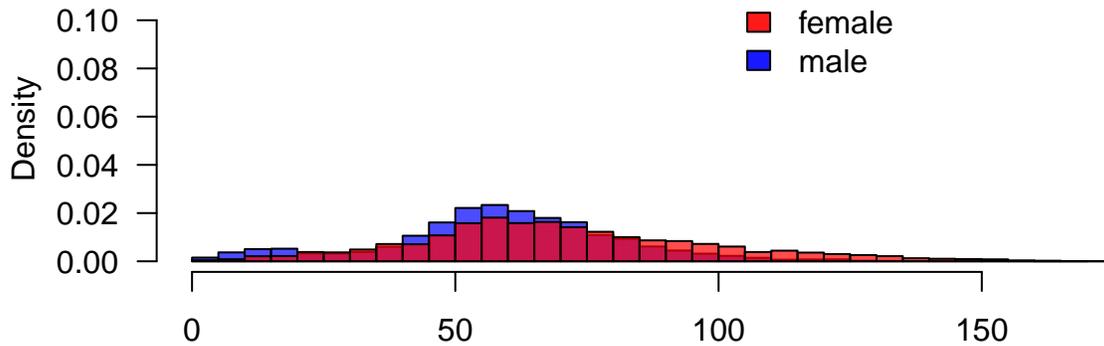
Figure 4: Two histograms of weekly working hours.

first histogram disappears (and is even slightly reversed) when one considers all work activities.

A.6.2 Analysis of Conditional Wage Distributions

As in Section 5, we consider the wage distributions of the same three covariate combinations by visual inspection. In contrast to the graphs in Figure 2, we can observe that the wage distributions of men and women overlap to a much greater degree, if one only considers the ordinary work-time concept. Naturally, the wages are also a lot higher as the working time according to the introduced concept is a superset of the conventional definition of working time.

For the first graph, we see that the graphs are again not adhering to the standard positive skew
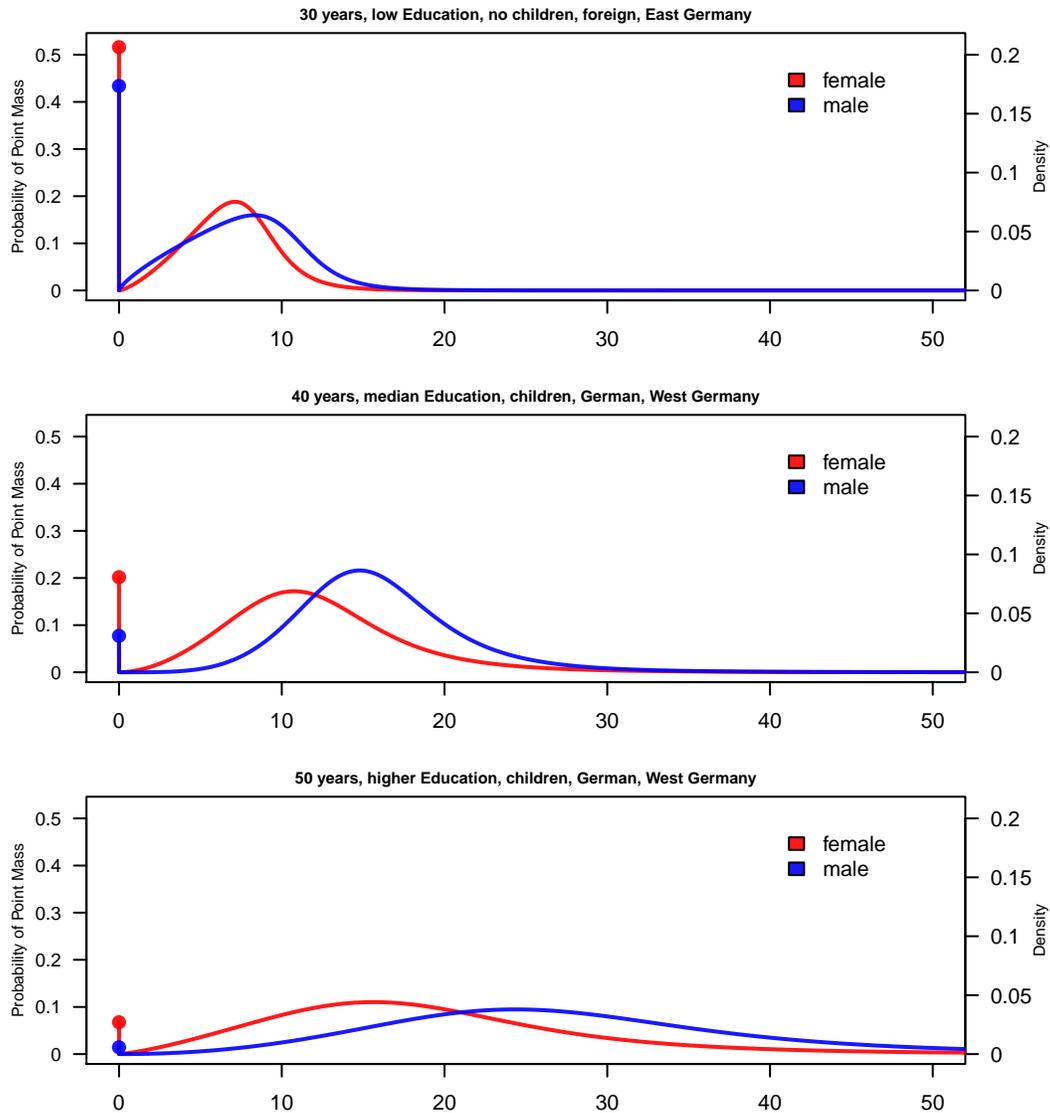
Figure 5: Distribution estimates of males (blue) and females (red).

normally found in most income related distributions. However, the skew is lower than in the case in Figure 5. Nonetheless, also with these graphs it is observable that there are differences beyond location and scale between the wage distribution of men and women.

This is also true for the second wage distribution, where women show a higher skewness as well as a higher coefficient of variation. Additionally we can observe that women are much more likely to be receiving no wages at all than men.

The third set of distributions displays a much wider distribution than originally. Especially for women, we observe a much lower skewness, as for a lot of the women in this subgroup the working time is significantly reduced if aspects like care work are not considered.

In Table 5 we display the equivalent measures to those displayed in Table 1 in the article.

	D_{1m}	D_{1f}	D_{2m}	D_{2f}	D_{3m}	D_{3f}
μ	4.33[3.52;5.22]	3.35[2.55;4.31]	14.82[13.67;16.15]	9.87[8.92;10.89]	28.33[25.83;31.51]	18.33[16.59;20.44]
G	0.26[0.20;0.32]	0.22[0.17;0.27]	0.18[0.15; 0.20]	0.25[0.22; 0.28]	0.27[0.22; 0.36]	0.32[0.27; 0.40]
A(ϵ=0.5)	0.44[0.38;0.50]	0.50[0.40;0.59]	0.10[0.08; 0.12]	0.23[0.19; 0.29]	0.06[0.05; 0.09]	0.14[0.11; 0.17]
A(ϵ=1.0)	0.91[0.87;0.94]	0.93[0.89;0.96]	0.42[0.36; 0.48]	0.73[0.64; 0.81]	0.19[0.15; 0.23]	0.45[0.37; 0.54]
A(ϵ=1.5)	0.99[0.99;0.99]	0.99[0.99;0.99]	0.93[0.91; 0.95]	0.98[0.97; 0.99]	0.70[0.62; 0.78]	0.93[0.89; 0.96]
CE(ϵ=0.5)	2.44[1.79;3.17]	1.68[1.05;2.52]	13.38[12.29;14.63]	7.58[6.37; 8.72]	26.49[24.01;29.44]	15.82[14.18;17.76]
CE(ϵ=1.0)	0.40[0.25;0.63]	0.22[0.11;0.47]	8.66[7.42; 9.81]	2.68[1.75; 3.78]	22.93[20.61;25.61]	10.07[8.23;12.14]
CE(ϵ=1.5)	0.05[0.04;0.07]	0.03[0.02;0.05]	0.97[0.70; 1.33]	0.19[0.12; 0.31]	8.39[6.22;10.84]	1.18[0.72; 1.92]
P(α=β=1)	0.06[-0.22;0.42]		1.11[0.72;1.55]		2.30[1.43;3.26]	

Table 5: Some distribution measures for 3 conditional wage distributions

The Table also shows that the expected wages are much higher both for men and women with the discrepancy between wages of men and women generally reduced in contrast to the comprehensive wage definition used in the article. For the first group, it may even be noted that the credible intervals for males and females overlap, although this is more likely down to the relatively high uncertainty of the estimates in structured additive distributional regression than an actual equivalence in the expected wages.

Concerning the Gini coefficient, we find generally find slightly reduced measures, although this trend is not true for all six distributions and mostly non-significant. The same applies for the Atkinson indices.

For the certainty equivalents, we see that they are naturally also higher. The discrepancy has generally decreased, as we will discuss in a bit more detail below.

Lastly for the polarisation, we find that this has decreased for the first group and less so for the second, while for the third group the polarisation measure increases slightly, with all these variations non-significant though. One aspect which strikes the eye, is that the credible intervals become negative which, as Gradín (2000) points out can happen, due to the way that the distributions are conceptualised as spikes with an error correction account for the within group variation.

	$D_{1.}$	$D_{2.}$	$D_{3.}$
μ	1.29[0.94; 1.80]	1.50[1.32; 1.72]	1.54[1.33; 1.79]
CE(ϵ=0.5)	1.45[0.87; 2.47]	1.77[1.49; 2.13]	1.67[1.44; 1.95]
CE(ϵ=1.0)	1.75[0.74; 3.99]	3.22[2.20; 5.10]	2.28[1.83; 2.89]
CE(ϵ=1.5)	1.38[0.82; 2.19]	5.03[2.85; 8.84]	7.00[4.02;12.52]

Table 6: Ratios of male and female wage distribution measures

To get a measure of discrimination, we use the ratios between male and female wage measures, as displayed in Table 6. As would be expected the measures decrease relative to the measures we

find in Table 2 using the comprehensive wage definition. Nonetheless, we still the same general relationship that the certainty equivalents portray a higher degree of discrimination than would be portrayed by simple comparison of expected incomes.

Using the conventional concept, we also show the results if we aggregate the various subgroups:

	μ	CE(ϵ=0.5)	CE(ϵ=1.0)	CE(ϵ=1.5)
Δ_a	1.60	2.04	4.65	8.06

Table 7: Average ratios of male and female wage distribution measures

Analogously to above, we find that although the magnitude is smaller, the general finding that the inclusion of risk in the assessment of gender discrimination generally increases the magnitude estimates is robust to the definition of work.

A.7 Misspecification Testing

A.7.1 Use of MCMC samples for the misspecification test

To implement the test, we use 10,000 bootstrap draws from the algorithm discussed in Section 4.3. For our wage distributions have two independent MCMC samples, one for π_0 and one for a, b and c. Hence, we have a two dimensional sample set $\mathcal{M} = (1, \ldots, M_1) \times (1, \ldots, M_2)$, with $M_1 = M_2 = 1000$. A given random bootstrap sample b thus takes a random set of realisations from the MCMC output $\{m_1; m_2\} = \mathbf{m} \in \mathcal{M}$ to yield $\pi_b^{m_1}$, $a_b^{m_2}$, $b_b^{m_2}$ and $c_b^{m_2}$.

A.7.2 Simulation study

We briefly assess the misspecification test from Section 4.3 by means of three simple simulation studies in order to validate its performance. In each simulation study, we consider a simple framework with one explanatory variable which has a linear effect on all the predictors of the Dagum distribution, i.e.

$$g(\eta^{\theta_k}) = \beta_0^{\theta_k} + \beta_1^{\theta_k} x, \tag{19}$$

where x is an integer from the interval $[1, 10]$ and g is the log-link.

For the simulations we use 1,000 observations and 1000 bootstrap repetitions and contrast the results for a true specification, as specified above, with a misspecified parametric model. For the misspecification we use a log-normal distribution with mean and coefficient of variation equivalent to that of the Dagum specification.

| | Sim. Study 1 | | Sim. Study 2 | | Sim. Study 3 | |
Sim.Run	$\mathcal{H}_0$ TRUE	$\mathcal{H}_0$FALSE	$\mathcal{H}_0$ TRUE	$\mathcal{H}_0$FALSE	$\mathcal{H}_0$ TRUE	$\mathcal{H}_0$FALSE
1	0.30	0.00	0.02	0.00	0.76	0.12
2	0.58	0.00	0.60	0.00	0.66	0.09
3	0.40	0.00	0.60	0.00	0.40	0.03
4	0.21	0.00	0.82	0.00	0.84	0.07
5	0.71	0.00	0.64	0.00	0.45	0.08
6	0.40	0.00	0.92	0.00	0.63	0.07
7	0.07	0.00	0.05	0.00	0.95	0.10
8	0.92	0.00	0.98	0.00	0.81	0.09
9	0.28	0.00	0.45	0.00	0.54	0.10
10	0.10	0.00	0.80	0.00	0.92	0.08
11	0.91	0.00	0.57	0.00	0.62	0.09
12	0.21	0.00	0.50	0.00	0.59	0.09
13	0.47	0.00	0.00	0.00	0.90	0.09
14	0.29	0.00	0.71	0.00	0.37	0.12
15	0.66	0.00	0.15	0.00	0.72	0.10
16	0.15	0.00	0.13	0.00	0.48	0.09
17	0.92	0.00	0.15	0.00	0.48	0.10
18	0.73	0.00	0.23	0.00	0.92	0.12
19	0.49	0.00	0.41	0.00	0.46	0.10
20	0.98	0.00	0.21	0.00	0.67	0.09
21	0.06	0.00	0.31	0.00	0.80	0.09
22	0.29	0.00	0.68	0.00	0.82	0.10
23	0.26	0.00	0.73	0.00	0.84	0.11
24	0.49	0.00	0.82	0.00	0.97	0.10
25	0.39	0.00	0.57	0.00	0.57	0.11
26	0.79	0.00	0.40	0.00	0.60	0.11
27	0.91	0.00	0.06	0.00	0.80	0.10
28	0.20	0.00	0.54	0.00	0.44	0.10
29	0.35	0.00	0.92	0.00	0.56	0.09
30	0.43	0.00	0.05	0.00	0.42	0.16
μ	0.47	0.00	0.47	0.00	0.67	0.10
σ	0.28	0.00	0.31	0.00	0.18	0.02
Q_1	0.26	0.00	0.17	0.00	0.50	0.09
Q_2	0.40	0.00	0.52	0.00	0.64	0.10
Q_3	0.70	0.00	0.70	0.00	0.82	0.10

Table 8: Results from Simulation Studies for Misspecificaiton Test

The result from our simulation studies are displayed in Table 8. We the p-value for each simulation run as well as the mean of the p-values of all simulation runs (μ), their standard deviation (σ) as well as the three quartiles (Q_1, Q_2, Q_3).

In the first two columns, we display the results for the a scenario where we have negligible parameter uncertainty such that the standard deviation of the posterior distribution is 1% of its expectation. As can be observed the p-values in the first column roughly follow a uniform distribution, as we would expect, while the second column repeatedly rejects the null.

In the third and fourth column, we display a scenario with moderate parameter uncertainty such that the standard deviation of the posterior distribution is 5% of its expectation. As for the first

two columns we see that the results are able to clearly distinguish between the correct and the false specification.

In the last two columns, we assume considerable parameter uncertainty with the standard deviation of the posterior distribution is 50% of its expectation. Given this large uncertainty, the model specification test is less likely to reject the false hypothesis. With higher parameter uncertainty, the test is thus conservative.

Overall, the simulation study indicates that the test generally works although its power is mitigated by large parameter uncertainty.

A.8   Model Selection

Subsequently to asking whether one parametric model fits at all let us look at a handful of parametric models and consider their fit. Following our estimation approach we will consider the modelling of the point mass for zero wages and positive wages separately. Here we will consider only one generic predictor for all parameters. While it would in theory be possible to specify the predictor for each parameter individually, this would yield a high dimensional model selection problem, which given the lack of automated routines is not feasible.

A.8.1   Variable selection for the logit model

For the logit model we simply consider three alternatives:

M_1^0 As the first model, we use the model used in the paper.

M_2^0 For comparison we will consider the model, using a linear approach to modelling the impact of age, i.e. use a parabola $(age + age^2)$, with the other covariates specified as before.

M_3^0 Lastly, we will consider a model which only considers two education levels only differentiating between those with and without university education.

For comparison of the models we will use the DIC, which has been shown to be an adequate model selection criterion in Klein et al. (2015). The results are displayed in Table 9.

	males	females
M_1^0	301850	413997
M_2^0	337540	445537
M_3^0	336439	451336

Table 9: DIC for zero wages model approaches

For the modelling of income we consider the the following five models:

M_1^+ As the first model, we use the model used in the paper.

M_2^+ For comparison we will consider a model, using a linear approach to modelling the impact of age as above for all three parameters of the Dagum distribution.

M_3^+ Thirdly, we will consider a model, using only two education levels as above for all three parameters of the Dagum distribution.

In contrast to the Dagum distribution we will also consider the log-normal distribution.

M_4^+ As a first log-normal specification, we will use the same specification as above for both parameters of the log-normal.

M_5^+ As a first log-normal specification, we will use the same specification as above for μ, while considering σ as a nuisance parameter and keeping it constant, as is standard in the conventional literature.

The results are displayed in Table 10.

	males	females
M_1^+	22316.8	20687.2
M_2^+	23632.2	21864.9
M_3^+	23723.4	22946.8
M_4^+	24418.4	20931.5
M_5^+	24474.4	21239.3

Table 10: DIC for zero wages model approaches