# A New Semiparametric Approach to Analysing Conditional Income Distributions

Alexander Sohn*[1], Nadja Klein[1] and Thomas Kneib[1]

[1]Chair of Statistics, University of Göttingen

17th July 2014

**Abstract**

In this paper we explore the application of structured additive distributional regression for the analysis of conditional income distributions in Germany following the reunification. Using a bootstrapped Kolmogorov-Smirnov test we find that conditional personal income distributions can generally be modelled using the three parameter Dagum distribution. Additionally our results hint at an even more pronounced effect of skill-biased technological change than can be observed by standard mean regression.

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

## 1 Introduction

Following the publication of the fourth report on poverty and wealth (BfAS, 2013) a public debate on the extent and the nature of income inequality in Germany (re)erupted. While the dust hasn't fully settled yet, there is an increasing consensus that with some delay Germany is partially catching up with what Paul Krugman (2007) called the Great Divergence of incomes. While numerous studies in the economic literature have been dedicated to exploring the issue, we have to admit that "we know relative little about the determinants of residual inequality" (Acemoglu, 2002).

It is the aim of this paper to put residual inequality, or rather of conditional income distributions (CIDs), in the centre of the analysis. Specifically, we investigate conditional labour income distributions of men in Germany dependent on age, region and education for the years 1992 and 2010. We thus attempt to go beyond the analysis of single point measures of the distribution,

*Corresponding author: asohn@uni-goettingen.de

We thank the DIW for the data and their friendly support.

i.e. mean, median, etc. For this purpose we introduce structured additive distributional regression (Klein et al., 2014) and explore whether it has the scope to aide the analysis of income distributions. Specifically, we investigate whether CIDs can be modelled using the Dagum distribution and compare this approach to the standard modelling of CIDs using the log-normal distribution.

The structure of this paper is as follows: In the next two sections, we introduce the parametric modelling approach for CIDs with a special focus on the estimation of Dagum distributions and log-normal distributions and outline the methodology of structured additive distributional regression. Using the latter we provide estimates for the conditional income distributions for males with respect to the three explanatory variables age, educational attainment and region. In section 4, we employ a bootstrapped Kolmogorov-Smirnov test to check whether the Dagum distribution and/or the log-normal distribution provide an adequate fit for the analysis of our CIDs. Subsequently, we go on discussing some aspects of inference in a distributional regression framework in our application with a specific focus on the relation of skill-biased technological change to the variables age and region. In the last section we conclude.

# 2 Conditional income distributions

Using the data available in the German Socio-Economic Panel (SOEP) database[1], we consider the personal labour income as defined in the gross market income definition from Bach et al. (2009). Thereby our income definition entails wage income (including social security contributions) both from the private and the public sector as well business income from agriculture and forestry, unincorporated enterprise and self-employment. However, contrary to Bach et al. (2009) we exclude capital income. Our labour income definition thus entails practically all income derived from the factor labour. Consequently, we implicitly incorporate both changes in wage rates and changes in working time[2]. Since we aim to analyse the evolution of labour related income inequality at large, this seems the most appropriate definition to use. For more elaboration on the data see the appendix.

## 2.1 Parametric conditional income distributions

Following one of the most popular decomposition categories, namely decomposition by population groups, we will condition our income distributions on various demographic variables - namely

[1]For an elaboration of the SOEP database see Wagner et al. (2007), Wagner et al. (2008) and Bach et al. (2009).

[2]Working time is of high importance as there was a steep rise in unemployment as well as part-time and marginal part-time work in the period under consideration (Biewen and Juhasz, 2012).

region, education and age.[3] We consider region as a binary variable differentiating between the geographical region of the former Federal Republic of Germany and the former German Democratic Republic (entailing both former East and West Berlin).[4] Following Acemoglu (2002) we consider education as a binary variable as well which is unity for everybody who has obtained at least a university degree and zero otherwise. Lastly age is considered in a different manner. In the literature age has generally be considered by a finite number of groups. For example Dustmann and Schönberg (2009) split up their sample into three age groups for their decomposition.[5] Yet, Morduch and Sicular (2002) point out that age should more appropriately be considered as a continuous variable.[6] They also point to the problem that as categories or variables are added the number of distributions which need to be estimated increases in a multiplicative manner. Thus given the usual finite number of observations (in the order of thousands or tens of thousands) a direct estimation of the conditional distribution quickly becomes unstable. Consequently regularisation is required. This regularisation is achieved with the distributional regression approach which is discussed in more detail below. Inherent to the approach which we will pursue is the assumption that we can adequately model the CIDs by a parametric distribution. While we acknowledge that "the use of the parametric approach to distributional analysis runs counter to the general trend towards the pursuit of non-parametric methods, [...]" (Cowell, 2000, .145) we perceive the parametric approach as a form of regularisation itself which by imposing a structure lends stability to the estimation process. Moreover, we concur with Morduch and Sicular (2002, p.93) that it is often "necessary to impose more structure in order to draw sharp conclusion." And last but not least it should be noted that parametric models are better suited for robustness checks (see Silber, 1999, p.8). Naturally, the applicability of any parametric approach hinges on the "agreement between the model being identified and the actual observations" (Dagum, 1977). In other words it is critical to find a parametric model which is able to provide a sufficiently "good fit of the whole range of the distribution" (ibid.) for all the covariate sets of interest. The more diverse the sets of covariates under consideration are, the more the CIDs are likely to differ. With

[3]Mainly for reasons of comparability with the existent literature, which unfortunately has focused heavily on the distribution of male incomes/wages, we chose to exclude the conditional income distributions of females. Yet it should be noted that preliminary analyses have shown arguably more interesting dynamics for female incomes than for the male side.

[4]A model with a finer geographical resolution would possibly yield interesting new results and further work should be done on the improved incorporation of geographical information.

[5]For an overview, we provide histograms on the histograms for income distribution conditioned on these three age groups and the other three binary variables in the appendix.

[6]While the difference may become only of academic interest from a certain resolution onwards, the rather coarse differentiation with very few age groups disregards important developments within the age groups. Taking the first age interval of Dustmann and Schönberg (2009) as an example we would have no notion of the direct income distribution effects in the early twenties when vocational training typically ends or the mid/late twenties when students typically leave university. All these important labour market dynamics and the associated changes to the income distribution are ground to analytical dust by the coarse structure of the age categorisation.

increasingly diverse sets of covariates, more flexible distributions are thus likely to be required.[7]

A lot of ink and paper has been dedicated to the description of the aggregate income distribution of single countries in a parametric manner.[8] Borrowing from this literature we have tried various parametric distributions. One fundamental problem we encountered was the frequently bimodal nature of the CIDs already reported in various other contexts (see Cowell et al., 1996).This structure is due to the inclusion of the whole population in the specified age range irrespective of their employment status. To our knowledge this bimodal structure is not accounted for in any of the standard aggregate income distributions. To take account for this problem we applied a rough-and-ready trisection to our data, truncating all recipients of an income below 4,800€[9] from the main group, such that the parametric income distribution was only estimated for incomes above this level.[10] While clearly far from ideal, this artificial truncation is in line with much of the current literature which implicitly pursues this truncation by only considering full-time employees. The incomes below the level are then in turn divided up into zero and above-zero incomes. Using standard sequential logit estimation we determine the probabilities for a zero income, a low income (i.e. that the income is greater than zero but reaches no more than 4,800€), called precarious income from hereon, or whether the income falls into the income category of the truncated income distribution above 4800€.

Each CID thus takes the form of a mixture of two discrete probability masses and a continuous distribution:

$$f(y \mid \pi_0, \pi_{pr}, \theta_1, \ldots, \theta_K) = \mathbb{1}_{\{y=0\}}\pi_0 + \mathbb{1}_{\{0<y\leq 4800\}}\pi_{pr} + (1 - \pi_0 + \pi_{pr})t(y - 4800 \mid \theta_1, \ldots, \theta_K),$$

where $\mathbb{1}_{\{y=0\}}$ is an indicator function which takes unity if the income is zero, while $\mathbb{1}_{\{0<y\leq 4800\}}$ takes unity if the person receives a precarious income. The corresponding probabilities are $\pi_0$ and $\pi_{pr}$. $t(y - 4800 \mid \theta_1, \ldots, \theta_K)$ denotes the truncated conditional income distribution function which we assume to be parametric. To improve the fit of the two distributions and to evade problems with identification, we shifted the truncated income distribution to the right such that their support is restricted to the domain $(4800, \infty)$. Concerning the choice of the parametric distributions several distributions we follow the literature on statistical size distributions in economics (Kleiber and Kotz, 2003).[11] From the various distributions proposed, we have decided to concentrate on two parametric distributions. The log-normal distribution is a natural starting choice not least

[7]Naturally, more flexibility generally requires more parameters which may lead to estimation problems. This aspect is treated in more detail in the distributional regression section.

[8]Kleiber and Kotz (2003) as well as Chotikapanich (2008) provide an excellent overview.

[9]This figure was chosen on grounds of the so called "400€ -jobs" which falls under the category of minor (and consequently atypical) employment which is exempt from social security.

[10]The associated mathematical problems with such a rough-and-ready truncation are discussed in Dagum (1977). This emphasises that more elaborate methods are required in the future.

[11]Naturally other distributions are also conceivable. Yet further research is required on this front.

for its appealing theoretical properties and its parameters' interpretability. While at least for the aggregate distribution it is well known to have problems to fit the the upper bound of the distribution (see among others Atkinson, 1975), the log-normal distribution is by far still the most widely used distribution for conditional income distributions, especially when only used implicitly like for tobit regression (e.g. Card et al., 2013). Consequently, we consider the log-normal distribution as the benchmark distribution of the economic literature on conditional income distributions. Against this distribution, we contrast the Dagum distribution (Dagum, 1977). The Dagum distribution belongs to the beta-type distributions and is a member of the Pearson system (see Kleiber and Kotz, 2003). While it has three parameters which are not as readily interpretable as the two of the normal distribution, it is generally acknowledged that for aggregate income distributions it provides a much better fit than the log-normal distribution.[12]

# 3 Estimating conditional income distribution

For the estimation of the truncated CIDs we employ structured additive distributional regression as proposed by Klein et al. (2014).

Let the truncated CID be given by a parametric density $t(y \mid \theta_1, \ldots, \theta_K)$. For structured additive distributional regression we also assume that each of the K parameters $\theta_k$ can be expressed as an additive composition of the explanatory variables. More specifically we write:

$$g_k(\theta_k) = \eta_k = X_k\beta_k + \sum_{j=1}^{J_k} s_j(z_{j,k}), \tag{1}$$

where $g_k$ is a monotonic link function, $X_k$ is a design matrix containing the explanatory variables considered as linear effects and $z_{j,k}$ denotes the $j$-th covariate considered as non-linear effects. $\beta_k$ notifies the corresponding estimator for $x_k$ and $s_j(z_{j,k})$ is the smooth function of the $j-th$ continuous covariate considered in a non-linear way.

Applied to our study of CIDs, we thus aim to obtain predictors for every parameter of the two parametric truncated CIDs under consideration, the log-normal distribution and the Dagum distribution.

[12]It should be noted that next to the Dagum distribution, several other distributions have been proposed in the literature, like the three-parameter Singh-Maddala distribution (Singh and Maddala, 1976), the four-parameter Generalised beta distribution of Second Kind (McDonald, 1984) or the five-parameter Generalised beta distribution (McDonald and Ransom, 2008). The choice for the Dagum distribution was based on a preliminary study which is described in some more detail in Section C.3 of the appendix.

## 3.1 Log-normally distributed conditional income distributions

In the first setting we assume the truncated CIDs to follow a log normal distribution:

$$t(\tilde{y} \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma\tilde{y}} e^{-\frac{(\ln \tilde{y} - \mu)^2}{2\sigma^2}}, \tag{2}$$

with $\tilde{y} = y - 4800$. The predictors for the truncated conditional income distribution are estimated in a structured additive framework such that

$$\hat{\mu} = \eta_\mu = s_{1\mu}(age) + Hs_{2\mu}(age) + Os_{3\mu}(age) + HOs_{4\mu}(age), \tag{3}$$

$$\log(\hat{\sigma}) = \eta_\sigma = s_{1\sigma}(age) + Hs_{2\sigma}(age) + Os_{3\sigma}(age) + HOs_{4\sigma}(age), \tag{4}$$

where $s$ denotes a smooth function such that we model the effect of age in a non-linear way. We thereby follow the notion of Lemieux (2003) that the relation between earnings and experience (or age) is not linear. $H$ is a binary variable which is unity if we consider the CID for people with higher education. $O$ is also a binary variable which is unity if the CID is for people living in the Eastern part of Germany. For $\mu$ we employ a unity link function, while for $\sigma$ a log-link is employed to ensure positivity of $\sigma$.[13]

The main difference to standard mean regression is found in Equation 4. Note that not only do we estimate the conditional mean of the log-transformed normal distribution but also it's second parameter with respect to the explanatory variable, rather than keeping it constant. Consequently, we are able to model the conditional distributions in a much more flexible manner.

## 3.2 Dagum distributed conditional income distributions

In the second setting we assume the CIDs to follow a Dagum distribution such that for $y > 0$

$$t(\tilde{y} \mid a, b, p) = \frac{ap\tilde{y}^{ap-1}}{b^{ap}[1 + (\tilde{y}/b)^a]^{p+1}}. \tag{5}$$

Analogously to above we would estimate each of the parameter predictors in an additive manner:

$$\log(\hat{a}) = \eta_a = s_{1a}(age) + Hs_{2a}(age) + Os_{3a}(age) + HOs_{4a}(age), \tag{6}$$

$$\log(\hat{b}) = \eta_b = s_{1b}(age) + Hs_{2b}(age) + Os_{3b}(age) + HOs_{4b}(age), \tag{7}$$

$$\log(\hat{p}) = \eta_p = s_{1p}(age) + Hs_{2p}(age) + Os_{3p}(age) + HOs_{4p}(age), \tag{8}$$

[13]Note that this implies that while for the predictor of Equation 3 we have an additive connection between the explanatory variables, this is multiplicative for Equation 4.

While for the log-normal distribution a direct interpretation of the parameters is possible, this is more intricate for the Dagum distribution. However, using the parameter estimates it is straight forward to obtain estimates for any desired distribution measure, like the mean, standard deviation. Also estimates for other economic measures of interest like inequality measures like the Gini coefficient or the Theil Index can easily be calculated for a given set of parameter estimates. See Section C.2 in the appendix for further discussion.

### 3.3 The estimation algorithm

The frequentist estimation procedure implemented in the `gamlss` package in `R` and will not be discussed in depth here. It employs a backfitting algorithms for the maximisation of the penalised likelihood, which are described in detail in Rigby and Stasinopoulos (2005). It must be noted though that as these models are highly complex, the are several pitfalls which are considered in Section C of the appendix. Bearing those aspects in mind, we turn to the estimation of the CIDs for males in the years 1992 and 2010.

## 4 Assessing conditional income distributions

Before going on to interpreting the the estimates, we assess whether the parametric CIDs we estimate, using the log-normal and the Dagum distribution respectively, provide adequate fits to the data. We do so by testing the hypothesis that

$$H_0 : f(x) = f_0(x, \theta),$$

where $f(x)$ is the observation generating p.d.f. and $f_0(x, \theta)$ is the parametric distribution thought to model the data.

We test the hypothesis by using a bootstrapped version of the Kolmogorov-Smirnov Test which uses Monte Carlo simulations to obtain the distribution of the test-statistic as suggested by Andrews (1997). The test statistic for this test is given by

$$D_n = \sup_x \mid F_0(x, \theta) - S_n(x) \mid, \tag{9}$$

where $F_0(x, \theta)$ denotes the c.d.f. of our parametric fit for the truncated CIDs. Using the parameter estimates from Equations 3-4 and 6-8 for the log-normal and the Dagum distribution respectively, these can be easily obtained. $S_n(x)$ denotes the empirical cumulative distribution function for

observations $x_1, \ldots, x_n$, with $n$ being the sample size of the given subpopulation under consideration. The distribution of this test-statistic was then obtained by parametric bootstrap whereby we used 100,000 simulations samples of size $n$ for each subpopulation yielding a distribution of the test-statistic. Using this procedure we obtained the p-values for each subpopulation given in Table 3 and 4.

The results for the log-normal distribution are displayed in Figure 1 and 2, for the years 1992 and 2010 respectively. For a p-value we expect to see a 5% share of observations to show a test-statistic with a corresponding p-value greater than 0.05. However, the share of subpopulations for which this p-value is surpassed is much higher. Hence, the results show that the log-normal distribution is inapt to model the conditional income distributions under consideration.

The results for the Dagum distribution are displayed in Figure 3 and 4, for the years 1992 and 2010 respectively. In contrast to the log-normal distribution, the rejection rate is much lower. On average over both time periods we get an average rejection rate of 0.069, which is just above the 0.05 we would expect. While it must be noted that for the males without higher education in the West in 1992 we generally have slightly too high shares of rejections, they do portray that there is no systematic problems with the modelling of these income distributions by the Dagum distribution. We therefore conclude that structured additive regression employing the Dagum distribution adequately models the truncated parts of the CIDs under consideration.

| Age | LowEduc.West | HighEduc.West | LowEduc.East | HighEduc.East |
|---|---|---|---|---|
| 21 | 0.000 | | 0.000 | |
| 22 | 0.000 | 0.987 | 0.000 | |
| 23 | 0.000 | | 0.000 | 0.984 |
| 24 | 0.000 | | 0.000 | 0.948 |
| 25 | 0.000 | 0.989 | 0.000 | |
| 26 | 0.000 | 0.845 | 0.000 | 0.995 |
| 27 | 0.000 | 0.994 | 0.000 | 0.815 |
| 28 | 0.000 | 0.000 | 0.000 | 0.975 |
| 29 | 0.000 | 0.000 | 0.000 | 1.000 |
| 30 | 0.000 | 1.000 | 0.000 | 0.000 |
| 31 | 0.000 | 0.000 | 0.000 | 0.998 |
| 32 | 0.000 | 0.000 | 0.000 | 0.984 |
| 33 | 0.000 | 0.000 | 0.000 | 0.991 |
| 34 | 0.000 | 0.000 | 0.000 | 0.000 |
| 35 | 0.000 | 0.000 | 0.000 | 0.000 |
| 36 | 0.000 | 0.000 | 0.000 | 0.000 |
| 37 | 0.000 | 0.000 | 0.000 | 0.000 |
| 38 | 0.000 | 0.000 | 0.000 | 0.000 |
| 39 | 0.000 | 0.000 | 0.000 | 0.000 |
| 40 | 0.000 | 0.997 | 0.000 | 0.000 |
| 41 | 0.000 | 0.000 | 0.000 | 0.000 |
| 42 | 0.000 | 0.000 | 0.000 | 0.000 |
| 43 | 0.000 | 0.000 | 0.000 | 0.000 |
| 44 | 0.000 | 0.000 | 0.000 | 0.000 |
| 45 | 0.000 | 0.000 | 0.000 | 1.000 |
| 46 | 0.000 | 0.000 | 0.000 | 1.000 |
| 47 | 0.000 | 0.994 | 0.000 | 0.000 |
| 48 | 0.000 | 0.000 | 0.000 | 0.000 |
| 49 | 0.000 | 0.000 | 0.000 | 0.999 |
| 50 | 0.000 | 0.000 | 0.000 | 0.000 |
| 51 | 0.000 | 0.000 | 0.000 | 0.000 |
| 52 | 0.000 | 0.000 | 0.000 | 0.000 |
| 53 | 0.000 | 0.000 | 0.000 | 0.000 |
| 54 | 0.000 | 1.000 | 0.000 | 1.000 |
| 55 | 0.000 | 0.000 | 0.000 | 0.996 |
| 56 | 0.000 | 0.971 | 0.000 | 0.000 |
| 57 | 0.000 | 0.990 | 0.000 | 0.999 |
| 58 | 0.000 | 0.957 | 0.000 | 0.000 |
| 59 | 0.000 | 0.998 | 1.000 | 1.000 |
| 60 | 0.000 | 1.000 | 0.000 | 0.999 |
| Share of p<0.05 | 1.000 | 0.600 | 0.975 | 0.500 |

Table 1: Log-normal: P-values from KS-Test for Males 1992

| Age | LowEduc.West | HighEduc.West | LowEduc.East | HighEduc.East |
|---|---|---|---|---|
| 21 | 0.000 | | 0.000 | |
| 22 | 0.000 | | 0.000 | |
| 23 | 0.000 | | 0.000 | |
| 24 | 0.000 | 0.990 | 0.000 | 0.992 |
| 25 | 0.000 | 1.000 | 0.000 | 0.978 |
| 26 | 0.000 | 0.000 | 0.000 | |
| 27 | 0.000 | 0.998 | 0.000 | 0.994 |
| 28 | 0.000 | 0.000 | 0.000 | |
| 29 | 0.000 | 0.000 | 0.000 | 1.000 |
| 30 | 0.000 | 0.000 | 0.000 | 0.000 |
| 31 | 0.000 | 0.000 | 0.000 | 0.000 |
| 32 | 0.000 | 0.000 | 0.000 | 0.000 |
| 33 | 0.000 | 0.000 | 0.000 | 0.000 |
| 34 | 0.000 | 0.000 | 0.000 | 0.997 |
| 35 | 0.000 | 0.000 | 0.000 | 0.865 |
| 36 | 0.000 | 0.000 | 0.000 | 0.947 |
| 37 | 0.000 | 0.000 | 0.000 | 0.000 |
| 38 | 0.000 | 0.000 | 0.000 | 0.994 |
| 39 | 0.000 | 0.000 | 0.000 | 0.000 |
| 40 | 0.000 | 0.000 | 0.000 | 0.999 |
| 41 | 0.000 | 0.000 | 0.000 | 1.000 |
| 42 | 0.000 | 0.000 | 0.000 | 1.000 |
| 43 | 0.000 | 0.000 | 0.000 | 0.000 |
| 44 | 0.000 | 0.000 | 0.000 | 0.984 |
| 45 | 0.000 | 0.000 | 0.000 | 0.000 |
| 46 | 0.000 | 0.000 | 0.000 | 0.000 |
| 47 | 0.000 | 0.000 | 0.000 | 0.000 |
| 48 | 0.000 | 0.000 | 0.000 | 0.000 |
| 49 | 0.000 | 0.000 | 0.000 | 0.000 |
| 50 | 0.000 | 0.000 | 0.000 | 1.000 |
| 51 | 0.000 | 0.000 | 0.000 | 0.000 |
| 52 | 0.000 | 0.000 | 0.000 | 0.000 |
| 53 | 0.000 | 0.000 | 0.000 | 0.000 |
| 54 | 0.000 | 0.000 | 0.000 | 0.000 |
| 55 | 0.000 | 0.000 | 0.000 | 0.000 |
| 56 | 0.000 | 0.000 | 0.000 | 0.000 |
| 57 | 0.000 | 0.000 | 0.000 | 0.000 |
| 58 | 0.000 | 0.000 | 0.000 | 0.000 |
| 59 | 0.000 | 0.000 | 0.000 | 0.000 |
| 60 | 0.000 | 0.000 | 0.000 | 0.000 |
| Share of p<0.05 | 1.000 | 0.850 | 1.000 | 0.550 |

Table 2: Log-normal: P-values from KS-Test for Males 2010

| Age | LowEduc.West | HighEduc.West | LowEduc.East | HighEduc.East |
|---|---|---|---|---|
| 21 | 0.100 | | 0.113 | |
| 22 | 0.522 | 0.606 | 0.011 | |
| 23 | 0.039 | | 0.545 | 0.693 |
| 24 | 0.906 | | 0.283 | 0.265 |
| 25 | 0.163 | 0.961 | 0.816 | |
| 26 | 0.105 | 0.464 | 0.772 | 0.985 |
| 27 | 0.037 | 0.854 | 0.456 | 0.166 |
| 28 | 0.363 | 0.732 | 0.002 | 0.189 |
| 29 | 0.166 | 0.266 | 0.269 | 0.135 |
| 30 | 0.055 | 0.148 | 0.397 | 0.777 |
| 31 | 0.445 | 0.925 | 0.509 | 0.740 |
| 32 | 0.336 | 0.756 | 0.260 | 0.318 |
| 33 | 0.904 | 0.537 | 0.326 | 0.218 |
| 34 | 0.027 | 0.409 | 0.807 | 0.239 |
| 35 | 0.758 | 0.015 | 0.498 | 0.634 |
| 36 | 0.641 | 0.823 | 0.372 | 0.688 |
| 37 | 0.147 | 0.398 | 0.117 | 0.747 |
| 38 | 0.757 | 0.290 | 0.021 | 0.559 |
| 39 | 0.387 | 0.814 | 0.170 | 0.569 |
| 40 | 0.211 | 0.342 | 0.185 | 0.810 |
| 41 | 0.610 | 0.598 | 0.557 | 0.705 |
| 42 | 0.832 | 0.957 | 0.884 | 0.898 |
| 43 | 0.505 | 0.031 | 0.306 | 0.631 |
| 44 | 0.302 | 0.651 | 0.159 | 0.498 |
| 45 | 0.543 | 0.548 | 0.909 | 0.203 |
| 46 | 0.058 | 0.620 | 0.289 | 0.472 |
| 47 | 0.344 | 0.432 | 0.343 | 0.678 |
| 48 | 0.066 | 0.885 | 0.580 | 0.135 |
| 49 | 0.481 | 0.256 | 0.056 | 0.014 |
| 50 | 0.477 | 0.337 | 0.093 | 0.530 |
| 51 | 0.950 | 0.164 | 0.030 | 0.416 |
| 52 | 0.178 | 0.696 | 0.056 | 0.051 |
| 53 | 0.860 | 0.166 | 0.017 | 0.242 |
| 54 | 0.100 | 0.051 | 0.033 | 0.037 |
| 55 | 0.156 | 0.856 | 0.029 | 0.092 |
| 56 | 0.530 | 0.389 | 0.134 | 0.921 |
| 57 | 0.210 | 0.126 | 0.163 | 0.864 |
| 58 | 0.894 | 0.013 | 0.576 | 0.479 |
| 59 | 0.057 | 0.310 | 0.043 | 0.700 |
| 60 | 0.540 | 0.604 | 0.752 | 0.849 |
| Share of p<0.05 | 0.075 | 0.075 | 0.200 | 0.050 |

Table 3: Dagum: P-values from KS-Test for Males 1992

| Age | LowEduc.West | HighEduc.West | LowEduc.East | HighEduc.East |
|---|---|---|---|---|
| 21 | 0.127 | | 0.934 | |
| 22 | 0.759 | | 0.970 | |
| 23 | 0.300 | | 0.553 | |
| 24 | 0.150 | 0.114 | 0.934 | 0.647 |
| 25 | 0.543 | 0.729 | 0.366 | 0.331 |
| 26 | 0.361 | 0.082 | 0.965 | |
| 27 | 0.130 | 0.659 | 0.935 | 0.822 |
| 28 | 0.159 | 0.744 | 0.134 | |
| 29 | 0.341 | 0.561 | 0.278 | 0.790 |
| 30 | 0.236 | 0.714 | 0.424 | 0.680 |
| 31 | 0.552 | 0.417 | 0.263 | 0.807 |
| 32 | 0.794 | 0.960 | 0.821 | 0.539 |
| 33 | 0.275 | 0.393 | 0.482 | 0.408 |
| 34 | 0.208 | 0.912 | 0.261 | 0.080 |
| 35 | 0.756 | 0.352 | 0.543 | 0.185 |
| 36 | 0.113 | 0.538 | 0.011 | 0.006 |
| 37 | 0.563 | 0.713 | 0.518 | 0.792 |
| 38 | 0.437 | 0.407 | 0.762 | 0.639 |
| 39 | 0.080 | 0.968 | 0.266 | 0.670 |
| 40 | 0.162 | 0.638 | 0.326 | 0.639 |
| 41 | 0.821 | 0.238 | 0.976 | 0.660 |
| 42 | 0.532 | 0.592 | 0.077 | 0.762 |
| 43 | 0.772 | 0.435 | 0.703 | 0.292 |
| 44 | 0.314 | 0.328 | 0.264 | 0.163 |
| 45 | 0.152 | 0.193 | 0.380 | 0.344 |
| 46 | 0.668 | 0.784 | 0.582 | 0.995 |
| 47 | 0.743 | 0.595 | 0.918 | 0.929 |
| 48 | 0.972 | 0.580 | 0.237 | 0.367 |
| 49 | 0.444 | 0.736 | 0.011 | 0.197 |
| 50 | 0.671 | 0.430 | 0.657 | 0.866 |
| 51 | 0.385 | 0.788 | 0.534 | 0.545 |
| 52 | 0.874 | 0.848 | 0.060 | 0.991 |
| 53 | 0.460 | 0.846 | 0.010 | 0.905 |
| 54 | 0.872 | 0.858 | 0.220 | 0.853 |
| 55 | 0.399 | 0.508 | 0.036 | 0.605 |
| 56 | 0.402 | 0.730 | 0.202 | 0.999 |
| 57 | 0.756 | 0.385 | 0.094 | 0.309 |
| 58 | 0.203 | 0.852 | 0.554 | 0.984 |
| 59 | 0.274 | 0.405 | 0.048 | 0.209 |
| 60 | 0.342 | 0.764 | 0.213 | 0.642 |
| Share of $p<0.05$ | 0.000 | 0.000 | 0.125 | 0.025 |

Table 4: Dagum: P-values from KS-Test for Males 2010

# 5 Analysing within-group inequality

From the estimated CIDs a variety of distribution measures can be deduced and analysed (see the Section F in the appendix for some measures). In this paper we focus on the matter of residual inequality and thus concentrate on the within-group inequality as measured by the Theil index of the CIDs. Note that we consider the whole CIDs again and not only the truncated CIDs. People who are unemployed and those in precarious employment with an labour income below 4,800€ are hence included in this analysis.

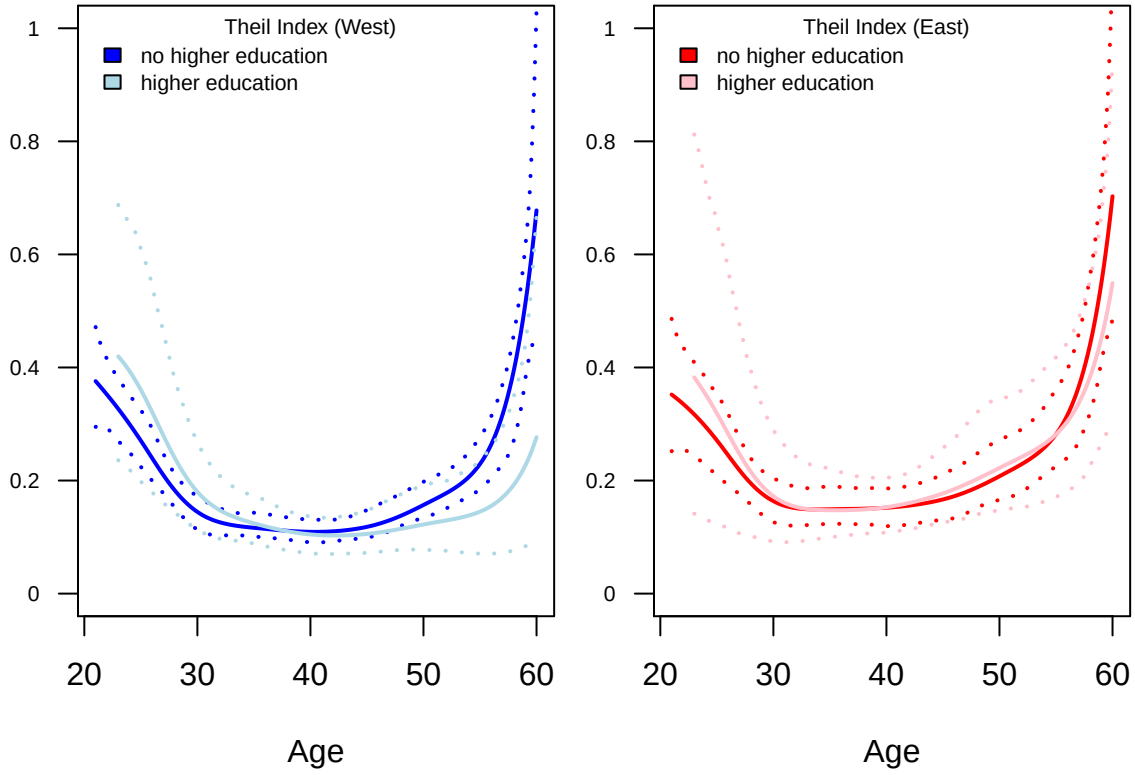


Figure 1: Theil Indices of CIDs in 1992

Figure 1 displays the Theil indices deduced from the CIDs for labour incomes in 1992. The solid line marks the maximum likelihood estimate, while the dotted lines are bootstrapped 95% confidence intervals. As we can see, the within-group changes over the age-span for both education levels and in West and East. Generally, we observe a U-shaped relation, so that within-group inequality is markedly higher for men below 30 and above 55. This is hardly surprising, since at a young age (due to education/vocational training) zero- and low incomes are common. At a higher age, retirement rates and reduced work-schemes increase substantially causing a wider dispersion of the CID and hence inflating the Theil index. Concerning the difference between men with higher education and men without it, we can observe that in the East there generally appears to be little difference, although it may be noted that while the intra-group inequality is higher for those with

higher education at an early age, it is lower towards the end of the age span. For the East this can also be observed, although the differences are slimmer. Both for East and West, it must be noted though that the differences are not significant at the 5% level. Between East and West, difference also appear to be slim, with a slight tendency of higher intra-group inequality in the West for both education levels and most ages.

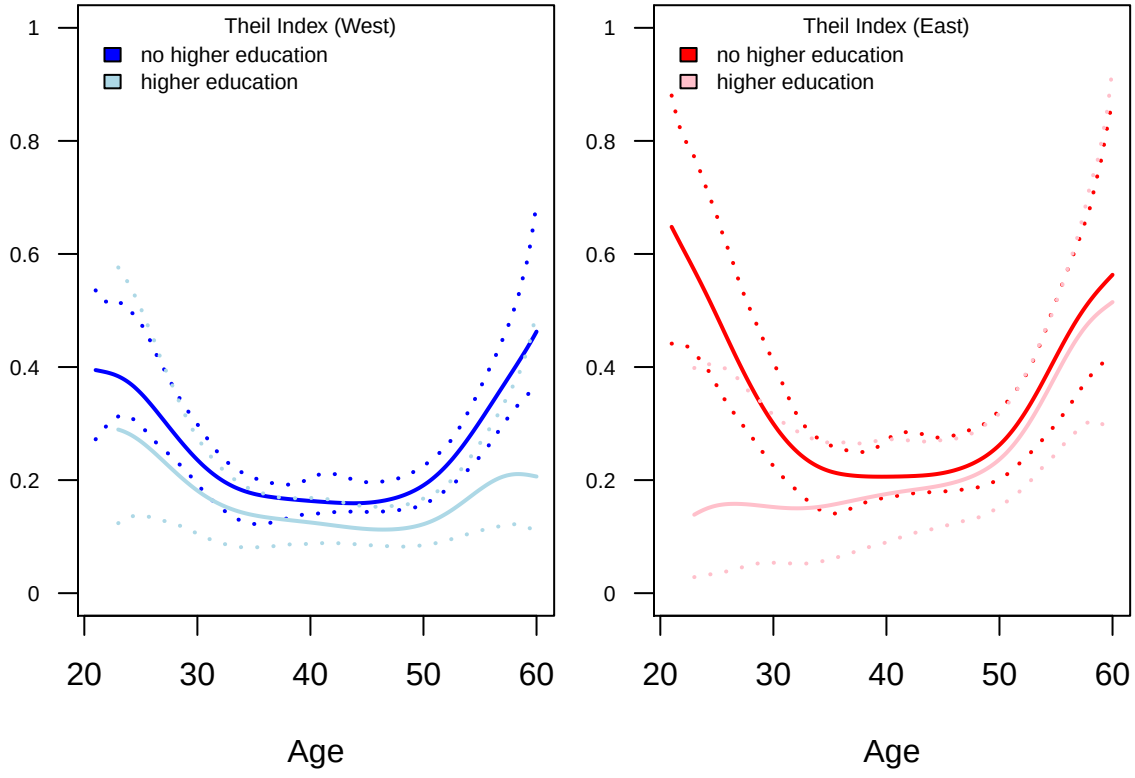


Figure 2: Theil Indices of CIDs in 2010

Figure 2 displays the corresponding Theil indices for labour incomes in 2010. As can be observed the general trend of the U-shape has persisted over time, with one exception. Men with higher education in the East do not display an upward swing for the younger ages. The most plausible explanation for this is that the East of Germany traditionally has one year of schooling less, so that it is much more frequent for people to start studying at the age of 18 (instead of 19 in the West) resulting in substantially higher numbers of people already in work by the age of 23. A second explanation is the low number of observations such that this finding may well also be an artefact of the sample. By contrast, men without higher education see a drastically increased within-group inequality. The main thrust of this increased inequality is the risen share of zero- and precarious incomes. This rise on the lower range of the distribution induces a much more pronounced positive skew to the CID and increasing the Theil index. Although less pronounced, a similar rise is observed for young men without higher education the West. By contrast at the other side of the age-span our estimation results indicate an opposite dynamic of decreasing within-group

inequality.

Concerning the impact of education on the Theil index, we can note that the differences of the within-group inequality also seem to follow a divergence, with the within-group inequality mainly on the rise for those without higher education. Especially young men in the East seem to be affected by a rise in intra-group inequality.

Commenting on these dynamics it is obviously of interest, whether the changes are statistically significant at the usual levels. As the wide confidence intervals for the Theil index indicate, statistically significant differences are rare. Given the complexity of the estimation of whole conditional distributions and the relative scarcity of data available, this is hardly surprising. It follows that the number of parameters and explanatory variables used in the regression is much more restricted for a given data size if reasonably precise estimates on distribution measures are desired. Given the number of factors which influence the level of an individual's wage and thus the conditional income distribution, the concept of inference and causality is much more intricate. Much more work needs to be done in this direction. For the moment, we thus consider the findings displayed here as rather advanced descriptive statistical methodology, which hints at developments rather than pinpointing at supposed causalities.

# 6 Conclusion and Outlook

At the outset of this article we highlighted the need for the analysis of conditional income distributions. Using structured additive distributional regression, we showed that it is possible to estimate conditional income distributions (CIDs) with respect to a set of variables, both continuous and discrete in an additive set-up. Specifically, we regressed German labour incomes on a continuous variable age, and two discrete binary variables education and region. For these CIDs we contrasted the log-normal distribution against the Dagum distribution and found that the latter provided an better fit. Subsequently, we considered the considered the development of residual inequality with respect to the three explanatory variables. We found conditional inequality, as measured by the Theil index of the CIDs, to be especially high among young men in the East for whom inequality also seems to have increased considerably between 1992 and 2010. If these findings were substantiated, it would mean that next to the nature of the much discussed literature on the skill-biased technological change has missed an important dimension. It has been noted that residual inequality seem to have risen in tandem with overall inequality (p.17 Acemoglu, 2002). However, our results indicate that the rise of this residual inequality, at least in Germany, is first and foremost driven by rising inequality among the unskilled workers without university education, most notably the young.

Clearly much work remains to be done in the field of modelling conditional income distributions/residual inequality and this working paper is no more than a first attempt at addressing the issue. The trisection we apply to include zero- and precarious incomes is somewhat haphazard and better ways to model the whole conditional income distribution are needed. One evasion of this problem is clearly only to include those in employment or even just those in full employment. Yet to us this exclusion of major parts of the population to us excluded important dynamics of income inequality. Having said that, the exclusion of the female side of the labour market is of course analogue and further research ought to aim to include both genders. In addition much more work needs to be done with regard to the incorporation of more explanatory variables. Taking the literature on the Mincer wage equations as a guideline, labour market experience should definitely be included, as well as industry of employment. Also borrowing from the literature on income decomposition the issue of institutional differences, trade union strength, should also be incorporated in the equations. Yet, as pointed out before, structured additive distributional regression is far less robust to additional variables than standard mean regression and consequently, new innovative ways of incorporating these variables into the regression framework without using up too many precious degrees of freedom are required.

So while much work remains to be done and many problems remain to be addressed, we believe that structured additive distributional regression holds much promise to broadening the perspective of income analysis.

# References


M. Abramowitz and I. A. Stegun (1972): Handbook of Mathematical Functions, Dover, New York.

D. Acemoglu (2002): Technical Change, Inequality and the Labor Market, in: Journal of Economic Literature, 40(1), pp. 7–72.

H. Akaike (1983): Information measures and model selection, in: Bulletin of the International Statistical Institute, 50, pp. 277–290.

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

A. B. Atkinson (1975): The economics of inequality, Clarendon Press, Oxford.

— (2003): Income Inequality in OECD: Data and Explanations, CESifo Working Paper 881.

S. Bach, G. Corneo and V. Steiner (2009): From Bottom to Top: the Entire Income Distribution in Germany, 1992-2003, in: Review of Income and Wealth, 55(2), pp. 303–330.

BfAS (2013): Der Vierte Armuts- und Reichtumsbericht der Bundesregierung, Bundesministerium für Arbeit und Soziales, Berlin.

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.

M. Biewen and A. Juhasz (2012): Understanding Rising Income Inequality in Germany, in: Review of Income and Wealth, 58(4), pp. 622–647.

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.

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

D. Chotikapanich (2008): Introduction, in: D. Chotikapanich (ed.), Modeling Income Distributions and Lorenz Curves, pp. ix–xii, Springer, New York.

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

F. A. Cowell, S. P. Jenkins and J. A. Litchfield (1996): The Changing Shape of the UK Income Distribution: Kernel Density Estimates, in: J. Hills (ed.), New inequalities, pp. 49–75, Cambridge University Press, Cambridge.

C. Dagum (1977): A New Model of Personal Income Distribution: Specification and Estimation, in: Economie Applicée, 30, pp. 413–437.

C. Domańki and A. Jedrzejczak (1998): Maximum likelihood estimation of the Dagum model parameters, in: International Advances in Economic Research, 4, pp. 243–252.

C. Dustmann and U. Schönberg (2009): Training and Union Wages, in: Review of Economics and Statistics, 91(2), pp. 363–376.

B. Efron (1979): Bootstrap Methods: Another Look at the Jackknife, in: The Annals of Statistics, 7(1), pp. 1–26.

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

N. M. Fortin and T. Lemieux (1998): Rank regression, wage distributions and the gender gap, in: Journal of Human Resources, 33(3), pp. 610–643.

R. Gibrat (1931): Les Inégalités Economiques, Sirely, Paris.

M. M. Grabka (2013): Codebook for the PEQUIV File 1984-2011, DIW Berlin, Berlin.

S. P. Jenkins (2007): Inequality and the GB2 income distribution, in: IZA Discussion paper series, No. 2831.

D. N. Joanes and C. A. Gill (1998): Comparing Measures of Sample Skewness and Kurtosis, in: The Statistician, 47(1), pp. 183–189.

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

N. Klein, T. Kneib, A. Sohn and S. Lang (2014): Bayesian Structured Additive Distributional Regression, in: Working Papers in Economics and Statistics, 2013-23.

R. Koenker and G. Bassett (1978): Regression quantiles, in: Econometrica, 46(1), pp. 33–50.

P. R. Krugman (2007): The conscience of a liberal, W.W. Norton & Co., New York, 1 ed.

T. Lemieux (2003): The "Mincer Equation" Thirty Years after Schooling, Experience, and Earnings, in: Center for Labor Economics -University of California, Working Paper No. 62.

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. B. McDonald (1984): Some Generalized Functions for the Size Distribution of Income, in: Econometrica, 52, pp. 647–663.

J. B. McDonald and M. R. Ransom (2008): The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality, in: D. Chotikapanich (ed.), Modeling Income Distributions and Lorenz Curves, pp. 147–166, Springer, New York.

P. W. Mielke and E. S. Johnson (1974): Some generalized distributions of the second kind having desirable application features in hydrology and meteorology, in: Water Resources Research, 10, pp. 223–226.

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

R. A. Rigby and D. M. Stasinopoulos (2005): Generalized Additive Models for Location, Scale and Shape, in: Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(3), pp. 507–554.

S. J. Sheather and M. C. Jones (1991): A reliable data-based bandwidth selection for kernel density estimation, in: Journal of the Royal Statistical Society: Series B (Statistical Methodology), 53(3), pp. 683–690.

J. Silber (1999): Introduction - Thirty Years of Intensive Research on Income Inequality Measurement, in: J. Silber (ed.), Handbook of Income Inequality Measurement, vol. 1-18, Kluwer Academic, Boston.

S. K. Singh and G. S. Maddala (1976): A Function for Size Distribution of Incomes, in: Econometrica, 44, pp. 963–970.

R. Skidelsky (2010): Keynes: The Return of the Master, Public Affairs, New York.

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 (2012): Verbraucherpreisindizes für Deutschland - Lange Reihen ab 1948, in: Preise, November 2012.

M.-P. Victoria-Feser (1995): Robust methods for the analysis of income distribution models with applications to Dagum's model, in: C. Dagum and A. Lemmi (eds.), Income distribution, welfare, inequality and poverty, pp. 225–239, JAI Press, Greenwich.

S. Vollmer, H. Holzmann, F. Ketterer and S. Klasen (2013): Distribution Dynamics of Regional GDP per Employee in Unified Germany, in: Empirical Economics, 44(2), pp. 491–509.

G. G. Wagner, J. R. Frick and J. Schupp (2007): The German Socio-Economic Panel Study (SOEP) - Scope Evolution and Enhancements, in: Schmollers Jahrbuch, 127(1), pp. 139–169.

G. G. Wagner, J. Göbel, P. Krause, R. Pischner and I. Sieber (2008): Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland – Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender), in: AStA Wirtschafts- und Sozialstatistisches Archiv, 2(4), pp. 301–328.

# A Data

## A.1 Sample

For our analysis we use the SOEP Database. In order to avoid distortion, we follow Bach et al. (2009) and only employ samples A-F, both for 1992 and 2010. We thus explicitly exclude the high-income sample, which allows us account for the upper tail of the distribution more accurately.

It also should be noted that our cross-sectional approach has several weaknesses. As Atkinson (2003) points out, single years are can be bad representatives of longer periods such as decades and can be highly misleading. Secondly, we our cross-sectional approach does not allow us to exploit the panel structure provided by the SOEP. In the future we plan to incorporate fixed and random effects in our analysis of the German income structure. But since many of the recent studies on the German income distribution also use cross-section, we consider a cross-sectional approach to be worthwhile.

Bach et al. (2009) only consider adults of 20 years or older. We extend this age restriction such that we only consider adults between 21 and 60 years of age. The reasoning behind this restriction is that we want to observe changes in the annual gross income over the period of standard employment. We thus exclude the time when most people finance themselves largely by pension payments, very similar to the exclusion of young people who are financed by their parents.

## A.2 Gross Market Income

We employ the definition proposed by Bach et al. (2009), excluding capital incomes though. Thereby, only the first two of the following three income components are incorporated to add up the individual's income.

- *Wage income* is the payment of wages and salaries received by the individual from all his employers as well as the employers' social security contributions. To obtain the annual income from wages and salaries we use the Variable *I11110$$* from the PEQUIV File (Grabka, 2013), which entails all the income from a dependent employment relationship.[14] For employees in the private sector we then add 20% in 1992 and 2010 to account for the employers' social security contributions.For civil servants we account for their *Vollversorgung* by adding 47% to their annual labour income for both 1992 and 2010. It should be noted that these are momentary rough-and-ready measures, which we intend to refine later.

[14]Note that we took the income information for a given year from the subsequent SOEP questionnaire, as the annual income in a questionnaire is naturally given for the previous year.

- *Income from business activity* entails includes income from unincorporated business enterprise and from self-employment activities, as well as taxable income from agriculture and forestry. This source of income is also accounted for in *I11110$$* from the PEQUIV File, or rather *ISELF$$* therein.
- *Capital income* entails incomes from interest and dividends as well as renting and leasing and is not considered.

## A.3 Unconditional income distributions

The two personal income distributions which we will consider in the following are displayed in Figure 3. Note that we inflated incomes in 1992 to be valued in 2010 Euros using the consumer price index (Statistisches Bundesamt, 2012).

From the graphic alone, no striking differences can be noted. One can observe a slight fall in the density of the lowest income bin, as the share of people with an income below 5,000€ per annum declines by over ten percent from 1992 to 2010. Concerning the other bins, differences are very slim. Looking carefully one can notice a slight upward shift of the densities from 50,000€ per annum.

Looking at some summary statistics, we observe an increase in the mean income from 29,200€ in 1992 to 31,000€ in 2010. The median also increased in the same time period from 25,400€ to 26,400€. Aggregate inequality, as measured by the Gini coefficient, somewhat surprisingly even falls slightly from 0.503 to 0.497. This implies a 1% fall in inequality as measured by the Gini coefficient and is somewhat counter-intuitive given the literature on rising inequality.

The reason for this slightly puzzling result is found in the concealing nature of our unconditional analysis which amalgamates various developments. Among other things it hides the differences due to developments in East and West Germany, changes in the income distribution due to rising female labour market participation and changes due to the changing demographic structure of the German population.[15] What this highlights is the importance of considering not only unconditional income distributions but to consider conditional income distributions (CIDs), that is to decompose the unconditional income distribution. This is standard in the literature.

Yet standard income decomposition by definition takes a macro-perspective in the sense its primary purpose is the explaining the contribution of various changes to the changes of the aggregate income distribution. The analysis we will conduct in the following differs from classical income inequality decomposition as it takes a fundamentally different micro-perspective. The focus is

[15]See below for a first analysis of conditional income distributions dependent on gender, age and region which highlight the very different dynamics of the development of the conditional income distributions.

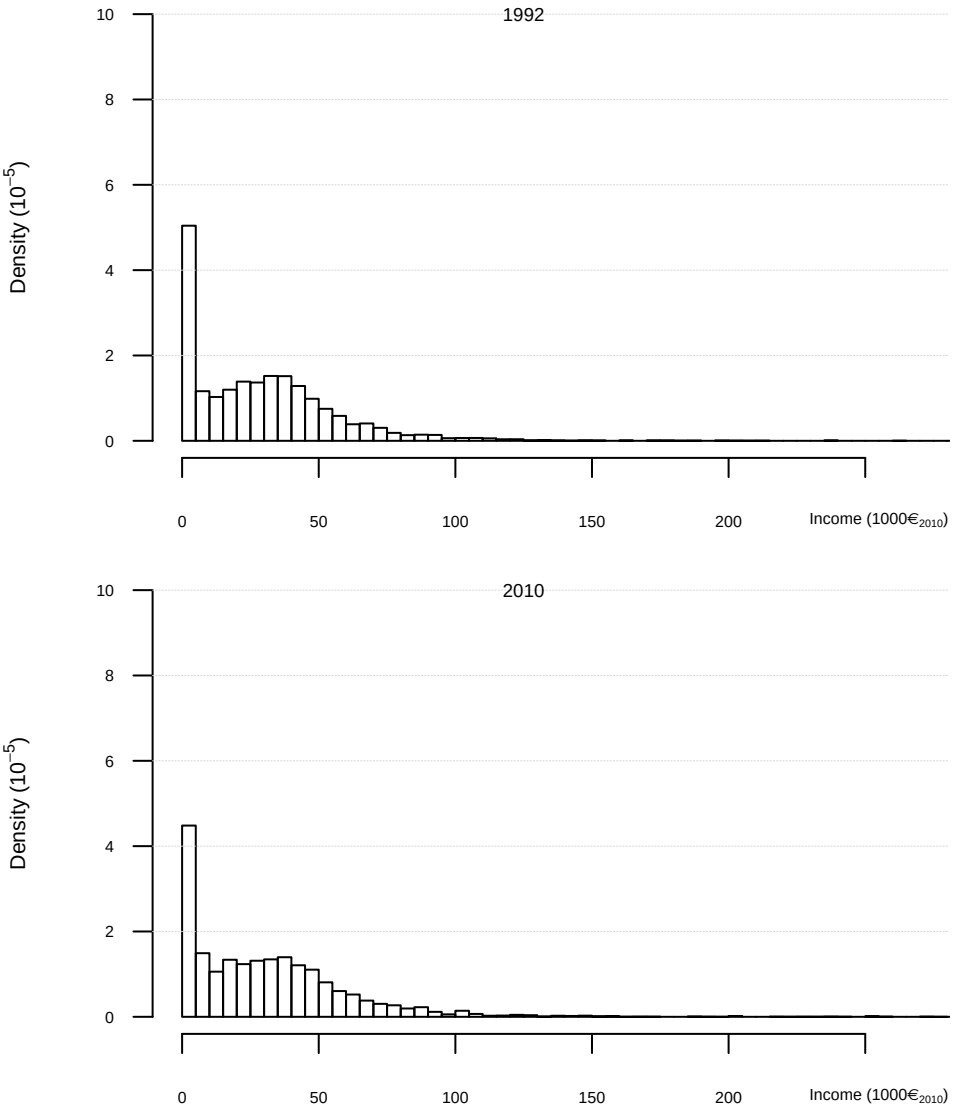


Figure 3: Unconditional German Income Distribution

directed towards the changes of the income distributions of subpopulations. The impact of these changes on the aggregate income distribution is secondary.[16] The main impetus of our research is thus the analysis of changes experienced by subpopulations rather than the population at large.

[16]Distributional regression based estimation of CIDs can also be used for classical decomposition. Yet this aspect won't be discussed in detail here.

# B Graphical analysis of dependent income distributions

In this section we display the graphics, of the decomposition following a two level education definition from Acemoglu (2002) and the three age groups from Dustmann and Schönberg (2009) for 1992 and 2010 for East and West. Note that we use the same scale on the y- and x-axis as in the unconditional distribution above, such that the density on the left is in the order of $10^{-5}$ while the income is denoted in 1000s of Euros with purchasing power of the year 2010. Two aspects shall be highlighted. First, while these conditional income distributions are far more coarse with respect to the explanatory variable than the one we obtain by distributional regression, they already give an indication for the adequacy of the various distributions we display in for each CID. Note that the black line indicates a kernel density estimate which uses Sheather and Jones (1991) for bandwidth selection. Second, it can be observed that at this still highly aggregated perspective we obtain considerably varying CID estimates for different subpopulations.

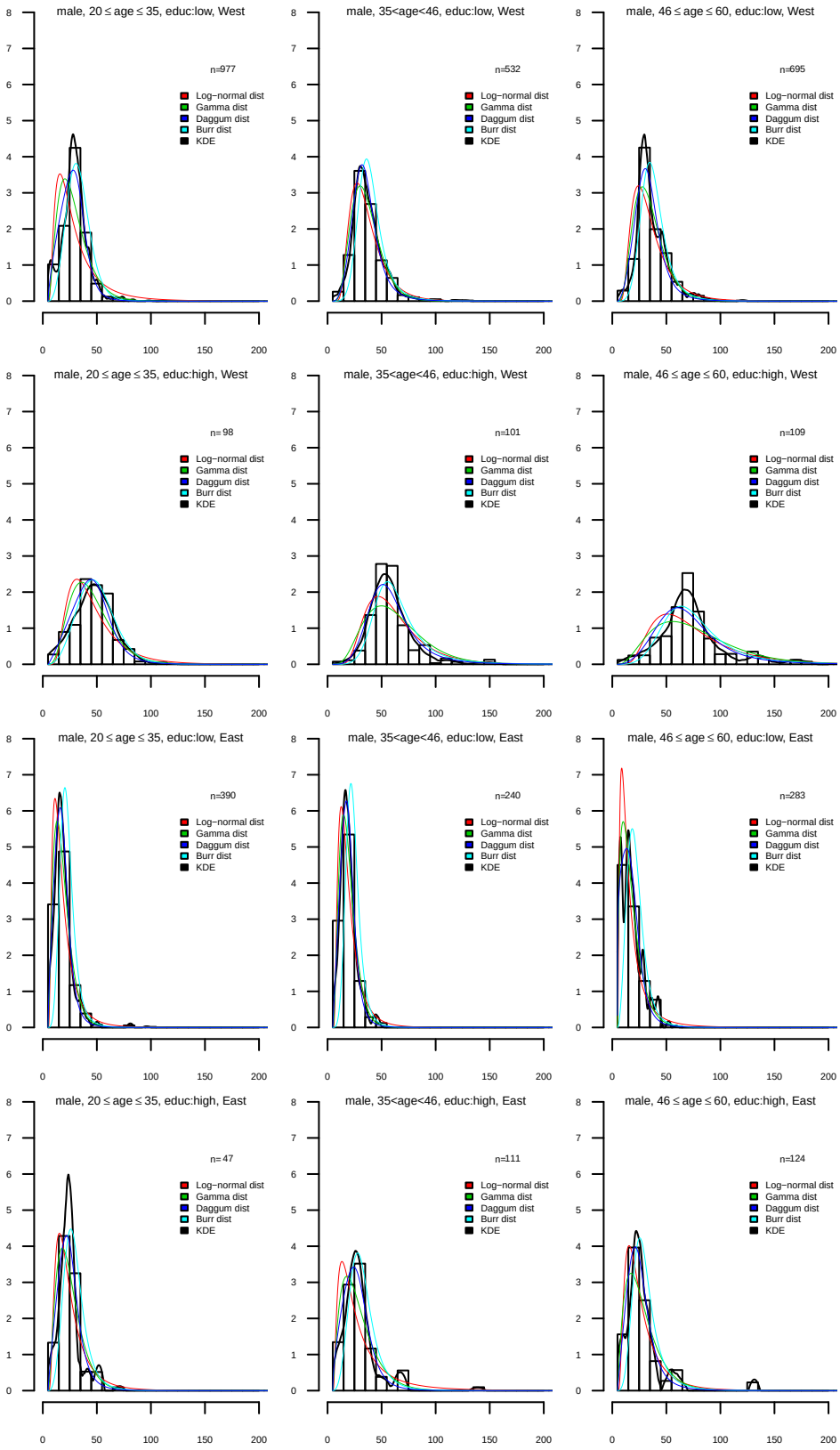


Figure 4: Income Distribution of Males in 1992

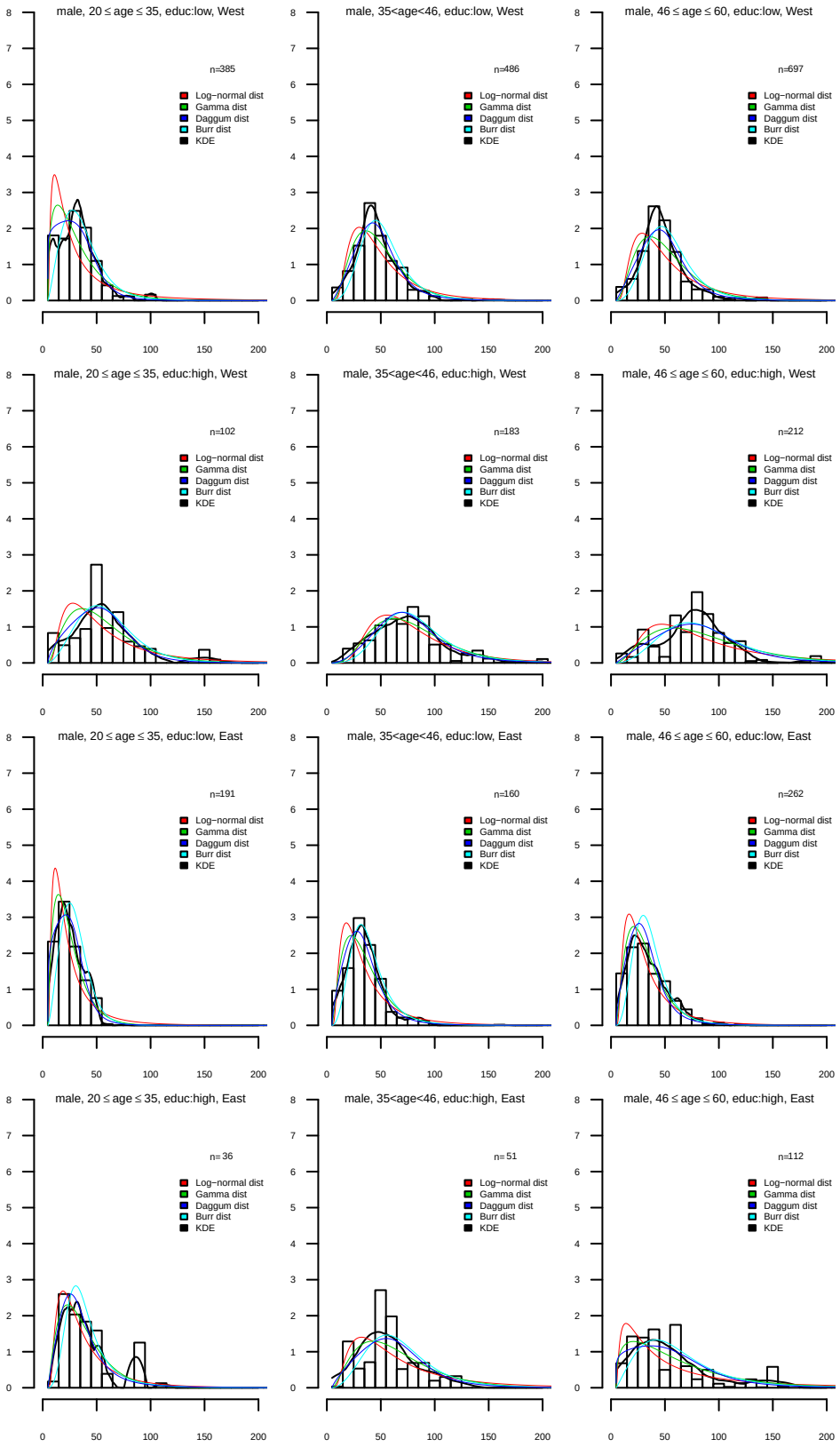


Figure 5: Income Distribution of Males in 2010

# C Structured additive distributional regression

Structured additive distributional regression provides a semiparametric framework for the estimation of conditional distributions. It is closely related to the idea of Generalised Additive Models of Location, Scale and Shape (GAMLSS) first proposed by Rigby and Stasinopoulos (2005). The main reason for using the term additive distributional regression is that for the modelling of CIDs we are not only interested in the direct interpretation of the parameters of location, scale and/or shape of a distributional, but rather auxiliary measures of the estimated conditional distributional. Thus the term distributional regression phonetically seemed more appropriate.

The estimation of the CIDs relies on parametric distributions. In contrast, the effect of explanatory variables on the parameter estimate entails non-parametric procedures, such as splines. For our estimation of the non-linear effects of age, we employ penalised B-splines (see Eilers and Marx, 1996) and use the `gamlss` package in`R` to numerically maximise the penalised likelihood. One aspect which must be paid particular attention is the selection of the smoothing parameters that determine the roughness of function estimates and for which Rigby and Stasinopoulos (2005) propose several alternatives. For our estimations we use the generalised Akaike Information criterion (see Akaike, 1983) to select an appropriate set of smoothing parameters. For given smoothing parameters we then maximise the penalised likelihood using the RS algorithm. For further information we refer to Rigby and Stasinopoulos (see 2005, pp.535-541).

Naturally, the estimation of whole distributions is much more involved than mean regression, as several interdependent predictors have to be estimated simultaneously.

It must be pointed out, that as the estimation strategy relies on numerical methods it is in principle liable to the standard problems associated therewith (convergence, local maxima, etc.). Caution is thus required in the estimation process. In addition, while the Fisher information matrix of the Dagum distribution in principle readily available and can used in the estimation to obtain asymptotic standard errors. However, again several problems can occur, for example if the covariance matrix is not positive definite. For more information see (Rigby and Stasinopoulos, 2005). In any case, due to the fact that we are estimating whole conditonal distributions, rather than just the mean, small sample inference is likely to be of greater relative importance than for mean regression. In a frequentist framework, non-parametric bootstrapping methodology (Efron, 1979) allows for the estimation of confidence bands for the parameters themselves as well as the auxiliary measures. While this can be computationally intensive, it is straight forward to implement. However, as this requires the repetitive estimation, the stability of the estimation process must be ensured (see above). Alternatively the issue of parameter uncertainty can be addressed in a Bayesian framework (see for example Klein et al., 2014).

Similarly, much work remains to be done on model selection. As Nick Longford plastically points out: "The new models are top of the range mathematical Ferraris, but the model selection that is used with them is like a sequence of tollbooths at which partially sighted operators inspect driver's licenses and road worthiness certificates." Thus models must for the moment be primarily selected on the grounds of economic reasoning.

More specifically to our estimation of Dagum distributions Kleiber and Kotz (2003, p.218) point to problems with likelihood based estimation. Domański and Jedrzejczak (1998) show in a simulation study that simple ML estimation, while consistent, is liable to biased estimation for small to moderate sample sizes. While more work must be done on this issue, preliminary simulations we conducted show that while these biases persist for small samples, their impact on the auxiliary measures we use is in general within an acceptable range even for very small sample sizes. Nonetheless, work on more robust estimation strategies would be necessary to make GAMLSS estimation of CID less sensitive to isolated observations. Especially the problem of the score function becoming unbounded can cause severe problems in the estimation process. Previous work by Victoria-Feser (1995) already gives a methodological framework which may be applied to account for these problems.

Despite these problems which will have to be addressed in the future, GAMLSS offers new perspectives for the analysis of conditional income distribution which we turn to now. For further discussion on GAMLSS contrasting it with mean regression, quantile regression and distribution regression see section D.

## C.1 Inclusion of point masses

For modelling the whole CID, we employ two point masses next to the truncated CID. We thus partially discretize the continuous income distribution similar to Fortin and Lemieux (1998). Naturally, a better representation of the distribution is desirable but will have to be left to further research. Yet for most auxiliary measures, like the mean or the Theil index, it is straight forward to incorporate the point masses.

For example, for the mean of the CIDs, we can simply use:

$$\hat{\mu}_{CID} = \hat{p}_{pr}\hat{\mu}_{pr} + \hat{p}_{t(y)}\mu_{t(y)}, \tag{10}$$

where $\hat{p}_{pr}$ and $\hat{p}_{t(y)}$ are the probability of precarious incomes and the probability the income falling into the truncated CID. $\hat{\mu}_{pr}$ is the estimated mean income of those with precarious incomes. For simplicity we do not condition on age but only on education and region for the estimation of $\hat{\mu}_{pr}$. The mean of the truncated CID is denoted by $\mu_{t(y)}$. In case of the Dagum distribution it is given

by Equation (13).

While analytical solutions are in principle available, we used brute computational capacity for the calculation. For the Theil index of each CID we use an approximation of the estimated Dagum distribution whereby we discretize it into 100,000 bins on the range 4,800€ to 10,004,800€ evaluating the density at the centre of each bin. We then use the standard formula for the Theil index weighting each value from the discretization as well as the mean with precarious incomes as well as zero-incomes with their corresponding density. For the latter, it should be noted that we approximate the zero-incomes by a symbolic cent, as the Theil index is only defined for positive incomes. For the Gini coefficient, which we also calculated, but didn't display, we proceeded analogously.

For the quantile ranges and the shares above the threshold values we can simply incorporate the point masses into the distribution and determine the quantile and p-values for the desired level. It should be noted that the point masses for the precarious incomes cause the slight discontinuities for the interquantile ranges observed in Figure 10 and 12.

## C.2 The Dagum distribution

In the subsequent estimations of the conditional distributions we will therefore primarily consider the Dagum distribution.The distribution was introduced as an income distribution by the Italian Camilo Dagum in 1977. It is nested in the Generalised Beta distribution of Second Kind (Mielke and Johnson, 1974) and its probability density function is given by

$$f(y) = \frac{apx^{ap-1}}{b^{ap}[1+(x/b)^a]^{p+1}},\ y > 0, \tag{11}$$

which yields the cumulative density function

$$F(y) = \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-p},\ y > 0, \tag{12}$$

where $a,\ b,\ p > 0$.[17]

Following the notation from Kleiber and Kotz (2003), $b$ is a scale parameter while $a$ and $p$ are shape parameters. As we can see in Equation (11) the parameters impact on the density is intricate as there are strong interrelations among the three parameters. Thus the direct economic interpretation of the parameters is limited. However, as the estimated parameters specify the conditional distribution, we can assess the CID. For this purpose, we able to borrow from the

[17]Note that we are using the notation and parametrisation of Kleiber and Kotz (2003), which is slightly different from the one to the parametrisation of Stasinopoulos and Rigby (2007) whose package we apply for estimation.

wide range measures applied in the economic literature to assess a size distribution with respect to the question at hand. First and foremost the literature tends to consider the distribution's first moment, i.e. the mean. For the Dagum distribution the first moment is given by:

$$\mu = \frac{bB(p+1/a, 1-1/a)}{B(p,1)}, \quad \text{for } a > 1, \tag{13}$$

where $B$ denotes the beta function as defined in Abramowitz and Stegun (1972, p.258). Note that the moment only exists for $1 < a$. This measure of location generally forms the backbone of econometric analysis and will be considered in detail. Another location measure is the mode, which is given by

$$\text{mode} = b\left(\frac{ap-1}{a+1}\right)^{1/a}, \quad \text{if } ap > 1, \tag{14}$$

and is at zero otherwise. While we will not analyse the mode of conditional distributions in detail, it is important to note that the Dagum distribution can thus be both unimodal and zeromodal. The importance of this attribute was already noted by Dagum (1977). We will see later on that for conditional distributions the ability to model zeromodal distributions is a key requirement.

However, while location measures are quintessential to any analysis of income distributions, their account of distributions is naturally limited. Thus additional measures will have to be considered. Thus we consider additional (standardised) moments like the standard deviation and the skewness of the distribution, which are given in Equations 15 and 16 respectively.

$$\sigma = b\sqrt{\frac{B(p+2/a, 1-2/a)}{B(p,1)} - \left(\frac{B(p+1/a, 1-1/a)}{B(p,1)}\right)^2}, \quad \text{for } a > 2, \tag{15}$$

$$\text{skewness} = \frac{B^2(p,1)\lambda_i - 3B(p,1)\lambda_2\lambda_1 + 2\lambda_1^3}{[B(p,1)\lambda_2 - \lambda_1^2]^{3/2}]}, \quad \text{for } a > 3, \tag{16}$$

where $\lambda_i = B(p+i/a, 1-i/a)$. For $a \leq 2$ and $a \leq 3$ the second and third moment do not exist respectively. In addition to these moments we also consider standard inequality measures which comprise both these aspects.

The two most widely used inequality measures are the Gini coefficient and the the of generalised entropy measures. The Gini coefficient can easily be obtained by

$$\text{G} = \frac{\Gamma(p)\Gamma(2p+1/a)}{\Gamma(2p)\Gamma(p+1/a)} - 1, \tag{17}$$

where $\Gamma$ denotes the gamma function as defined by Abramowitz and Stegun (1972, p.255).

Similarly we can obtain the generalised entropy measures. In the following we will concentrate on

the Theil index which is given by

$$\mathrm{I}(1) = \frac{\psi(p+1/a)}{a} - \frac{\psi(1-1/a)}{a} - \Gamma(p+1/a) - \Gamma(1-1/a) + \Gamma(p) + 1, \tag{18}$$

where $\psi(z) = \frac{d}{dz}\log\Gamma(z)$ is the digamma function (see Jenkins, 2007).

Interquantile ranges are the third inequality measure we will consider, which can be obtained directly from Equation (12). In specific, we follow Dustmann and Schönberg (2009) and consider the differential between the $85^{\text{th}}$ and the $50^{\text{th}}$ as well as between the $50^{\text{th}}$ and $15^{\text{th}}$ percentile.

## C.3 Other distributions considered for CIDs

The key question for modelling the conditional income distribution is how to model the truncated income distribution above the 4800€ threshold? Pursuing a parametric approach the choice of the distribution is paramount. Borrowing from the aggregate distribution literature, several distributions come to mind: The natural starting choice for the truncated income distribution is the log-normal distribution, which was popularised by Gibrat (1931). Next to it's appealing theoretical properties it has the advantage that it's parameters are directly interpretable. While at least for the aggregate distribution it is well known to have problems to fit the the upper bound of the distribution (see among others Atkinson, 1975), the log-normal distribution is by far still the most widely used distribution for conditional income distributions, especially when only used implicitly like for tobit regression (e.g. Card et al., 2013). Next to the standard normal we consider the gamma distribution, which is another frequently used two-parameter distribution in the aggregate distribution literature. The other two distributions which we consider belong to the distributions of the beta-type which are members of the Pearson system (see Kleiber and Kotz, 2003). In specific we consider the two the Dagum distribution (Dagum, 1977) as well as the Singh-Maddala distribution (Singh and Maddala, 1976). While these two distributions do not belong to the family of exponential distributions and thus cannot be modelled by standard mean regression, like generalised linear models or generalised additive models, they are found to provide a much improved fit to the unconditional income distribution. As McDonald (1984) points out, the more flexible Generalised Beta distribution of Second Kind with 4 parameters or the even the 5 parameter Generalised Beta distribution might provide an even better fit. However, while for unconditional income distributions the additional parameter estimation required may still prove feasible, we found this to be no longer the case for CIDs such that we have to restrict ourselves to sparsely defined distributions to a much greater extent. [18] Hence, we only considered only the log-normal, gamma, Dagum and Singh-Maddala distribution. In a preliminary study of the quality

[18] These problems are already noted in Biewen and Jenkins (2005).

of fit of the four CIDs, the very coarse conditioning on three age levels (20-35, 36-45, 45-60), two education levels (without and with higher education) and East/West divide. For each subsample we compute the four parametric CID by maximum likelihood and compare it to the parametric distribution to the empirical distribution function. Using the Kullback-Leibler divergence to a kernel-density estimate[19] of the conditional income distribution we found the Dagum distribution to provide the best fit. Consequently we selected the Dagum distribution as the most appropriate CID. However since the choice of the is pivotal to the modelling of CIDs by structured additive distributional regression much more work must be done on this issue though.

[19]We used an automatic bandwidth selection from Sheather and Jones (1991) and an Epanechnikov kernel.

# D Other regression approaches

## D.1 Mean regression

Standard analysis with regard to conditional incomes focusses on the conditional mean and generally treats additional parameters as nuisance parameters. While additional aspects of the conditional income distribution like heteroskedasticity or varying shapes can be partially incorporated by the use of link functions, the flexibility remains constrained by the variability of only one predictor. Nonetheless, using GLM or GAM conditional income distributions of the exponential family (e.g. log-normal or gamma) can in principle be modelled. For reasons of comparison we fit an generalised additive model of the same form as for our GAMLSS equations:

$$\log(y) = s_1(age) + Hs_2(age) + Es_3(age) + HEs_4(age) + \varepsilon, \tag{19}$$

where $y$ denotes the truncated income-vector excluding all incomes up to 4,800€ and where the elements of $\varepsilon$ are assumed to be independently distributed following a distribution from the exponential family, e.g. the normal distribution centred around zero with a constant variance $\sigma^2$. The thusly estimated means and Theil indices for the whole CIDs are displayed in Figure 6. We can observe that the expected incomes are not obviously worse than in the GAMLSS approach. Nonetheless, aspects where scale and shape are important, like for the Theil index, we can see that the log-normal distribution misfits and thus portrays much poorer estimates for the Theil indices of CIDs.

## D.2 Mixed models

The inclusion of random effects in the distribution allows for random intercepts and/or random slopes for different groups. Yet one core requirement for modelling CIDs is the ability to use continuous variables, like age as explanatory factors influencing the distribution. As random effects can only be assigned to a finite number of groups and thus not in accordance with a continuous variable mixed models are not considered in detail here.

## D.3 Other Distribution regression

A third alternative for the modelling of CIDs is the new class of so called distribution regressions (see Chernozhukov et al., 2013, p.10). It stems from the literature of survival functions and relies on link functions to model the conditional cumulative distribution function. Further in-depth

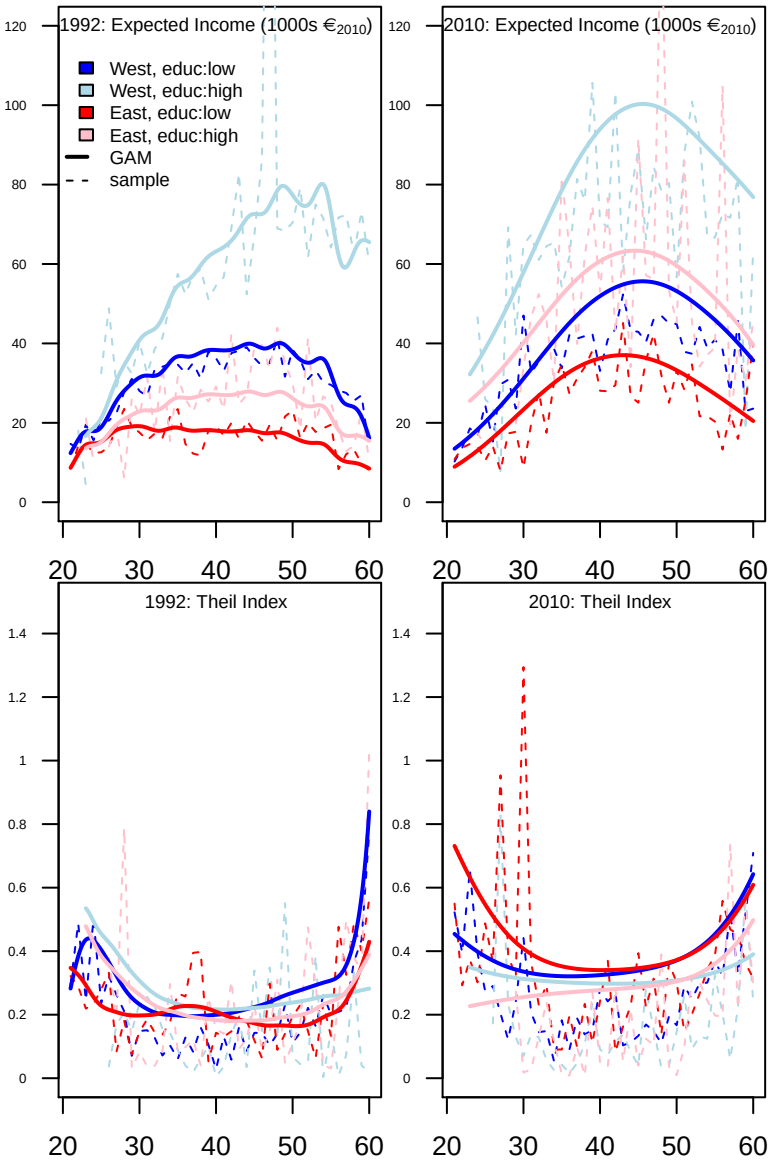


Figure 6: Means and Theil Indices of whole CID obtained by GAM

comparisons with this model class is needed. Nonetheless, in principle the flexibility of the CID is constrained as only one predictor is used, similarly to the case of mean regression.

## D.4 Quantile Regression

Quantile Regression, pioneered by Koenker and Bassett (1978) allows for the estimation of conditional quantiles. In principle we are thus able to estimate any quantile $q_\tau$ of $Y$:

$$y = s_{1,\tau}(age) + Hs_{2,\tau}(age) + Es_{3a,\tau}(age) + HEs_{4,\tau}(age) + \varepsilon_\tau, \tag{20}$$

While this alternative, which has been used for the analysis of income distributions by Machado and Mata (2005), offers a non-parametric alternative to the estimation of CIDs, some possible problems with such an approach should be pointed out.

It is the nature of quantile distribution that it is intrinsically focussed towards the estimation of specific quantiles within the conditional distribution rather than the conditional distribution at large. While for some purposes this bears several advantages (for example interquantile ranges are estimated with greater ease and less assumptions) it can prove a disadvantage if a comprehensive analysis of the CIDs is required. If enough data is available the problem is solely of computational nature as a large number of conditional quantiles can be estimated, constructing the conditional distribution thereof. Yet, as is generally the case, the scarcity of data is a major concern in which case quantile regression as well as any non-parametric approximation becomes highly unstable. The constraints that a parametric approach imposes can in that case aide to get a better approximation of the CID, if and only if the imposed constraints are applicable.

Connected to this aspect, is the role of censored data. As Bach et al. (2009) point out, the SOEP under-represents the top incomes. While it is in principle possible to account for this censored data in non-parametric approaches as well, the same argument as before applies, that a parametric approach can lend structure to the appropriate modelling of censored data. While the non-parametric approach employing quantile regression thus offers several advantages of GAMLSS, most notably the greater flexibility in modelling the CID, this flexibility can especially for small sample sizes hamper the estimation process as it fails to lend the required stability. Whether this additional flexibility provided by our parametric approach is of more use than harm, hinges on the question of whether an appropriate parametric form for the CIDs can be found.

# E The truncated CID

## E.1 Auxiliary measures for interpretation of the truncated CID

Figures 7 and 8 display several standard measures for size distributions, namely the expectation of the conditional distribution, its standard deviation and skewness as well as the Theil Index which incorporates all three moments and is a well known measure for inequality. The thick lines display the resulting measure from the GAMLSS estimation of the conditional distribution, while the dashed lines display the measure obtained directly from the sample for a given age, education level and region. It is well noted in the literature that the sample measures for moments are biased, especially for samples resulting from distributions other than the normal (Joanes and Gill, 1998, see). While we have accounting for this sampling bias in the standard manner[20] it is likely that some bias remains. Interestingly, in 1992 the skill premium in East Germany is relatively small such that on average incomes of men with a degree in the East are still lower than those of men without higher education in the West. The distributions' standard deviations show higher dispersion with higher age and rising mean incomes. We can also see that from the late thirties onwards the dispersion of men from the East with higher education exceed those without higher education in the West. Even more striking are differences in the skewness, which show that especially in later years the higher spread is anything but symmetric but largely caused by some high incomes. In comparison the skewness of the other distributions seems negligible. Nonetheless, we can observe a similar trend as for the standard deviation, i.e. increasing skewness as age and incomes progress. It should also be noted that skewness is systematically higher in the East for both education levels. This indicates that in the East more incomes are notably clustered at the lower end of the income range.The resultant findings for the Theil index for these CIDs show that the thus measured inequality is greater in the East than in the West, a finding which is contrary to the popular perception. However, as pointed out above, the nature of this higher inequality is such that a large share of incomes is clustered within the lower range of incomes with few very high incomes. While the perception of the income distribution which is probably mainly driven by the dispersion of the inner quantiles (e.g. $15^{\text{th}}$-$85^{\text{th}}$), i.e. that of the 'ordinary man', the nature of income inequality in the East is very different to that of the West (see below) as well as the inter-group differences, which we do not discuss in detail here.

However, the truncated distribution we have considered so far is only telling part of the story. As a matter of fact, much of the literature on income inequality is hampered by such a partial perspective, where only parts of the population (only males, full-time employees, etc.) and/or only

[20]For the standard deviation we have used the adjustment proposed by the cov.wt function from the stats package in R. For the skewness we have used the sample adjustment which is found for weighted data in the statistical software SAS. For the Theil index we have refrained from correcting the bias, following the ineq package in R.

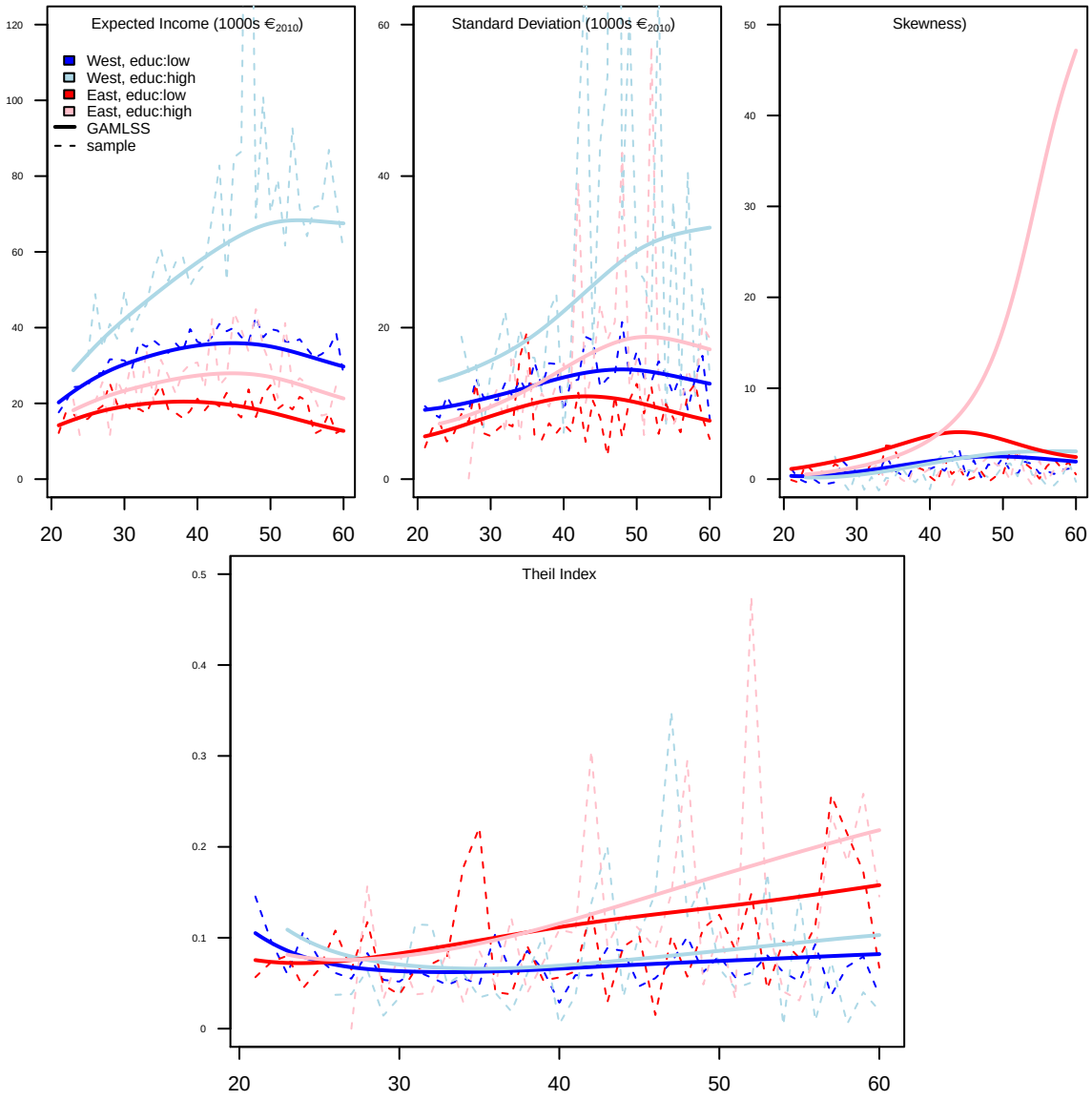


Figure 7: Measures for truncated CIDs males in 1992

parts of the income range (only incomes of people in employment, i.e. those above zero). While such partial perspectives provide scientific insight on their own a comprehensive account of income inequality must consider the whole income distribution, despite the analytical problems associated with such an approach.

The most noticeable difference of the expectation for the dependent income distributions displayed on the top left of Figure 8 is the surpassing of the average incomes of men with higher education in the East of those without higher education in the West. Also the rift of incomes between East and West for men without higher education has also narrowed. This convergence in mean incomes already pointed out by Vollmer et al. (2013) and others. Especially for the older generations in the East the increase in income dispersion is dramatic. For those with higher education this is

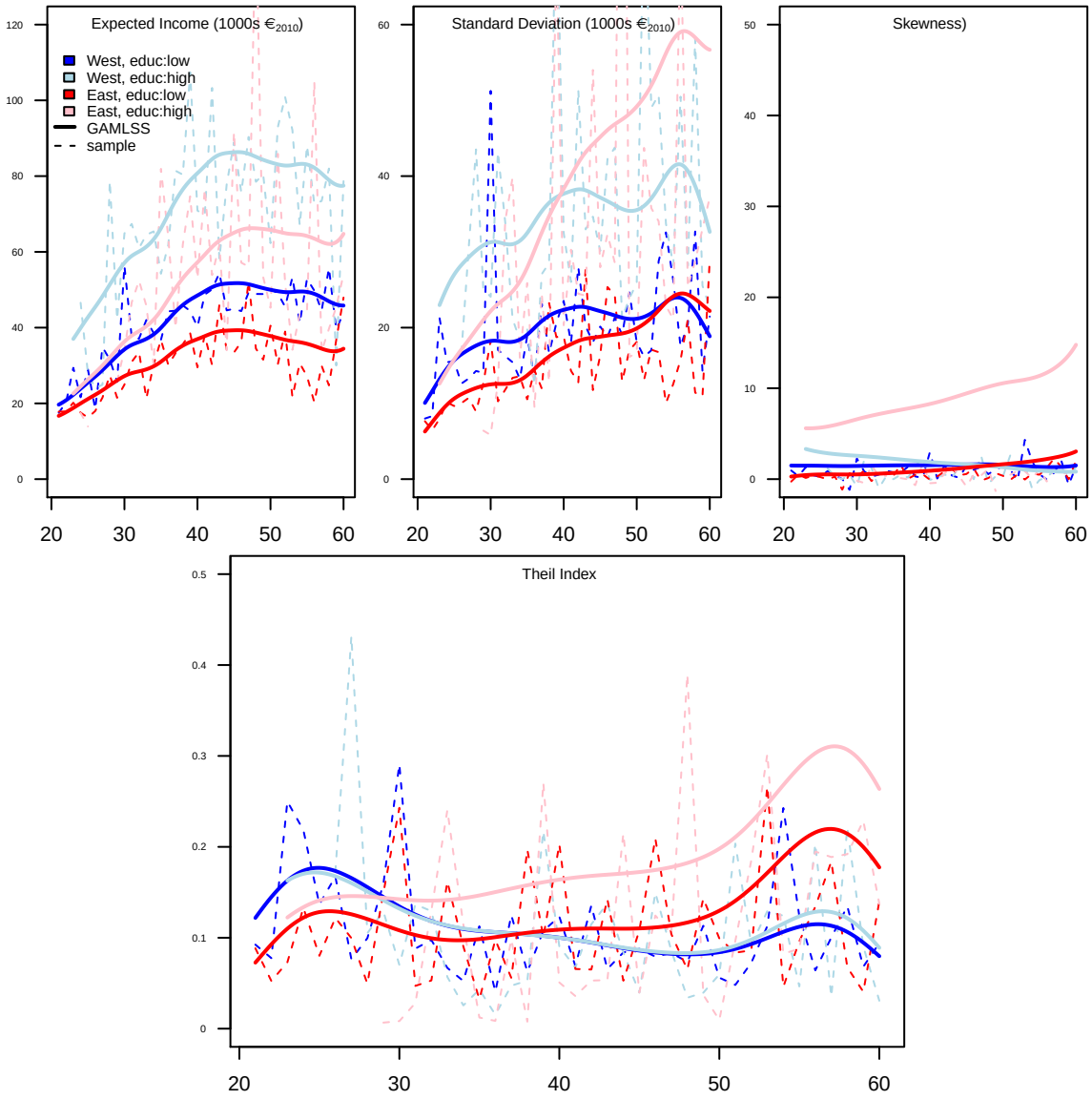


Figure 8: Measures for truncated CIDs of males in 2010

combined with a persistent high skewness again causing this group to show high within-inequality as indicated by the Theil index. With respect to the convergence of the whole income distribution (rather than just the mean) one has to concede that true convergence is still some way off. For the truncated income distribution, mean convergence across the age groups is driven by the catching up of incomes for some with a large part of the male population in the East left wanting as indicated by the higher standard deviation and skewness.

# F Conditional income distributions in Germany

In this section, we display our estimation results for the estimation of the CIDs with respect to the explanatory variables for the years 1992 and 2010. The results are then presented in the following manner: First we display the parameter estimates. As pointed out in Section 2.1, we model income distributions in a two-step procedure, such that point masses for zero-incomes and the truncated low incomes are truncated from the rest of the income distribution which is considered by distributional regression. This implies that we have five parameters for each CID (2 for the point masses and three for the Dagum distribution which is used to model the truncated CID). As these estimates can only partly be interpreted with regard to income inequality, we subsequently display some auxiliary measures of the resultant estimated conditional income distributions.

## F.1 Whole conditional income distributions of males in 1992

Figure 9 displays the parameters estimates.[21] From the parameter estimates for zero-incomes we can observe the expected pattern of an inverted U-shape for all four education/region combinations, as no-employment situations are more frequent at a young age during or directly after education as well as towards the end of the age span as retirement sets in. For precarious incomes we generally observe a decline over the age-span as extremely low-paid or part-time occupations tend to cease towards the later stages of the life-span. In addition there Note that the probability seems to change over age and different education levels. Combining these two measures our estimates thus portray a picture whereby the probability mass below 4,800€ generally also follows a U-shape.

As mentioned above, the direct interpretation of the parameters of the Dagum distribution is intricate. Yet it is possible to use the parameters to calculate auxiliary measures of interest, like the conditional mean. While the analysis of the income distribution above our truncation, i.e. the conditional Dagum distribution, is of interest itself, we will not analyse it in detail here.[22] Instead we will go on to discuss some auxiliary measures of the whole CID, which entails both the conditional dagum distribution as well as the two point masses which are truncated.

Figure 10 displays six measures on the whole conditional distributions. The thick lines display the resulting measure from the distributional regression estimation of the conditional distribution, while the dashed lines display the measure obtained directly from the sample for a given age, education level and region. Note that we do not display confidence bands for our estimates **warum eigentlich nicht?**. The results displayed here should thus be regarded as rather preliminary and

[21] For higher education we only display the parameters from the age of 23 onwards as beforehand very few students are likely to have completed their higher education degree.

[22] Some auxiliary measures are provided in the appendix.

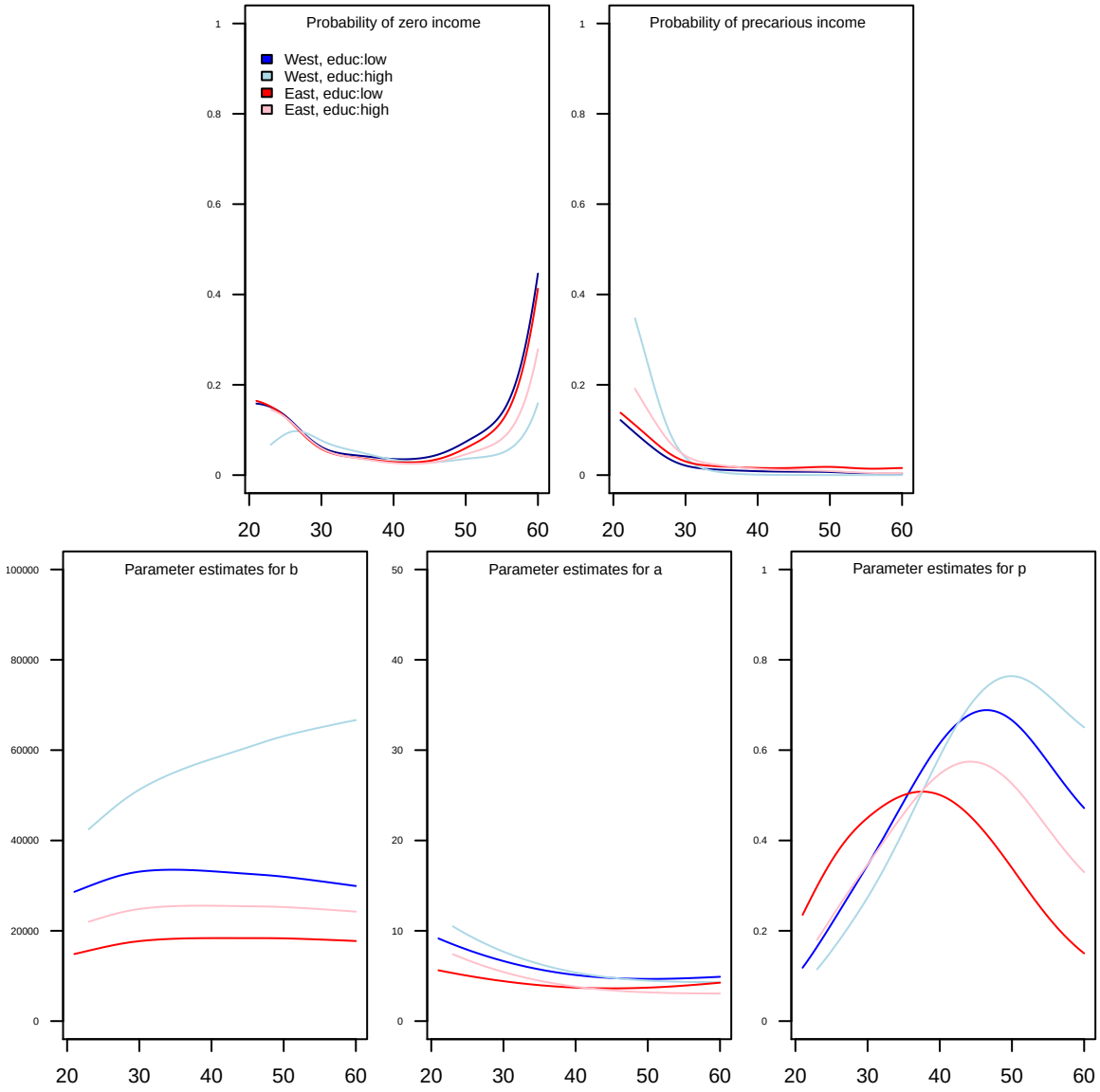


Figure 9: Parameter estimates for Males in 1992 as functions of age

of a rather exploratory nature only.

On the top left we display the conditional mean incomes. The general results are little surprising showing the positive relation between education and mean income, age and mean income as well as higher mean incomes in the West. This analysis is standard and it suffices to note that the comparison with the conditional means from the sample indicates that we are able to model the conditional mean using CIDs. Naturally, the estimation procedure is more cumbersome and requires more (possibly erroneous) assumptions than standard mean regression and is hence not ideal for the analysis of conditional means. Especially the parametric assumption of the conditional distribution can (and in fact does) lead to some bias in the estimation. See Section C in the appendix for more details. However, this assumption also allows us to analyse additional aspects

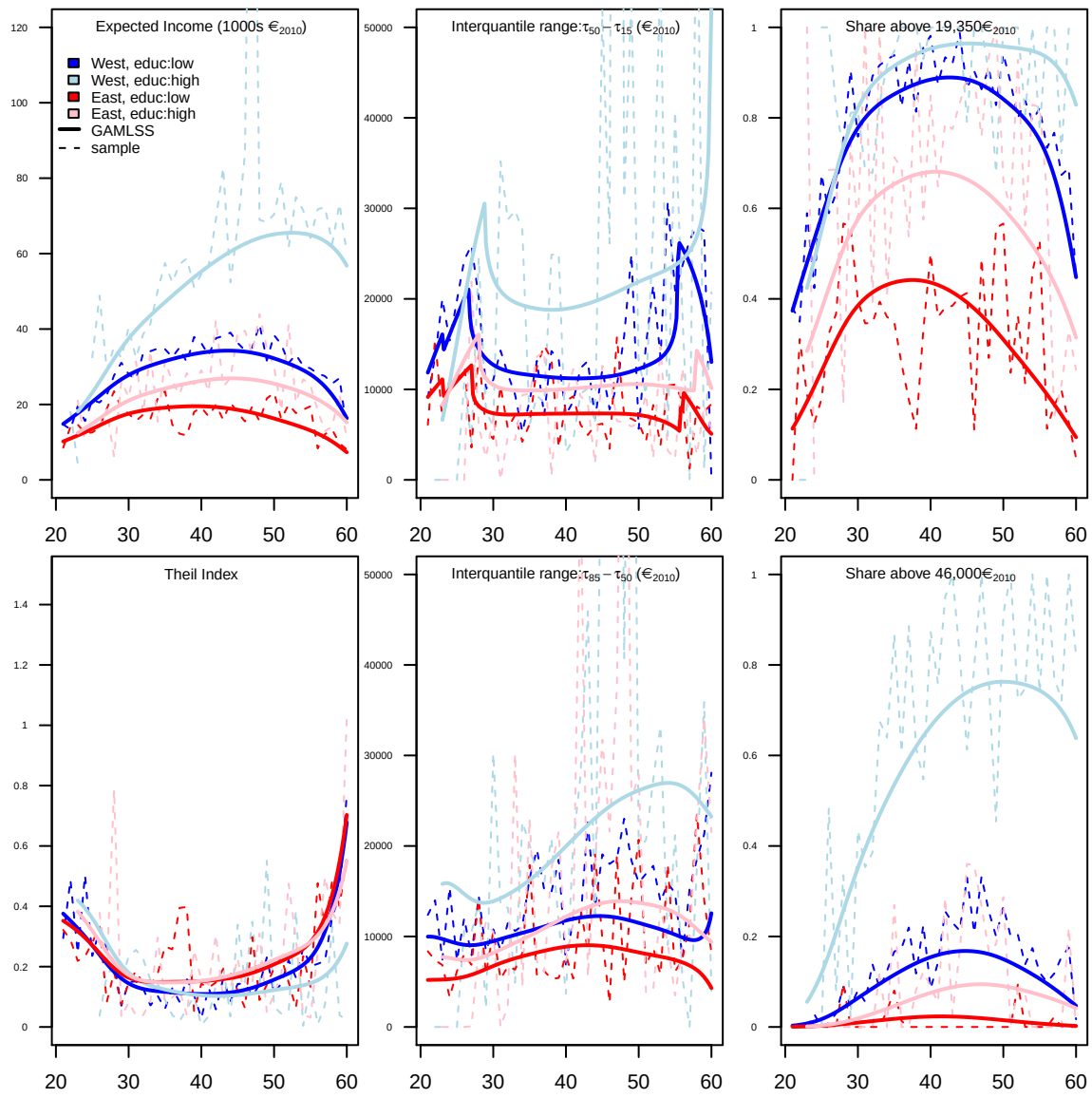


Figure 10: Measures for whole CIDs of males in 1992 as functions of age

of the conditional income distribution, three of which we will turn to now.

As a first measure we display our estimates for the within-group inequality as measured by the Theil Index. As can be observed we generally have a U-shape which is little surprising given the stark differences due to large parts of the subpopulations in education and retirement for the two extremes of the age-range respectively. Again the comparison with the estimates obtained directly from the sample we can observe that our estimates based on the Dagum distribution generally seem to provide an adequate fit not only for the conditional mean but also for the inequality within the finely defined subpopulations.

The second measure of inequality which we display are two interquantile ranges, namely the differ-

ence between the 50$^{th}$ and the 15$^{th}$ percentile as well as the difference between the 85$^{th}$ and the 50$^{th}$ percentile. Such a measure excludes the developments in the extreme ends of the income distribution, thus concentrating on the core income range of a subpopulation. Although other methods are better tailored towards this estimation (see Section D.4 in the appendix) our estimates show that they do resemble the empirical findings from the sample. Concerning the general trend we see that little surprisingly the both interquantile ranges are highest for the well paid group of highly educated men in the West and lowest for the group of men without higher education in the East. However, while for the former there seems to be considerable change over the age-range, the latter has two rather flat interquantile ranges. It may also be noted that the second interquantile range, i.e. $\tau_{85} - \tau_{50}$ is generally greater than the $\tau_{50} - \tau_{15}$ interquantile range for a given subgroup. The reason for this is mostly found in the nature of the Dagum distribution which is right skewed. The exceptions to this pattern, like for example the high difference between the 50$^{th}$ and the 15$^{th}$ quantile of men in the West above 50 years of age, are mostly driven by high zero- or precarious incomes for substantial parts of the subpopulation.

Next to these two inequality measures, which by definition are relative concepts, we will also provide an "absolute" income distribution measure. Thereby we use the c.d.f. to determine the shares of the subpopulation with an income greater than the following thresholds: 0, 19,350€[23] and 46,000€[24]. Since the first was already provided in Figure 9, we will only displayed the latter two. Generally, the results seem to resemble the development of the conditional mean income although the changes over the age span are much more pronounced, which can be explained by the development of within-group inequality over this dimension, which as was pointed out earlier follows a U-shape.

## F.2 Conditional income distributions of males in 2010

Contrary to the previous analysis we will now first and foremost concentrate on the inter-temporal differences rather than the intra-temporal differences.

Looking at the parameters we can observe that the share of people with zero-income has seemingly increased dramatically for young men without higher education in the East. Similarly for precarious incomes we also get a starkly risen estimated share of incomes. This alone shows a substantial change in the shape of some conditional income distributions, which cannot be grasped by sole mean income analysis. One the other side of the age-range we can observe that the share of people

[23]This is the annual gross market income as would be obtained if a German full-time employee (35 hours) would be paid at the level of the French minimum wage (salaire minimum interprofessionel de croissance) for the year 2010.

[24]This is the amount of money which Keynes ascribes to be enough to turn the human mind away from pecuniary requirements (see Skidelsky, 2010, p.142).

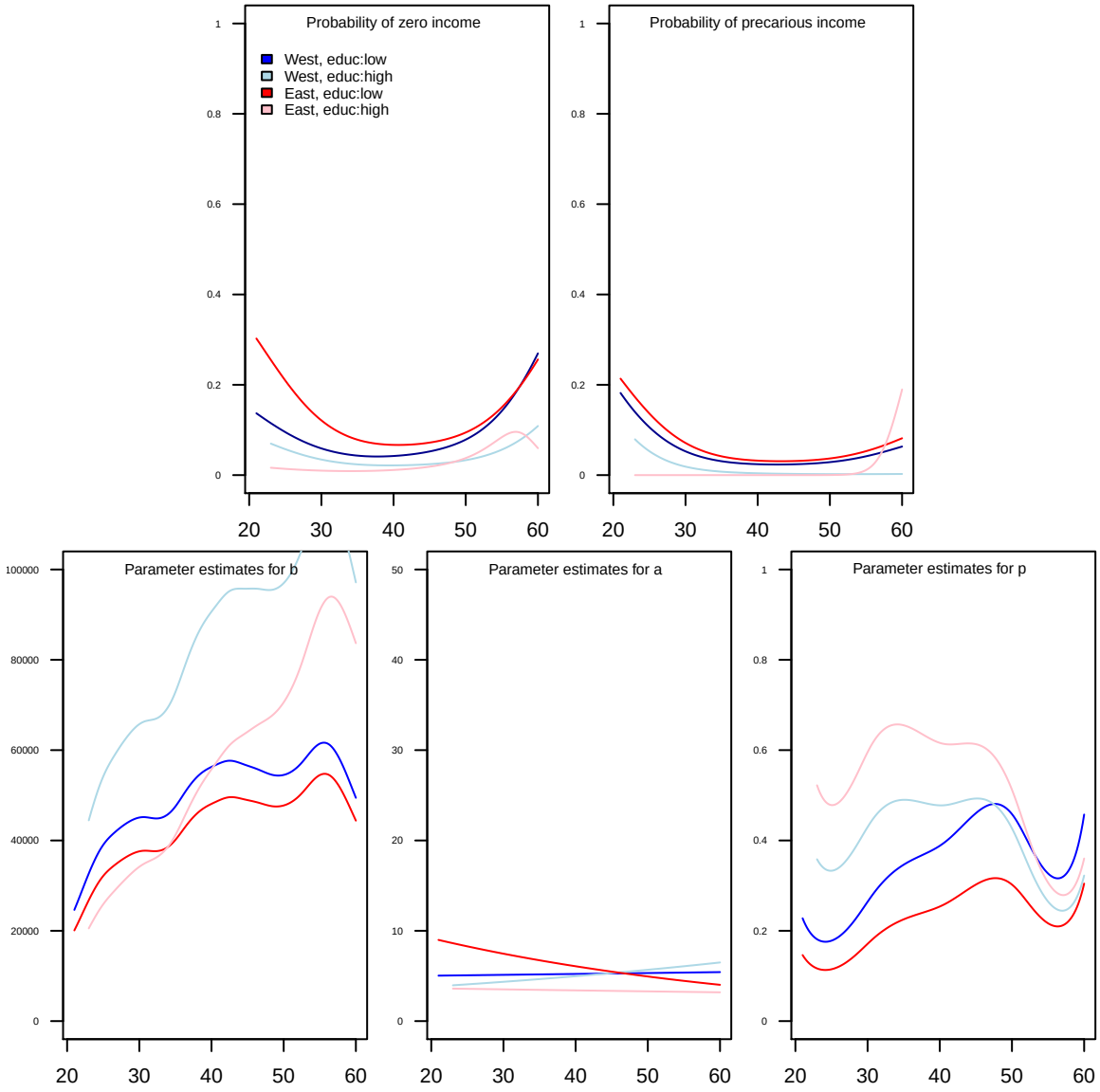


Figure 11: Parameter estimates for Males in 2010 as functions of age

with zero income has decline for all four education-region combinations, which is most likely driven by the increased standard retirement age and consequently later use of early retirement schemes. For the parameters of the Dagum distribution we again refer to Section E.1 in the appendix and go on to the auxiliary measures for the whole CID.

Analogously to 1992 we observe the same positive relation between both living in the West and having higher education with income. A comparison with the means from the sample shows that our measure fits quite well, which is even better than for 1992. In fact it should be noted that the results from the Kolmogorov-Smirnov test indicate that the fit of the Dagum distribution in 2010 is even slightly better than in 1992, where the rejection of the null was slightly higher.

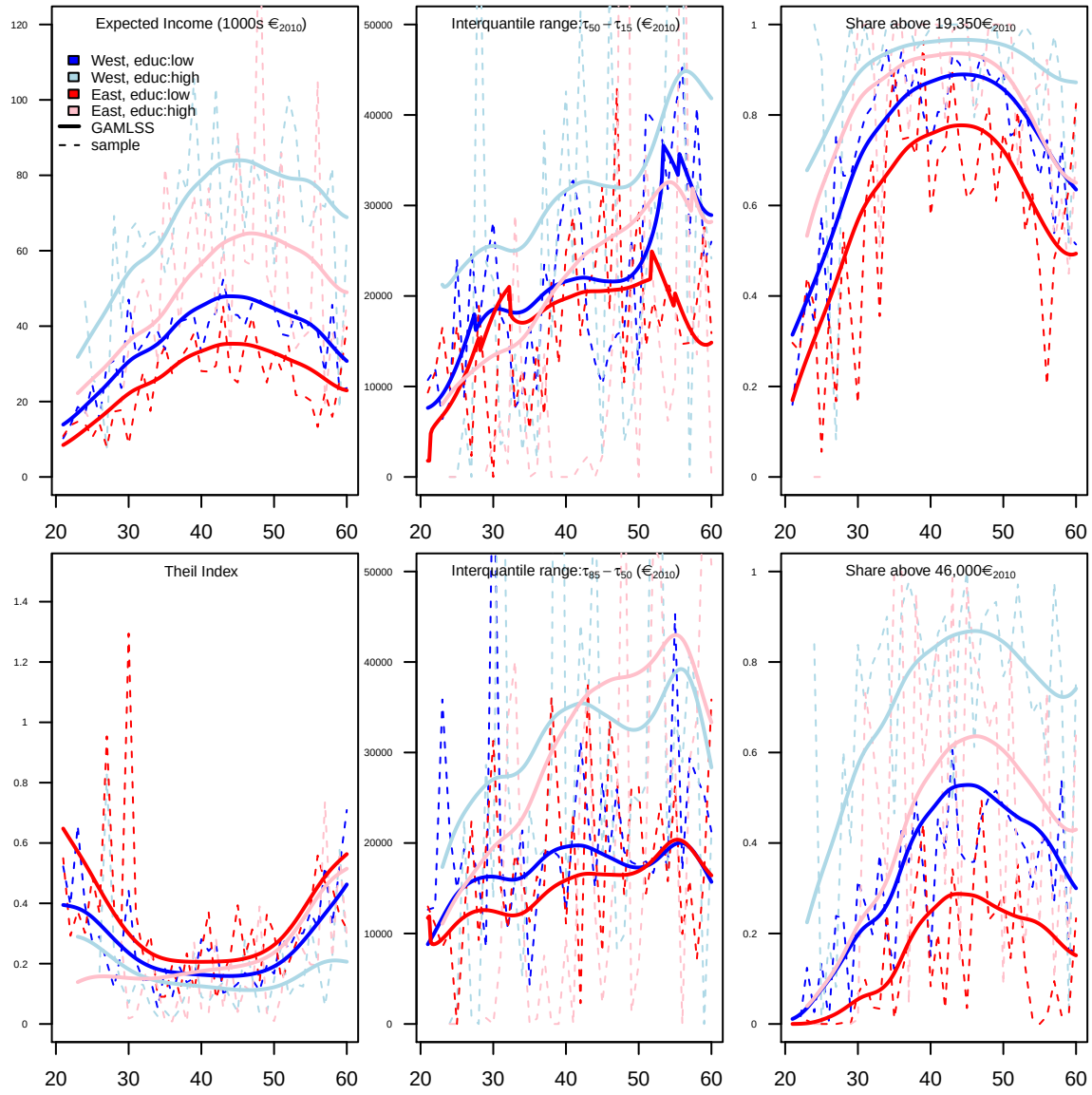


Figure 12: Measures for whole CIDs of males in 2010 as functions of age

For the Theil index, we also find generally a U-shaped over the age-span with men from the East with higher education forming the exception. Again the estimates derived by the CIDs fit those obtained directly form the sample.

The interquantile ranges of the sample are very volatile making a comparison with our estimates difficult. However, again the fit seems to represent the data with most of the differences between the sample and the CID driven estimates attributed to the smoothing. The most striking feature of our estimates is the starkly risen $\tau_{85} - \tau_{50}$ difference which rises dramatically over the age-range even surpassing the corresponding interquantile range of men in the West with higher education. This shows that the greater Theil index observed for these groups is not solely driven by a heavy right hand tail but rather a different shape and dispersion in the core of the CID.

Concerning the thresholds, we again see a strong resemblance of the expected income, although especially for the first threshold, i.e. 19,350€, the differences between the four education-region groups are much less pronounced for most points in the age range.

It should be noted that all these observations can only be interpreted as expected parameter estimates, which naturally have various sources of uncertainty and possibly bias attached to them.