Article

Distributional Regression Techniques in Socioeconomic Research on the Inequality of Health with an Application on the Relationship between Mental Health and Income

Alexander Silbersdorff [1],* and Kai Sebastian Schneider [2]

[1] Economics Faculty, Georg-August-Universität Göttingen, 37073 Göttingen, Germany
[2] Department of Clinical Psychology, PFH Private University of Applied Sciences, 37073 Göttingen, Germany; kschneider@pfh.de
*   Correspondence: asilbersdorff@uni-goettingen.de

check for updates

Abstract: This study addresses the much-discussed issue of the relationship between health and income. In particular, it focuses on the relation between mental health and household income by using generalized additive models of location, scale and shape and thus employing a distributional perspective. Furthermore, this study aims to give guidelines to applied researchers interested in taking a distributional perspective on health inequalities. In our analysis we use cross-sectional data of the German socioeconomic Panel (SOEP). We find that when not only looking at the expected mental health score of an individual but also at other distributional aspects, like the risk of moderate and severe mental illness, that the relationship between income and mental health is much more pronounced. We thus show that taking a distributional perspective, can add to and indeed enrich the mostly mean-based assessment of existent health inequalities.

1. Introduction

The relationship between income and health is one of the most widely researched issues in health economics and epidemiology, and has seen countless articles discussing its nature. To this end, various health measures have been employed from objective measures, such as life expectancy [1] and physiological outcomes [2,3] to subjective measures, such as single-item measures [4,5] or composite health scores, like the SF-12 [6] (The SF-12 [7] is a short version of the original SF36 [8]). In terms of the statistical methodology, we can equally observe an array of approaches.

1.1. A Review of the Methodological Literature

Most of the heavy empirical ploughing has been carried out by two kinds of workhorse methods. On the one hand, many studies have employed concentration curves and indices [9–12]. On the other hand, health outcomes are contrasted between different groups with varying incomes (and potentially other covariates) by means of relative frequencies [13], odds ratios [14], or other effect sizes derived from regression estimates [15,16]. This is mostly done by employing conventional regression approaches from the framework of generalised linear models.

Recent years have seen numerous methodological advances, while improved data availability and increased computational capacities have made their application to the analysis of socioeconomic inequalities in health feasible. For concentration curves and indices, the use of level-based concentration curves, instead of rank-dependent ones, have been proposed [17]. In the statistical literature, numerous regression techniques have been developed that go beyond the still-dominant standard generalised linear models in the epidemiological literature, like quantile regression [18], expectile regression [19], conditional transformation models/distribution regression [20,21], recentered influence functions [22], and generalised additive models of location, scale, and shape/structured additive distributional regression models [23,24]. This set of approaches that contemplate not only the conditional expectation but further distributional aspects has already found its ways in the literature on socioeconomic inequalities in health [25–28], but much of the potential of the application of these methodological advances is still dormant.

1.2. A Review of the Literature on Income and Mental Health

A lot of research exists on the relationship between income and mental health, and this section aims to briefly summarize the important findings. One should keep in mind that the research in the field differs in important aspects, such as in the used definitions of mental health (i.e., the presence of mental disorders, continuous scores for constructs like life satisfaction and emotional well-being), the operationalization of income (absolute income, relative income, changes in income, income subsumed in variables like socioeconomic status), or in the subject of interest (individuals or groups).

The prevalence of mental disorders is found to be higher among individuals with lower income [29,30] or among those with lower socioeconomic status [31–33]. Equivalently, increased odds for mental disorders are found for individuals with lower levels of income [14] or socioeconomic status [13,32,34–36]. Wirtz et al. [37] reported that a higher income is associated with higher MCS scores (higher MCS scores indicated good mental health, and vice versa), whereby MCS scores are a composite mental health score which we also use in this publication (see Section 2). Additionally, Wood et al. [38] showed that low income predicts mental distress since it functions as an indirect proxy for social rank. McMillan et al. [39] did not find a relation between income and mood or anxiety disorder, but did find one with drug abuse and suicide attempts. The latter was also found by Lee et al. [40]. Sareen et al. [41] revealed that having a low income was a financial barrier for accessing mental health services.

Weich and Lewis [42] found that unemployment and measures of poverty (which were built amongst other variables with the help of household income and savings) had predictive ability for the prevalence of mental disorders, but ongoing subjective financial strain seemed to be an even better predictor for the maintenance and onset of mental illness. Lorant et al. [43] researched the direction of effects, and found that income reductions, as well as increasing financial strain, increased the risk of showing depressive symptoms. On the other hand, increasing income, as well as reduced financial strain seemed not to be related to decreasing depressive symptoms.

Income inequality has also been found to be associated with mental illness. In a systematic review, Pickett and Wilkinson [44] published evidence for a strong association between income inequality and mental disorders, whereas a meta-analysis by Ribeiro et al. [45] revealed small effect sizes. Suicide could be an exception and seems to be more common in more equal societies [46].

Lucas and Schimmack [47] summarized that the correlation between subjective well-being and income is often found to be very small, yet small correlation coefficients can turn into large mean differences when comparing different levels of income. Matz et al. [48] found that the degree of matching between an individual's spendings and his/her personality had better predictive ability for life satisfaction than income. Even though these effects were found to be statistically significant, the effects may be too small to be relevant [49]. Kahneman and Deaton [15] found that life satisfaction and emotional well-being seemed to rise with increasing income, with emotional well-being satiating at

an annual household income of 75,000$ (USA). Additionally, Kushlev et al. [16] reported that a higher income was related with less daily reported feelings of sadness, but not with more daily reported feelings of happiness. Westerhof and Keyes [50] found significant effects when predicting mental illness with the help of income and other socio-demographic variables.

Boyce et al. [51] found the income rank to be more important for the evaluation of life satisfaction compared to absolute income. Yu and Chen [52] found that relative, as well as absolute income was related with negative aspects of emotional well-being, whereas only relative income was related to positive mental well-being. In contradiction to these findings, Sacks et al. [53] provided a review in which they stated that subjective well-being rose with increasing income. According to the authors, this holds for within-country (individual level) and between-country comparisons. The satiation thesis is questioned in this study, and absolute income is identified as more important than relative income.

Overall, we find that there is strong evidence for the assumption that income is related to constructs related to poor mental health, as well as constructs related to good mental health. The current state of research is characterized by disunity about the working mechanisms of the effects of income. On the other hand, there is high methodological cohesion which either sees assessments on the basis of ranks, or is based on arithmetic means.

1.3. Aims and Structure of the Paper

This paper aims to add to the young and evolving branch of literature of distributional regression approaches in health economics in two important ways. Firstly, it considers the relationship between income and mental health in a distributional framework. Secondly, it constitutes an exemplary application of the frequentist framework of generalised additive models of location, scale, and shape to the analysis of socioeconomic inequalities in mental health that is designed to guide applied researchers in the use of this powerful, but in some ways, also challenging statistical methodology. In this publication, we focus on the frequentist framework of generalised additive models of location scale and shape. Readers interested in the Bayesian approach of structured additive distributional regression are referred to Klein et al. [24], Silbersdorff [54], and Silbersdorff et al. [27].

The remainder of this publication is structured as follows: Firstly, we explain the mental health concept, variables, and data on which our analysis relies. Then we go on to provide a general discussion on the use of a distributional regression framework and contrast it to the conventional regression approach. In Section 3, we consider the association between mental health and income found in our data. Lastly, we draw conclusions obtained by our analysis. We thereby outline which findings are particularly generated by taking a distributional perspective.

2. Materials and Methods

2.1. Data

To investigate the relationship between mental health and income, we used data from the German Socio-Econocmic Panel (SOEP) [55]. The SOEP is a "[...] widely used long-running household panel study that seeks to provide a representative view of the entire population [...]" ([55], p. 1). We used only cross-sectional data for the year 2014. The survey for 2014 contained data from 16,037 households [56]. The SOEP data contains a large array of socio-demographic information and various income variables, as well as variables on individual health. Concerning the socio-demographic variables used in our analysis, we followed Silbersdorff et al. [27] and considered the age, marital status, nationality, educational attainment, and place of residence, as well as the household income. Additionally, we incorporated the employment status, as well as the place of living in terms of urbanity in our analysis, since both variables are common in the field of mental health research (i.e., [34,57–62]). Further information on these variables, as well as their SOEP identifiers, are provided in Appendix A.1.

In contrast to Silbersdorff et al. [27], we considered a mental health score, rather than a physical health score as the outcome variable. Specifically, we used the Mental Component Scale (MCS) score

obtained from the SF-12v2 questionnaire, which also contains the Physical Component Scale (PCS) score used by Silbersdorff et al. [27]. We see this measure as well-suited for our purpose for several reasons. Firstly, MCS scores claim to cover the whole spectrum of self-rated mental health—the MCS scores are designed to have a mean of 50 and a standard deviation of 10 (norm population), where higher scores indicate better mental health conditions compared to lower values [63]. According to Ware et al. [8] (p. 72), very high MCS scores represent an "absence of psychological distress and limitations in usual social/role activities due to emotional problems; (mental) health rated excellent", whereas very low MCS scores represent "frequent psychological distress; substantial social and role disability due to emotional problems; (mental) health in general rated poor". Secondly, the MCS scores are a composite score of several items covering the concepts of mental health, social functioning, vitality, and emotional aspects [64], and therefore cover mental health in its multifaceted nature. Thirdly, the MCS scale has good test-theoretical properties. Its internal consistency can be evaluated as acceptable [37], while its convergent and discriminatory validity, as well as its reliability can be evaluated as good [37,65,66]. Lastly, despite the ongoing discussions concerning the definition and measurement of mental health [67–72], we see this measure in line with the widely accepted definition of the World Health Organization (WHO) [73]: "mental health is a state of well-being in which every individual realises his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community". For further information on the MCS scale, also see Appendix A.1.

It should be noted at the outset that differential item functioning by education, age, and sex have been observed for the mental component score compared to the physical component score [74,75]. We condition on these variables to address this drawback.

A drawback of the data gathered by the SOEP is that it fails to include the institutionalized population. Therefore, the results cannot be seen as representative of the whole German population, but only of those outside any institutions—meaning that the results we portray with respect to the risk of very low self-rated mental health are presumably underestimated to some degree.

However, despite this drawback, we deem the analyses on socioeconomic inequalities with respect to mental health to be too important to be left void due to data-related deficiencies, and thus will pursue the following analyses on the basis of this somewhat imperfect data.

2.2. Conditional Health Assessment beyond the Mean

In order to facilitate the understanding of the distributional regression approach, which we propose in this publication, we aim to contrast it to the classical mean regression in an intuitive manner.

Let us consider the conditional health distribution for the simplified case, where we only regress MCS on the household income and omit other variables for the sake of simplicity.

In Figure 1, we display the empirically observed conditional mental health distribution for both men and women with a household income of around 15,000 Euro and 30,000 Euro, respectively, with the help of histograms, and contrast it with the obtained estimates from standard mean regression and distributional regression techniques. While the portrayed estimates are estimated for exactly 15,000 Euro and 30,000 Euro, the portrayed empirical distributions are drawn from individuals with incomes in between 12,500 Euro and 17,500 Euro, as well as 27,500 Euro and 32,500 Euro, respectively.

Classical mean regression—including generalised linear models (GLM)—yields information on the conditional arithmetic mean (displayed by the blue dot). This estimate is ideally suited for a comparison of the expected health outcome. In our simplified example, one could thus deduce the expected mental health score of an individual for any income level. From a mean regression-based analysis, we could thus infer that the expected mental health level with a low income is roughly 1.2 units lower than that of an individual with a high income. As is argued by Silbersdorff et al. [27], drawing a conclusion on the basis of this measure is somewhat problematic, as it gives equal emphasis to improving the health of the already healthy, rather than improving the health of the ill. This is

arguably problematic in any form of health-centered analysis, and particularly so in an analysis focused on health inequalities.
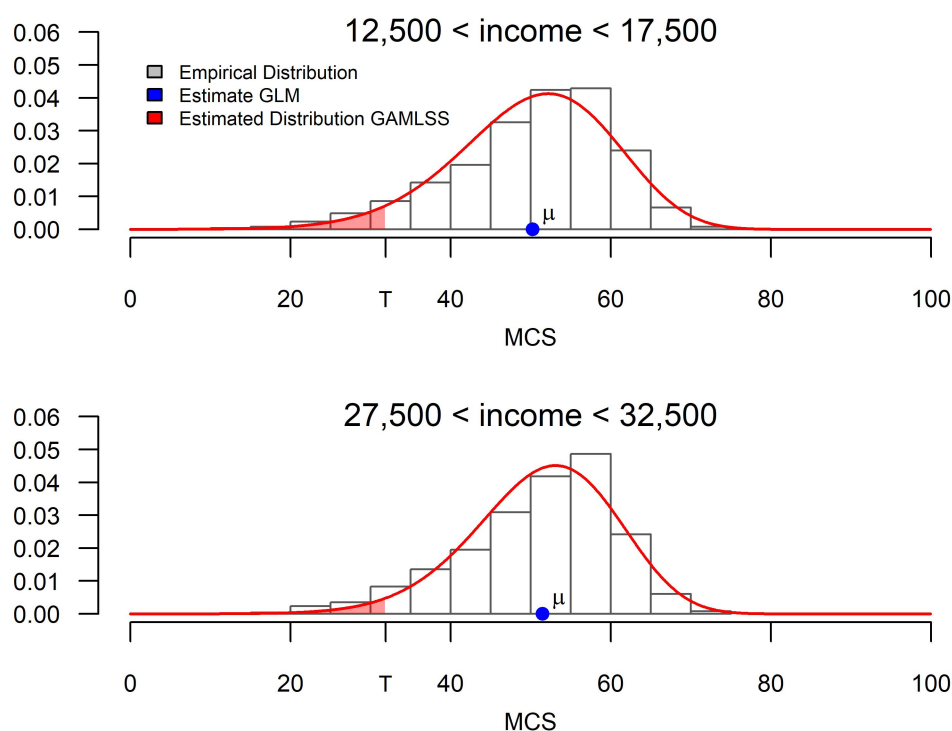

Figure 1. Comparison of estimates between GLM and GAMLSS.

Distributional regression, on the contrary, does not focus on a particular measure of the distribution, but directly aims to estimate the whole conditional response distribution for each group (displayed by the red function). From this conditional distribution, any desirable distribution measure can be calculated in principle. One can therefore explore the influence of an explanatory variable on any desired statistical distributional measures like the mean, variance, or skewness, as done by Silbersdorff et al. [27]. Moreover, we can consider the inequality and risk measures associated with that distribution. In our particular case of analysing health inequalities, we can deduce risk measures defined as the share of individuals falling below a certain threshold. For example, we could consider the risk of belonging to the lowest 5% of all individuals in the sample, which is displayed by the shaded red area underneath the distribution to the left of the point *T* (for threshold). Figure 1 shows that the shift in probability mass translates to a smaller probability for falling below the defined threshold for a person with high income. Furthermore, we can deduce the risk for any given income of falling below a globally defined threshold, which can be seen as the threshold to suffering from poor mental health.

While we consider only two risk measures in the following, it should be stressed that, in general, it is straightforward to use the estimated conditional mental health distributions to explore any number and kind of further risk measures—be they further thresholds or other distributional measures, like inequality indices. One major advantage of the distributional approach is thus the need for only one estimation process to estimate a host of distributional measures. This stands in contrast to equally feasible approaches to estimate each distributional measure individually that render an array of problems associated with simultaneous inference on several models [76].

The more comprehensive distributional perspective thus allows for a nuanced assessment of the income–health relationship which not only contemplates expected health but other distributional aspects as well, like risk measures. However, this naturally also implies a much greater model space that is accompanied by some pitfalls, just as the assessment on an array of separately estimated

models would be. In the following, we thus consider some vital components of the approach from a user's perspective, and give some guidelines for safe and sound usage. The underlying baseline recommendation for distributional regression users with limited or no experience is to commence your applied analysis by using simple models, rather than exploiting the extensive statistical artillery that has been made available over the past few years. The guidelines we give here are thus intended to provide a starting point for an applied researcher to use distributional regression in the context of health inequalities and who may want to get a first sense of the existence and magnitude of potential effects beyond the mean.) The estimation is done with the help of the GAMLSS-package [23] for the statistical software R [77].

2.3. Considering the Conditional Distribution Set-Up

In the distributional regression framework, the distribution of the response variable—here, the mental health distribution—is described by a distribution that is conditional on a set of explanatory socioeconomic and potentially further variables, i.e., $\mathcal{D}(Y \mid x_1, \ldots, x_K)$.

One assumes a parametric distribution, which parameter values depend on the explanatory variables, that is, $\mathcal{D}(\theta_1(x_1, \ldots, x_K), \theta_2(x_1, \ldots, x_K), \ldots, \theta_L(x_1, \ldots, x_K))$. Therefore, one relates its parameters $(\theta_1, \ldots, \theta_L)$ to the variables $(x_1, \ldots, x_K)$ via a regression predictor, which contains regression coefficients, in the form:

$$g_l(\theta_l) = \eta^{\theta_l}, \tag{1}$$

where θ_l denotes the *l*-th parameter of the distribution, g_l denotes the corresponding link function, and η^{θ_l} the predictor.

By relating all the parameters of the response distribution to a regression predictor, one models the response distribution conditional on the variables. After the estimation, the variables can be fixed at specific values to define groups that are of special interest. The resulting conditional response distributions can then be compared between the groups using different measures (i.e., risk measures, see Section 3.3.1) describing certain aspects of the conditional response distribution.

The analysis of socioeconomic inequalities using distributional regression requires the consideration of various modelling components, which we will discuss in the following.

2.3.1. The Conditional Health Distribution

The greatest strength, and potentially the greatest weakness of the distributional regression lies in the use of a parametric distribution for modelling the health distribution. If the selected parametric distribution provides a sufficiently good approximation to the response under consideration, the assumption will facilitate the estimation to a substantial degree. Choosing a distribution that provides a sufficiently close approximation to the response distribution has thus been found to provide more stability in the estimation process, meaning, among other things, higher estimation precision and smaller standard errors [78]. For the usually rather limited sample sizes available in health inequality research, this is of particular importance for the assessment of inequality in the tails, as data there is naturally very scarce. If, however, the selected distribution does not sufficiently approximate to the response distribution, the outcome will be very detrimental to the model fit, and may potentially yield misleading results. The choice of a good parametric distribution is thus key to any analysis employing a distributional regression framework.

One major problem with health scores like the MCS is that they follow a shape (negative skewness) that runs counter to most usual parametric distributions. Thus, Silbersdorff et al. [27] suggests the following linear transformation:

$$g_{MCS}(S) = S^* = \frac{S_0 - S}{S_{scale}}, \tag{2}$$

where S is short for score and denotes the untransformed MCS scores, and S^* denotes the transformed MCS scores. S_0 is a constant ensuring that S^* has positive support, and S_{scale} is a rescaling factor. Within this publication, $S_0 = 100$ and $S_{scale} = 10$ are used such that the distribution of S^* is positively skewed and the values are restricted to the interval $[0, 10]$. A contrast between the original and the transformed marginal distribution of the MCS score is shown in Figure 2.
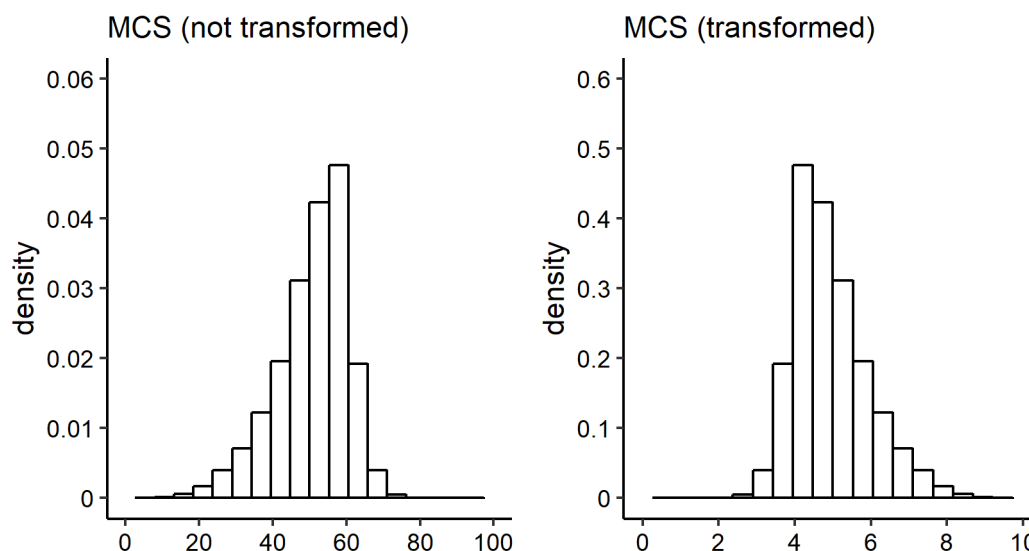

Figure 2. Histogram of MCS scores. Left: not transformed, right: transformed.

Using this transformation, we have done extensive comparisons based on information criteria and residual diagnostics on a host of potential distributions. A distribution is an option if it supports the permissible range of mental health values, here being $[0, 10]$. For an extensive account, see Appendix A.2. Here, we concentrate on three distributions—the gamma distribution (Ga) with two parameters linked to a regression predictor, the Box–Cox power exponential (BCPE) with three out of four parameters linked to a regression predictor, and the generalised beta distribution of the second kind (GB2) with all four parameters linked.

The comparison on the basis of various information criteria shows that the more complex three- and four-parameter distributions generally outperform the more simplistic gamma distribution with only two parameters.

Residual diagnostics reveal that this difference is due to the rigidity of the gamma distribution in the tails, which leads to inferior fits. Residual diagnostics plots are provided in Appendix A.3. However, the observed differences are minor, and all three distributions provide adequate fits in the sense that from the most flexible four-parameter distribution to the more rigid two-parameter distribution, all distributions yield similar results concerning the risk measures used for the assessment of socioeconomic inequalities in health.

Moreover, the more complex characteristics also feature various problems—first and foremost, there is decreased estimation stability due to the much increased model complexity. This not only leads to much wider confidence intervals, but also to some undesired statistical artifacts, like very high expected mental health outcomes for very low incomes.

With regard to the choice of the response distribution, we thus conclude that any of the three distributions provide a sufficiently good fit to model health outcomes in the distributional framework we employ here. While all three distributions are thus viable candidates to be used in the analysis of socioeconomic inequalities in health, we advise inexperienced users to stick to the simplest of the three distributions, namely the gamma distribution.

2.3.2. The Predictor Specification

The predictor specification defines the nature of the dissection of the population into the groups used for the analysis. In principle, all parameters are allowed to have separate predictors containing potentially completely different variables and structures (e.g., linear interaction terms, splines, local regression smoothers, ridge and lasso regression terms, neural networks, . . .) are implemented in the `GAMLSS` software [23]. However, by using simple linear predictors, the groups' differences are coerced to follow certain patterns that facilitate the estimation procedure, stabilize the estimation results, and potentially allow for a straightforward interpretation of the results. For the sake of simplicity, model comparability, and technical feasibility, we thus recommend using one generic straightforward linear predictor for all the distributions' parameters that takes the form:

$$\eta^{\theta_l} = \beta_0^{\theta_l} + \sum_{k=1}^{K} \beta_k^{\theta_l} x_k. \tag{3}$$

Even in this simplified predictor framework, it must be noted that the vector of all regression coefficients $\boldsymbol{\beta}$ entails parameters not only for one predictor, but for all L predictors required to specify the response distribution yielding $L \times (K+1)$ parameters, which can quickly yield a daunting model complexity for the kind of finite datasets and computational capacity that are available for research in health inequalities.

Thus, the use of distributional regression requires researchers to heed the advice of Box [79] to select variables economically and refrain from any form of "kitchen sink" regression, with the obvious difficulties for causal inference.

2.3.3. The Link Function and Other Technical Aspects

The link function g_l links the predictor to the parameters of the response distribution. It is first and foremost designed to ensure that the parameters are constrained to valid values. Moreover, they can transform the impact of the covariates on the parameter. While the effect of applying different link functions is far from negligible, we found that the default choices (see Table A1 of Appendix A.2) generally yielded reliable results (pending the quality of the distribution).

Equally, the additional inputs that are usually required (e.g., optimization options) are generally not critical, as long as one does not veer too far from the sensible options usually put down as a default, and as long as one sticks to simple two-parameter distributions and linear predictors. However, once the models get more intricate as many-parameter distributions or complex predictors are selected, the optimization routines will quickly run into the curse of dimensionality and require thoughtful usage.

Our bottom-line suggestion for the applied research on socioeconomic inequalities in health using distributional regression techniques is thus to keep things simple, in the sense that simple distributions (like the two-parameter gamma distribution) and simple predictors (like the linear predictor) should be employed. While more complex distributions and predictors may often yield models with better fit, the risk of technical or even inferential problems (if the models are not applied with the necessary care and background knowledge) does, in practice, outweigh the theoretical benefits.

The estimation of GAMLSS, as provided by the `GAMLSS` software [23] is based on the maximum likelihood principle [80]. For the sake of simplicity, assuming that no smoothing functions, random effects, or modeling techniques are present in the model, estimates are obtained by maximizing the log-likelihood defined by $\ell = \sum_{i=1}^{N} ln(f(y_i | \mu_i, \sigma_i, \nu_i, \tau_i))$, where each parameter is related to a regression predictor containing the parameters to be estimated [81]. For more complex modelling structures, estimates need to be obtained by maximizing a penalized log-likelihood given by $\ell_p = \ell - \frac{1}{2} \sum_{k=1}^{4} \sum_{j=1}^{J_k} \gamma_{kj}^T \mathbf{G}_{kj}(\lambda_{kj}) \gamma_{kj}$ [81]. The penalized log-likelihood is defined under the assumption that the model can be written as a random effect model with smoother $s(x) = \mathbf{Z}\gamma$, where $\mathbf{Z}$ is a design matrix constructed with the data (x) and γ is a parameter vector to be estimated under

the restriction of a quadratic penalty $\gamma_{kj}^T \mathbf{G}_{kj}(\lambda_{kj})\gamma_{kj}$, with penalty matrix $\mathbf{G}$ and hyperparameters λ that regulate the amount of smoothing [80,81]. The optimisation is carried out using a two-cycle backfitting algorithm. The user of the GAMLSS software can choose between two specific algorithms: The RS and CG algorithm [23]. While the RS algorithm is more stable and faster in general, the CG algorithm may outperform the RS algorithm in situations where the distribution parameters are highly correlated [81]. Using only linear predictors and the gamma distribution, the computational stability of the RS algorithm tends to outweigh the advantages of the CG algorithm, so we recommend using the former as a default option to inexperienced users. Detailed information on the estimation routines are provided by Stasinopoulos and Rigby [23], Rigby and Stasinopoulos [80], and Stasinopoulos et al. [81].

3. Results

Following the notions from the previous section, we will use the two-parameter gamma distribution to model the transformed MCS in a distributional regression framework. In this section, we solely display and discuss extended results (regression estimates, distributional measures, exemplary conditional densities) for this distribution, but also provide other distributions (i.e., the BCPE and GB2 distribution) in the Appendix A.4.

3.1. The Predictor

The following regression predictor set-up is applied to all parameters of the distribution, with θ denoting a generic parameter representing either of the gamma distribution's parameters, μ or σ:

$$
\begin{aligned}
\eta_\theta = \beta_0^\theta &+ \beta_1^\theta AGE + \beta_2^\theta AGESQ + \beta_3^\theta LOGINC + \beta_4^\theta GER \\
&+ \beta_5^\theta EDU_2 + \beta_6^\theta EDU_3 + \beta_7^\theta EDU_4 + \beta_8^\theta MAR_2 \\
&+ \beta_9^\theta MAR_3 + \beta_{10}^\theta MAR_4 + \beta_{11}^\theta EAST + \beta_{12}^\theta CITY \\
&+ \beta_{13}^\theta UNEMPLOYED,
\end{aligned}
\tag{4}
$$

where *AGE* denotes the age of the individual in years, and *AGESQ* denotes the squared age of the individual. *LOGINC* refers to the logarithm of the annual net equivalised household income. *GER* refers to a dummy variable indicating whether an individual is a German national or not. The variables EDU_1–EDU_4 represent the respondents' educational attainment measured with the International Standard Classification of Education (ISCED97) [82]. EDU_1 includes all individuals with ISCED levels 0, 1 and 2, representing pre-primary, primary, and lower secondary education; EDU_2 includes all observations with ISCED level 3 representing upper secondary education; EDU_3 includes all observations with ISCED levels 4 and 5; EDU_4 includes all observations with ISCED level 6. In Germany, this means that the first education level (EDU_1) contains the educational attainment of finishing kindergarten and/or Haupt-, Realschule, or the Gymnasium (ohne Oberstufe). The second education level (EDU_2) contains the educational attainment of finishing Berufsschule, Gymnasium (Oberstufe), or equivalent schooling levels. The third education level (EDU_3) contains the educational attainment of being awarded a Bachelor's or Master's degree. The fourth education level (EDU_4) contains the educational attainment of being awarded a doctorate or a habilitation. The variables MAR_1–MAR_4 represent the respondents' marital status, with MAR_1 containing individuals who are married and living together or who are in a same-sex partnership and living with their partner; MAR_2 contains all individuals who are married but living separately, divorced, having a dissolved registered partnership, or those who have a registered same-sex partnership but are living separately; MAR_3 contains all individuals who are single; and MAR_4 contains all individuals who are widowed. *EAST* denotes a dummy variable indicating whether an individual lives in former West or East Germany. *CITY* denotes a dummy variable indicating whether an individual lives in an urban area or not. *UNEMPLOYED* denotes a dummy variable indicating whether the respondent is unemployed or not.

Note that the predictor does not contain coefficients for the variables EDU_1 and MAR_1 since they function as the reference group, meaning their effect is subsumed in the intercept β_0^θ. Further information on how the above mentioned variables were built, as well as their theoretical background and SOEP references, are provided in Appendix A.1.

3.2. The Regression Coefficients

The results of the regression are displayed in Table 1, with the standard errors denoted in parentheses following the coefficients' estimate. The standard errors were computed on the basis of the variance–covariance matrix provided by the GAMLSS software [81]. It should be noted that these values are to be interpreted with caution. For this reason, we provide the more reliable bootstrap-based intervals for the analysis of the mental health and income association.

Table 1. Linear effects of η^μ and η^σ for MCS in the Gamma model, with standard errors in parentheses.

	Male				**Female**			
	η^μ		η^σ		η^μ		η^σ	
const.	1.771 ***	(0.045)	−1.003 ***	(0.172)	1.944 ***	(0.043)	−0.994 ***	(0.158)
MAR_2	0.018 *	(0.008)	0.115 ***	(0.028)	0.031 ***	(0.006)	0.075 ***	(0.022)
MAR_3	0.007	(0.006)	0.052 *	(0.025)	0.021 ***	(0.006)	0.042 .	(0.022)
MAR_4	0.043 **	(0.014)	0.036	(0.048)	0.025 **	(0.009)	0.028	(0.03)
GER	−0.019 **	(0.007)	−0.028	(0.027)	−0.032 ***	(0.006)	−0.035	(0.025)
UNEMPLOYED	−0.026 **	(0.010)	−0.048	(0.034)	−0.027 **	(0.009)	−0.094 **	(0.029)
$EDUC_2$	−0.007	(0.007)	−0.046 .	(0.027)	−0.020 ***	(0.006)	−0.038 .	(0.021)
$EDUC_3$	−0.004	(0.009)	−0.052	(0.032)	−0.002	(0.008)	−0.006	(0.028)
$EDUC_4$	−0.008	(0.008)	−0.101 ***	(0.031)	−0.016 *	(0.007)	−0.056 *	(0.026)
CITY	0.014 ***	(0.004)	0.012	(0.016)	0.005	(0.004)	−0.013	(0.015)
EAST	0.008 .	(0.005)	−0.036 .	(0.019)	0.003	(0.005)	0.001	(0.017)
AGE	0.003 ***	(0.001)	0.005	(0.003)	0.001	(0.001)	0.001	(0.003)
AGESQ	0.000 ***	(0.000)	0.000	(0.000)	0.000 **	(0.000)	0.000	(0.000)
LOGINC	−0.021 ***	(0.004)	−0.083 ***	(0.017)	−0.029 ***	(0.004)	−0.070 ***	(0.015)

Notes. ***, **, * and . refer to significance levels with $p \leq 0.001$, $p \leq 0.01$, $p \leq 0.05$ and $p \leq 0.1$ obtained from t-tests with $H_0 : \beta_j^\theta = 0$ and $H_1 : \beta_j^\theta \neq 0$.

Due to the intricate nature of the parameter interpretation, we have refrained from discussing the results at length. The interpretation is intricate, since the parameters of the response distribution cannot necessarily be equated with a distributional measure, like the expectation, variance, or skewness (which would be more or less straightforward to interpret). Distributional measures are often functions of several parameters of the response distribution. We used the gamma distribution in mean parameterisation with density:

$$f_Y(y|\mu,\sigma) = \frac{y^{1/\sigma^2-1}e^{-y/(\sigma^2\mu)}}{(\sigma^2\mu)^{1/\sigma^2}\Gamma(1/\sigma^2)}, \tag{5}$$

for $y > 0$, $\mu > 0$, $\sigma > 0$, with $E(Y) = \mu$ and $Var(Y) = \mu^2\sigma^2$.

It should, however, be pointed out that living in separation of a partner (MAR_2) has a highly significant coefficient for σ, indicating variations beyond the expected value of the distribution. Similar observations—to various degrees of significance—can be made for being unemployed, as well as for having received a higher education ($EDUC_4$). Last but not least, the household income ($LOGINC$) features highly significant effects on parameters for both men and women, and will be considered in more detail in the following.

3.3. A Distributional Perspective on the Association between Mental Health and Income

In order to assess the relationship between mental health and income, while accounting for the other variables, we followed Silbersdorff et al. [27] and employed the concept of "average Joe" and "average Jane" representing an ideal-typical man and woman, respectively, with average characteristics. In our case, the following variable combination was assumed: 52 years old, married, living in West Germany, having standard secondary education, having German nationality, being employed, and not living close to a city centre. In the first step, we consider the relationship between three distribution measures and income at large, and subsequently consider the resultant differences between 15,000€ and 30,000€ in more detail.

3.3.1. Visualizing the Mental Health and Income Relationship

Figure 3 shows the relation between income and the different distribution measures for "average Joe" and "average Jane".

The x-axis begins at 4700€, which roughly matches the income received on the basis of the German Social Security. The regular level of social security for a single person was set at 391€ per month in 2014. Approximately one percent of the survey participants had an income below that threshold, and reaches up to 100,000€ to exclude the economic elite, which is not adequately represented in the SOEP [83]. Furthermore, the x-axis is plotted on the log scale. The dashed lines show 0.95%-pointwise confidence intervals of the portrayed estimates. They were obtained by bootstrap sampling with 2500 bootstrap samples. Details on the bootstrap procedure and recommendations are provided in Appendix A.5.

The graph at the left portrays the expected outcome with varied incomes. The y-axis displays back-transformed MCS scores. It can be observed that the MCS scores increase with increasing income for males and females, as one would expect.

The right column of Figure 3 shows the effect of income on the conditional probability that a person will fall below specific threshold values T of the MCS scale. Specifically, we used the threshold of having a MCS score below the lowest quintile (denoted by $R_{0.2}$) and the lowest vingtile (denoted by $R_{0.05}$). Note that we used separate thresholds for men and women, with $T_{0.2}^{\male} \approx 45$, $T_{0.2}^{\female} \approx 42$, $T_{0.05}^{\male} \approx 34$ and $T_{0.05}^{\female} \approx 30$. Ware et al. [8] found that among individuals with MCS scores between 30–34, 89% reported symptoms related to depression, while this was 59% for individuals with MCS scores between 40–44. These risk measures are essential in our analysis that should take a perspective that goes beyond the mean. Modeling the whole conditional response distribution gives the possibility of deriving findings extracted from the resulting probability density functions. The idea of the chosen risk measures is to portray the risk of minor or major mental health issues (with $R_{0.2}$ entailing both, and $R_{0.05}$ only the latter).

Both risk measures show that the risk of falling below the respective thresholds decreases with increasing income. In particular, one can observe that for $R_{0.05}$, the individuals with very high incomes can practically eliminate their risk of being in a very bad mental health condition, which mirrors the finding from Silbersdorff et al. [27] on the PCS score. The association regarding the distributional measures for the BCPE and GB2 model is very similar, and has been displayed in Appendix A.4.
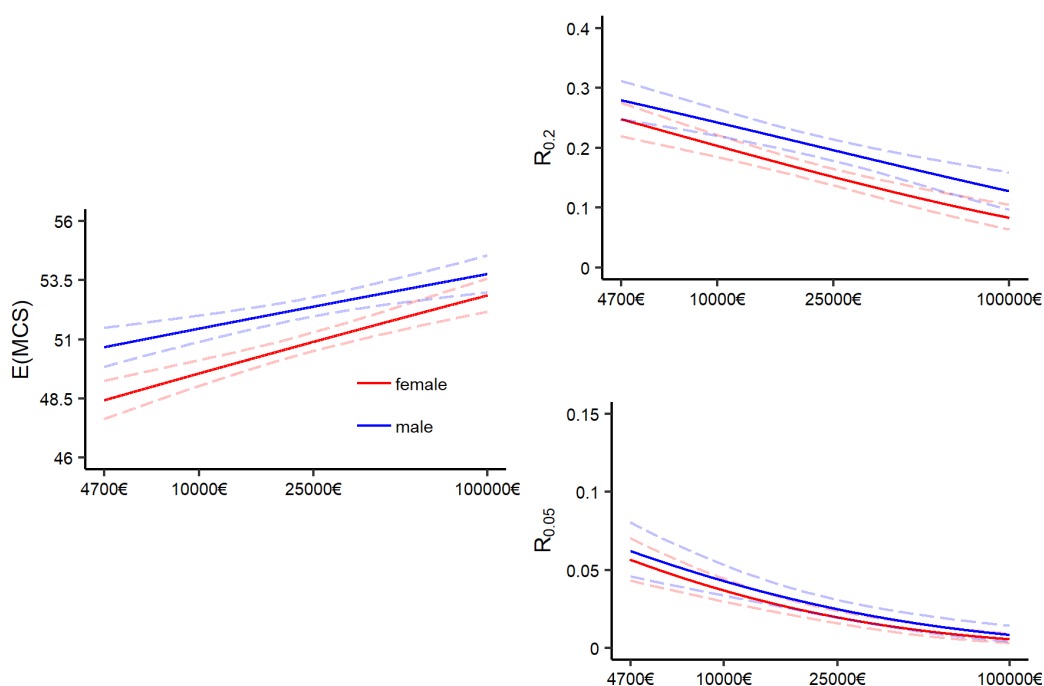
Int. J. Environ. Res. Public Health **2019**, *16*, 4009

12 of 28

Figure 3. Effect of income on distribution measures for the GA model: Ga ($\mu = \checkmark$, $\sigma = \checkmark$). **Left**: Effect of income on the expectation. **Right**: Effect of income on risk of falling below the lowest quintile and lowest vingtile.

3.3.2. Contrasting Mental Health for Two Income Levels

Let us now consider the difference between two income levels—15,000€, representing the 25th percentile or the median of the poorer half of the population, and 30,000€, representing the 75th percentile or the median of the better-off half of the population.

Figure 4 displays the estimated conditional probability density functions for average Joe (blue) and Jane (red) at an income of 15,000€ (solid) and 30,000€ (dashed). Additionally, Table 2 shows the corresponding distributional measures.

Table 2. Expectation and risk measures for average Joe and Jane at income levels of 15,000€ and 30,000€.

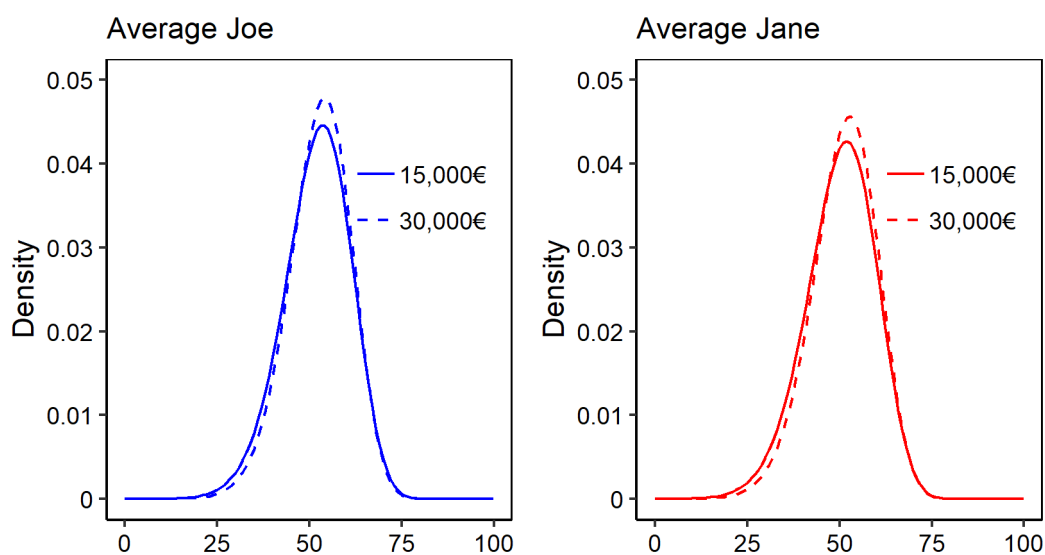|		15,000€		30,000€		Relative Difference	
male	μ	51.86	[51.4 ; 52.31]	52.56	[52.14; 52.98]	0.013	[0.008 ; 0.02]
	$R_{0.2}$	0.222	[0.203 ; 0.241]	0.187	[0.168 ; 0.205]	0.158	[0.105 ; 0.216]
	$R_{0.05}$	0.034	[0.027 ; 0.042]	0.022	[0.017 ; 0.028]	0.358	[0.256 ; 0.456]
female	μ	50.15	[49.72 ; 50.6]	51.16	[50.74 ; 51.57]	0.020	[0.014 ; 0.026]
	$R_{0.2}$	0.180	[0.165 ; 0.195]	0.141	[0.127 ; 0.156]	0.214	[0.161 ; 0.265]
	$R_{0.05}$	0.028	[0.023 ; 0.034]	0.017	[0.013 ; 0.021]	0.401	[0.311 ; 0.483]

Figure 4. Estimated conditional density of average Joe (blue) and Jane (red) at income levels of 15,000€ and 30,000€.

The portrayed densities reveal that differences are not equal across the distribution, but feature a relocation of probability mass from the lower end of the distribution to the centre of the distribution (when changing from 15,000€ to 30,000€). In other words, the change in income mainly incurs a depletion of the risk of having low or very low mental health scores, and changes only very little for the upper spectrum of the distribution.

Depending on the distribution measure, we thus found relative differences of varying magnitudes. The relative difference is the absolute difference of the measures for 15,000€ and 30,000€, divided by the measure for 15,000€. For males, the expected mental health outcome increased from 51.86 to 52.56, a relative difference of only 1.3%. For females, the corresponding change was 50.15 to 51.16, yielding a relative difference of 2.0%.

By contrast, considering the risk measure $R_{0.2}$ which focuses precisely on the lower end of the distribution, we saw a change from 0.222 to 0.187 for males, meaning that the average Joe with an income of 15,000€ with a risk of suffering from a minor or major mental health issue was expected to be 15.8% higher than that of a comparable man with an income of 30,000€.

This difference becomes even more pronounced when focusing on the more extreme $R_{0.05}$ measure that is thought to focus on the major mental health issues alone, and features a 35.8% and 40.1% change for men and women, respectively.

These results show that, like for PCS, the association between income and health is much more pronounced at the lower end of the health spectrum than it is for expected health.

4. Conclusions

In this publication we have provided some general insights into the use of distributional regression approaches on the issue of socioeconomic health inequality in general, and provided an application to the relationship between mental health and income in detail.

This publication complements the findings related to physical health provided by Silbersdorff et al. [27]. Taken together, both publications show that taking a distributional perspective can reveal insights that otherwise would not have been revealed. While Silbersdorff et al. [27] utilised a distributional regression approach within the bayesian framework, in this publication the frequentistic framework was followed. Independent of the framework, the aim was to estimate complete conditional (mental) health distributions rather than obtaining single estimates for the expected outcome. These

provide more detailed information on the relationship between income and (mental) health, which we made explicit by defining risk measures. Regarding this relationship between mental health and income, the risk measures showed that income is more strongly related to the risk of being in a poor mental health condition than to the expected mental health. Thus, this publication contributes to the literature stated in Section 1.2, since it disentangles the effect of income on mental health and gives a first indication that the expectation-based perspective may underestimate the importance of income, not only for physical health, but also in mental health.

Regarding methodological guidelines, we proposed to use simple distributional regression models with linear predictors and distributions with few parameters, like the gamma distribution with two parameters. Additionally, we proposed to use simple and intuitive measures to exploit the information generated by estimating full conditional response distributions. Subsequently, we focused on risk measures on the basis of this distribution, but other measures could also be considered. Looking also at more complex models, we find that while these often provide better model fits, they also feature problematic fits in the sparsely populated areas of the covariate space, as well as much wider confidence intervals due to the higher model instability. Following Box [84] it should be noted that frequentist models, as well as bayesian models suffer from under-accounting uncertainty due to the unconditional assumption of the model specification. The wider confidence intervals of the more complex models thus go some direction in correcting for this under-accounting, but do so only implicitly.).

Naturally, the proposed methods can be extended even further. For example, one could make use of the longitudinal nature of the data, yielding a deeper data foundation to the analysis [14]. The GAMLSS framework theoretically provides the required statistical repertoire, as random effects can be incorporated [81] and longitudinal approaches have been applied in other fields [85,86]. One could overcome the rather questionable separation between physical and mental health [87] and model them jointly by bivariate distributional regression, which is currently being developed [88]. While these extensions may provide interesting research avenues for the methodologically minded health inequalities researcher, we believe that for many applied researchers, a simple distributional regression approach will suffice to gain sound and interesting insights into the matter.

Even if using a relatively simple predictor and distribution with few parameters, it must be noted that the modelling of distributional regression is still a complex undertaking, yielding mostly indicative and approximate, rather than robust and precise results. Further evidence from theory and other modelling approaches is thus usually needed for any causal conclusions to be made.

Nonetheless, following Chalmers [89], new methods generating new empirical evidence on persistent questions are sometimes needed to generate progress in science. The distributional regression approach allows for looking at the full health distribution while conditioning on a set of variables, and therefore provides additional perspectives that should be considered for any comprehensive assessment on the much-discussed relationship between income and health.

Author Contributions: Conceptualization: A.S.; methodology: A.S. & K.S.S.; software: K.S.S.; validation: A.S.; formal analysis: A.S. & K.S.S.; investigation: K.S.S.; data curation, K.S.S.; writing—original draft preparation, A.S.; writing—review and editing: A.S. & K.S.S.; visualization: A.S. & K.S.S.; supervision: A.S.; project administration: A.S. & K.S.S.

Abbreviations

The following abbreviations are used in this manuscript:

GLM	Generalised Linear Model
GAMLSS	Generalised Linear Models of Location, Scale & Shape
MCS	Mental Component Scale
PCS	Physical Component Scale

Appendix A

Appendix A.1. Variables

In the following it is explained which variables provided by the SOEP were used. Each variable can be linked via its unique identifier to the data file containing it. The SOEP provides an online search-engine which allows for further information on the variables to be easily obtained by searching for the unique identifier: https://paneldata.org/soep-core (last access: 5 September 2019).

For the dependent variable, the Mental Component Scale (MCS) scores, contained in variable `MCS`, from the `HEALTH` file were used. The MSC scale of the SF-12v2 provide self-rated measurements of specific aspects of mental health. Using instruments based on self-reports are suited to measure aspects of health, since mental and physical health are facets of the health-related quality of life, which is a subjective matter [90]. Self-rated measurements are a standard tool in health research, and experience growing importance [37]. They may reveal insights that might not be gained with objective measurements—for example, according to Bährer-Kohler and Carod-Artal [91], many people state that they suffer from poor mental health even though they may not reach thresholds for official diagnoses. An observation's score is computed as a weighted average of the answers given on six questions. The weights are constructed with the help of factor loadings resulting from a factor analysis. The MCS scores provided by the SOEP are based on a four-factor model. Information on the algorithm for their computation is provided by Nübling et al. [64] and Andersen et al. [63]. The four factors provided by the SOEP are labelled as mental health (two items), role emotional (two items), social functioning (one item), and vitality (one item) [64]. Test-theoretical properties of the MCS scale are mentioned in Section 2 (p. 3).

In some parts, the same explanatory variables as used by Silbersdorff et al. [27] were incorporated. The authors used variables that are standard in the field of (physical) health research, such as annual net equivalised household income, age, education, marital status, a variable indicating whether an individual is a German national (or not), and a variable indicating whether an individuals lives in former West or East Germany. These variables are also frequently used in mental health research [14,34,37,92] and are in line with current academic textbooks ([93], pp. 76–77).

An individual's income is operationalised as the annual net equivalised household income. The annual net household income is given by `i1110215`. The variable is provided in the `bfpequiv` file, from the survey of year 2015. It is a backdating question, and therefore refers to the income of 2014. The annual net household income is adjusted with the help of Modified OECD Equivalence Weights to account for the number of individuals living in the household, as well as the children/adults ratio [94]. The formula for the computation recommended by the SOEP is used ([95], p. 38). In order to equivalise, the variables "Number of persons in household" (`d1110614`) and "Number of children in household aged 0–13" (`h1110114`) are used. Both variables are taken from the `bepequiv` file. Following Silbersdorff et al. [27] a log transformation (natural logarithm) on income is used. For simplification, within this publication, the term *LOGINC* is used to refer to the variable.

While income is the main variable of interest, all other variables included in the models are used to statistically control for their effects:

The variable age (*AGE*) is provided by `d1110114` in the `bepequiv` file. Based on simple descriptive analysis, (mean) PCS scores seem to decrease with increasing age, whereas (mean) MCS scores seem to stay approximately equal ([8], pp. 815–816, Table 8.4). Westerhof and Keyes [50] disentangled the effect of age on mental health (positive aspects) and mental illness with the help of four linear regression models with differing dependent variables. The authors found negative significant effects for age and age squared on mental illness, yet the effect of age squared vanished when control variables (e.g., age, educational status, employment status, gender, income, migration) were introduced. For age and age squared, positive significant effects were found on emotional well-being. Additionally, age—but not age squared—was found to have positive significant effect on psychological well-being. On the other hand, no effects were found for social well-being.

Following Silbersdorff et al. [27], the individuals' educational attainment was measured with the International Standard Classification of Education—ISCED97 [82]. The variable can be found in the `bepgen` file with the identifier `isced97_14`. The original six categories were reduced to four categories: EDU_1 includes all individuals with ISCED levels 0, 1, and 2, representing pre-primary, primary, and lower secondary education; EDU_2 includes all observations, with ISCED level 3 representing upper secondary education; EDU_3 includes all observations with ISCED levels 4 and 5; EDU_4 includes all observations with ISCED level 6. Related to Germany this means, that the first education level (EDU_1) contains the educational attainment of finishing Kindergarten and/or Haupt-, Realschule, Gymnasium (ohne Oberstufe). The second education level (EDU_2) contains the educational attainment of finishing Berufsschule, Gymnasium (Oberstufe) or equivalent schooling-levels. The third education level (EDU_3) contains the educational attainment of being awarded a Bachelor's or Master's degree. The fourth education level (EDU_4) contains the educational attainment of being awarded a doctorate or a habilitation.

As a dummy variable indicating whether an individual is a German national, the variable `bep129` from the `bep` file was used.

For the marital status, the six categories of `bep127` from the `bep` file were reduced to four categories. The four categories are: MAR_1, containing individuals who are married and are living together or who are in a same-sex partnership and living with their partner; MAR_2 contains all individuals being married but living separately, who are divorced or having a dissolved registered partnership, or those who have a registered same-sex partnership but are living separately; MAR_3 contains all individuals who are single; MAR_4 contains all individuals who are widowed.

As a variable indicating whether an individual lives in former West or East Germany, a dummy variable is built with the help of the variable state of residence (`l1110114`) from the `bepequiv` file. Differences in the prevalence of mental disorders between former West and East Germany were present in the past [96], but seem to have vanished nowadays [34].

To account for the well-known effects of being unemployed on mental health, a dummy variable with two levels is used: $UNEMPLOYED$. Therefore, the variable `bep09` from the `bep` file is used. Meta analyses on the evidence of negative effects of being unemployed on mental illness [58], as well as evidence of positive effects of being employed on mental health exist [61].

Differences in mental health when comparing individuals living in rural and and urban areas are well-known. Mental health tend to be worse for individuals living in urban areas compared to individuals living in rural areas [34,57,59,62]. Many different operationalisations of the degree of rurality/urbanity in research exist—for example, self ratings [59], number of inhabitants [62], and the degree of concentration of inhabitants [60]. In part, these effects may be grounded in special physical characteristics of urban areas such as traffic, pollution, noise, artificial light at night, availability of green areas, presence of crime, availability of drugs, and many more. In order to operationalise this concept, the variable `beh57` (from the `beh` file) is dichotomised to the variable $CITY$, indicating whether an individual lives closer than ten kilometres to the nearest metropolis (reference group) or not.

Only those individuals with full information on all variables are used. Individuals who reported that their income is zero were removed from the analysis since the log-transformation could not be applied (0.0008% of the sample). The sample for the analysis yields 22,678 individuals: 10,387 males and 12,291 females. 80% of the data was used as training data (and therefore 20% was used as test data).

Appendix A.2. Choosing a Distribution for the Conditional Health Distribution

Table A1 lists all used distributions tested as response distributions within this publication. Columns μ, σ, ν, τ indicate the used link functions, if the parameter is present in the respective distribution. Rigby et al. [97] provide a publication in which all distributions available in the `gamlss` package are described.

Table A1. Distributions used for the analysis.

Distribution	gamlss **Name**	μ	σ	ν	τ	**Page**
Box–Cox power exponential (BCPE)	BCPE	ident.	log	ident.	log	p. 291
Box-Cox-Cole-Green (BCCG)	BCCG	ident.	log	ident.	-	p. 282
Box–Cox-t (BCT)	BCT	ident.	log	ident.	log	p. 290
Dagum (Da)	GB2	log	log	log	= 1	p. 294
Gamma (Ga)	GA	log	log	-	-	p. 271
Generalised Beta type 2 (GB2)	GB2	log	log	log	log	p. 293
Generalised Gamma (GG)	GG	log	log	ident.	-	p. 285
Generalised Inverse Gaussian (GIG)	GIG	log	log	ident.	-	p. 287
Log Normal (LOGNo)	LOGNO	ident.	log	-	-	p. 275
Normal (No)	NO	ident.	log	-	-	p. 232
Pearson-Type-VI (PtVI)	GB2	log	= 1	log	log	p. 293
Singh-Maddala (SM)	GB2	log	log	= 1	log	p. 293
Weibull (WEI3)	WEI3	log	log	-	-	p. 280

The column named *page* of Table A1 gives the page of the publication on which the distribution is described. The probability density function, cumulative density function and, if defined, the expectation, variance, skewness, and kurtosis are provided. The publication can be accessed at: https://www.gamlss.com/distributions/; last access: 26 August 2019.

Tables A2 and A3 display information criteria of the models for the male and female sample. Both tables are sorted by the AIC in ascending order. The smallest five values in each column are marked by a dagger symbol (†). The notation can be understood as follows: BCT ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = -$) indicates that a model with BCT distribution as a response distribution is fitted. The BCT distribution is fully described by four parameters, where the check symbols ($\checkmark$) indicate that only the first three have been linked to the regression predictor, while τ is modelled only by an intercept.

Unfortunately, estimation routines fail when a GG distribution is used and all three parameters are linked to the regression predictor. As a consequence, GG ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark$) does not occur in the respective tables.

Table A2. Information criteria of estimated models with the male sample.

Model (Male)	AIC	BIC	GAIC ($\kappa = 4$)	TGDEV [a]
GB2 ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = \checkmark$)	20,734 [†]	21,128	20,846 [†]	5166 [†]
GB2 ($\mu = \checkmark, \sigma = \checkmark, \nu = -, \tau = -$)	20,760 [†]	20,971 [†]	20,820 [†]	5185 [†]
GB2 ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = -$)	20,763 [†]	21,066	20,849 [†]	5188
Da ($\mu = \checkmark, \sigma = \checkmark, \nu = -$)	20,769 [†]	20,973 [†]	20,827 [†]	5189
BCT ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = -$)	20,777 [†]	21,079	20,863	5184 [†]
Da ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark$)	20,781	21,076	20,865	5189
BCT ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = \checkmark$)	20,781	21,175	20,893	5187
BCCG ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark$)	20,783	21,078	20,867	5181 [†]
BCPE ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = -$)	20,784	21,086	20,870	5182 [†]
BCPE ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = \checkmark$)	20,786	21,179	20,898	5187
BCT ($\mu = \checkmark, \sigma = \checkmark, \nu = -, \tau = -$)	20,799	21,010 [†]	20,859 [†]	5197
GG ($\mu = \checkmark, \sigma = \checkmark, \nu = -$)	20,809	21,012 [†]	20,867	5190
BCCG ($\mu = \checkmark, \sigma = \checkmark, \nu = -$)	20,810	21,013 [†]	20,868	5192
BCPE ($\mu = \checkmark, \sigma = \checkmark, \nu = -, \tau = -$)	20,810	21,021	20,870	5195
SM ($\mu = \checkmark, \sigma = \checkmark, \tau = -$)	20,848	21,052	20,906	5221
SM ($\mu = \checkmark, \sigma = \checkmark, \tau = \checkmark$)	20,851	21,146	20,935	5220
GIG ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark$)	20,879	21,174	20,963	5233
PtVI ($\mu = \checkmark, \nu = \checkmark, \tau = \checkmark$)	20,903	21,198	20,987	5240
GB2 ($\mu = \checkmark, \sigma = -, \nu = -, \tau = -$)	20,915	21,034	20,949	5242
Da ($\mu = \checkmark, \sigma = -, \nu = -$)	20,918	21,031	20,950	5244
GIG ($\mu = \checkmark, \sigma = \checkmark, \nu = -$)	20,919	21,123	20,977	5248
PtVI ($\mu = \checkmark, \nu = \checkmark, \tau = -$)	20,929	21,133	20,987	5248
BCT ($\mu = \checkmark, \sigma = -, \nu = -, \tau = -$)	20,938	21,058	20,972	5249
BCPE ($\mu = \checkmark, \sigma = -, \nu = -, \tau = -$)	20,952	21,072	20,986	5250
BCCG ($\mu = \checkmark, \sigma = -, \nu = -$)	20,956	21,069	20,988	5246
GG ($\mu = \checkmark, \sigma = -, \nu = -$)	20,958	21,070	20,990	5246
SM ($\mu = \checkmark, \sigma = -, \tau = -$)	20,979	21,092	21,011	5267
GIG ($\mu = \checkmark, \sigma = -, \nu = -$)	21,006	21,119	21,038	5277
LOGNo ($\mu = \checkmark, \sigma = \checkmark$)	21,020	21,216	21,076	5281
PtVI ($\mu = \checkmark, \nu = -, \tau = -$)	21,028	21,141	21,060	5286
LOGNo ($\mu = \checkmark, \sigma = -$)	21,146	21,251	21,176	5328
Ga ($\mu = \checkmark, \sigma = \checkmark$)	21,267	21,464	21,323	5357
Ga ($\mu = \checkmark, \sigma = -$)	21,381	21,486	21,411	5400
No ($\mu = \checkmark, \sigma = \checkmark$)	22,040	22,237	22,096	5574
No ($\mu = \checkmark, \sigma = -$)	22,148	22,253	22,178	5613
WEI3 ($\mu = \checkmark, \sigma = \checkmark$)	23,305	23,501	23,361	5894
WEI3 ($\mu = \checkmark, \sigma = -$)	23,341	23,447	23,371	5909

Notes. [a] Test-Global-Deviance. [†] smallest five values in column. "$\theta = \checkmark$" indicates that the parameter is linked to the predictor. "$\theta = -$" indicates that the parameter is modelled solely by an intercept.

Int. J. Environ. Res. Public Health **2019**, *16*, 4009

19 of 28

Table A3. Information criteria of estimated models with the female sample.

Model (Female)	AIC	BIC	GAIC ($\kappa = 4$)	TGDEV [a]
BCPE ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = \checkmark$)	26,475 [†]	26,878	26,587 [†]	6617
BCPE ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = -$)	26,487 [†]	26,796 [†]	26,573 [†]	6603 [†]
GB2 ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = \checkmark$)	26,547 [†]	26,950	26,659	6622
BCCG ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark$)	26,555 [†]	26,857	26,639 [†]	6612 [†]
BCT ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = -$)	26,557 [†]	26,866	26,643 [†]	6612 [†]
BCPE ($\mu = \checkmark, \sigma = \checkmark, \nu = -, \tau = -$)	26,569	26,785 [†]	26,629 [†]	6623
BCT ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = \checkmark$)	26,576	26,979	26,688	6612 [†]
GB2 ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark, \tau = -$)	26,595	26,904	26,681	6613 [†]
GB2 ($\mu = \checkmark, \sigma = \checkmark, \nu = -, \tau = -$)	26,605	26,821 [†]	26,665	6624
GG ($\mu = \checkmark, \sigma = \checkmark, \nu = -$)	26,615	26,824 [†]	26,673	6627
BCCG ($\mu = \checkmark, \sigma = \checkmark, \nu = -$)	26,621	26,829 [†]	26,679	6629
BCT ($\mu = \checkmark, \sigma = \checkmark, \nu = -, \tau = -$)	26,623	26,839	26,683	6629
GIG ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark$)	26,626	26,928	26,710	6634
PtVI ($\mu = \checkmark, \nu = \checkmark, \tau = \checkmark$)	26,658	26,960	26,742	6639
PtVI ($\mu = \checkmark, \nu = \checkmark, \tau = -$)	26,659	26,867	26,717	6643
GIG ($\mu = \checkmark, \sigma = \checkmark, \nu = -$)	26,668	26,877	26,726	6634
LOGNo ($\mu = \checkmark, \sigma = \checkmark$)	26,723	26,924	26,779	6658
Da ($\mu = \checkmark, \sigma = \checkmark, \nu = \checkmark$)	26,753	27,056	26,837	6643
Da ($\mu = \checkmark, \sigma = \checkmark, \nu = -$)	26,764	26,973	26,822	6652
BCPE ($\mu = \checkmark, \sigma = -, \nu = -, \tau = -$)	26,800	26,922	26,834	6650
GG ($\mu = \checkmark, \sigma = -, \nu = -$)	26,824	26,939	26,856	6651
BCCG ($\mu = \checkmark, \sigma = -, \nu = -$)	26,827	26,942	26,859	6653
GIG ($\mu = \checkmark, \sigma = -, \nu = -$)	26,828	26,943	26,860	6653
BCT ($\mu = \checkmark, \sigma = -, \nu = -, \tau = -$)	26,829	26,952	26,863	6653
GB2 ($\mu = \checkmark, \sigma = -, \nu = -, \tau = -$)	26,837	26,960	26,871	6651
PtVI ($\mu = \checkmark, \nu = -, \tau = -$)	26,847	26,962	26,879	6659
LOGNo ($\mu = \checkmark, \sigma = -$)	26,902	27,009	26,932	6676
SM ($\mu = \checkmark, \sigma = \checkmark, \tau = \checkmark$)	26,903	27,205	26,987	6684
Ga ($\mu = \checkmark, \sigma = \checkmark$)	26,921	27,123	26,977	6712
SM ($\mu = \checkmark, \sigma = \checkmark, \tau = -$)	26,941	27,150	26,999	6694
Da ($\mu = \checkmark, \sigma = -, \nu = -$)	26,980	27,095	27,012	6674
Ga ($\mu = \checkmark, \sigma = -$)	27,077	27,185	27,107	6727
SM ($\mu = \checkmark, \sigma = -, \tau = -$)	27,122	27,237	27,154	6711
No ($\mu = \checkmark, \sigma = \checkmark$)	27,637	27,838	27,693	6902
No ($\mu = \checkmark, \sigma = -$)	27,799	27,907	27,829	6920
WEI3 ($\mu = \checkmark, \sigma = \checkmark$)	28,755	28,957	28,811	7190
WEI3 ($\mu = \checkmark, \sigma = -$)	28,798	28,906	28,828	7198

Notes. [a] Test-Global-Deviance. [†] smallest five values in column. "$\theta = \checkmark$" indicates that the parameter is linked to the predictor. "$\theta = -$" indicates that the parameter is modelled solely by an intercept.

Appendix A.3. Residual Diagnostics

Stasinopoulos et al. [81] (pp. 417–422) recommend to use diagnostic plots of normalised quantile residuals to assess the adequacy of estimated models. According to Stasinopoulos et al. [81] (p. 418), the "[...] main advantage of normalized (randomized) quantile residuals is that, whatever the distribution of the response variable, the true residuals always have standard normal distribution when the assumed model is correct". Hereafter, we refer to normalised quantile residuals simply as residuals.

The top row of Figure A1 shows kernel density estimates of the residuals from the models estimated with the male training data (red), as well as a standard normal distribution (blue). The bottom row shows corresponding quantile–quantile plots. The residuals of the GA model deviate most from a standard normal distribution. The residuals are positively skewed ($skew(\hat{\epsilon})_{male}^{GA} = 0.594$) and the kernel density estimate reveals a tail too thin for small residuals and a tail too fat for large residuals. Additionally, the distribution of residuals is leptokurtic $kurt(\hat{\epsilon})_{male}^{GA} = 3.43$ (the curtosis

of a standard normal distribution is 3). This is also resembled by deviations from the red diagonal shown in the quantile–quantile plots. Using BCPE and GB2 distributions led to residuals with a skewness close to 0 ($skew(\hat{\epsilon})_{male}^{BCPE} = -0.003$, $skew(\hat{\epsilon})_{male}^{GB2} = -0.019$). Furthermore, the distribution of residuals is less leptokurtic for the BCPE model ($kurt(\hat{\epsilon})_{male}^{BCPE} = 3.12$) and slightly platykurtic for the GB2 model ($kurt(\hat{\epsilon})_{male}^{GB2} = 2.907$). Even though the tails of the kernel density estimates for the BCPE and GB2 model look close to optimal, the quantile–quantile plots show that there is a remaining deviation from a standard normal distribution in the tails. Figure A2 shows the same type of plot for models estimated with the female training data. The general observable patterns are comparable to the models estimated with the male training data. The residuals resulting from the GA model are positively skewed ($skew(\hat{\epsilon})_{female}^{GA} = 0.413$) and slightly platykurtic ($kurt(\hat{\epsilon})_{female}^{GA} = 2.888$). The quantile–quantile plots reveal that there are deviations from the optimal in the lower tail. Using a BCPE or GB2 distribution instead again leads to distributions of residuals with a skewness close to 0 ($skew(\hat{\epsilon})_{female}^{BCPE} = -0.002$, $skew(\hat{\epsilon})_{female}^{GB2} = -0.022$). While the curtosis of the residuals of the BCPE model is very close to its ideal value of three ($kurt(\hat{\epsilon})_{female}^{BCPE} = 3.062$), the residuals of the GB2 model are platykurtic ($kurt(\hat{\epsilon})_{female}^{GB2} = 2.621$). Again, the tails of the kernel density estimates for the BCPE and GB2 model look close to optimal. Yet, the quantile–quantile plots show that there is a remaining deviation from a standard normal distribution in the tails. As a conclusion, it can be said that none of the used response distributions leads to residuals that clearly outperform the other. For all models, it holds that deviations of the distribution of residuals from a standard normal distribution are not of such magnitude that the models must be discarded.
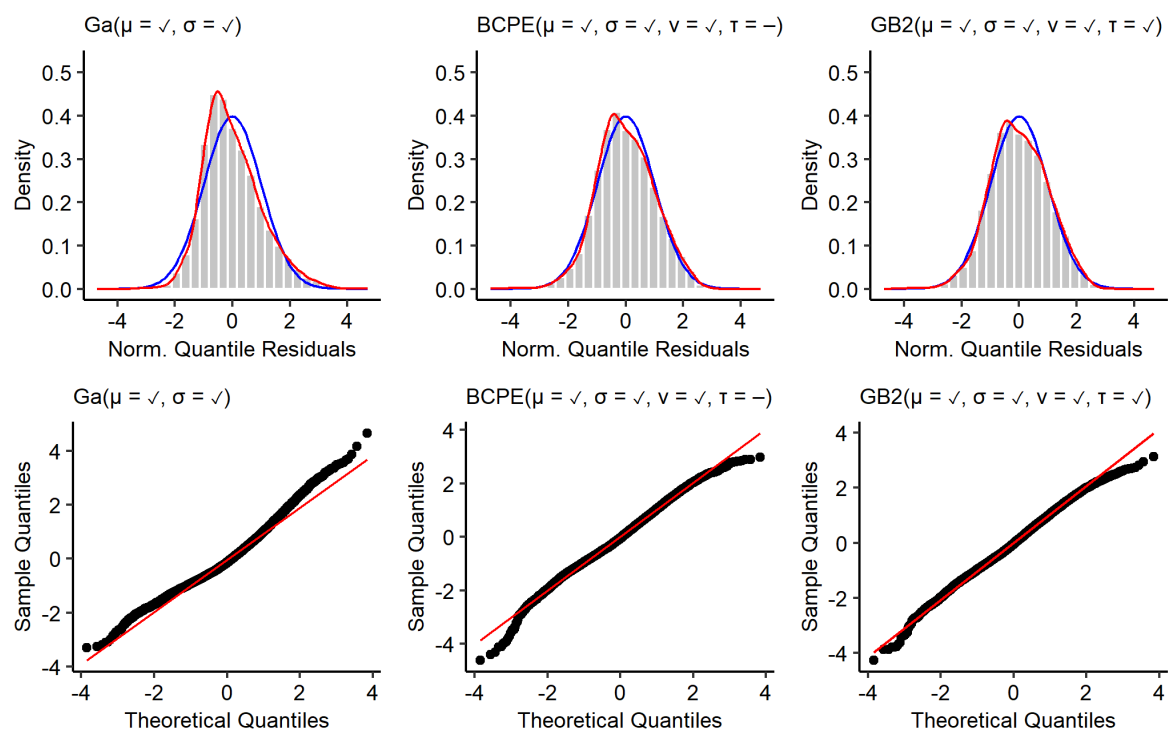

Figure A1. Residuals of models estimated with the male training data. **Top**: Histogram and kernel density estimate of normalised quantile residuals with standard normal distribution (blue). **Bottom**: Quantile–quantile plots.
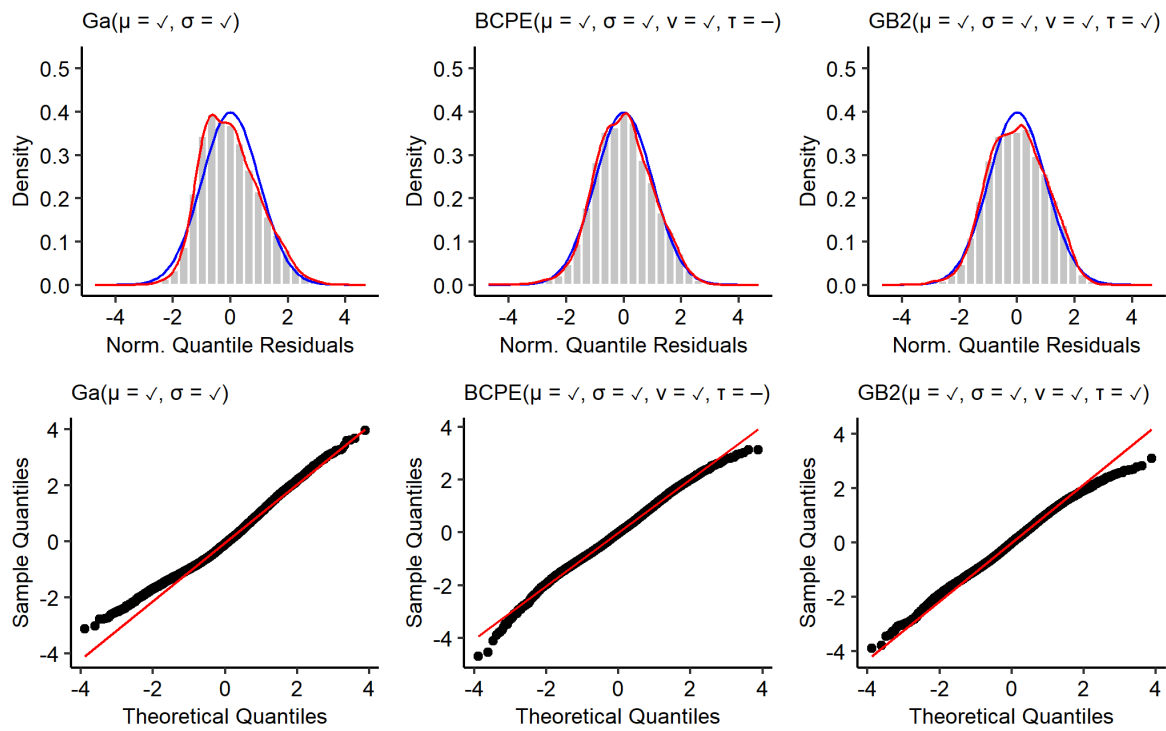
Figure A2. Residuals of models estimated with the female training data. **Top**: histogram and kernel density estimate of normalised quantile residuals with standard normal distribution (blue). **Bottom**: Quantile–quantile plots.

Appendix A.4. Distributional Measures of BCPE and GB2 Models

Figures A3 and A4 display the effect of income on different distributional measures for the BCPE and GB2 models. The general pattern is very similar to the GA model presented in Section 3.3.1. The expected mental health increases with increasing income, while the risks of falling below the discussed thresholds decreases. One can observe that the width of the bootstrap confidence intervals are wider for the BCPE and GB2 model compared to the GA model, which is a result of the increased model complexity.
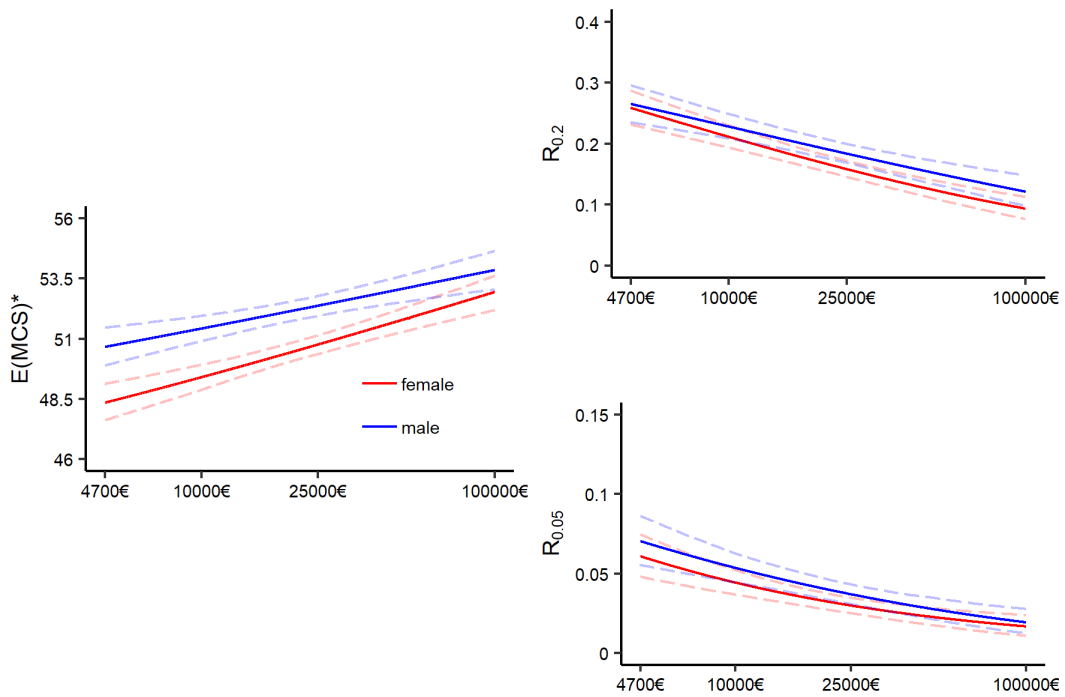
Figure A3. Effect of income on distribution measures for the BCPE-model: BCPE ($\mu = \checkmark$, $\sigma = \checkmark$, $\nu = \checkmark$, $\tau = -$). **Left**: Effect of income on the expectation; **right**: Effect of income on risk of falling below the lowest quintile and lowest vingtile.
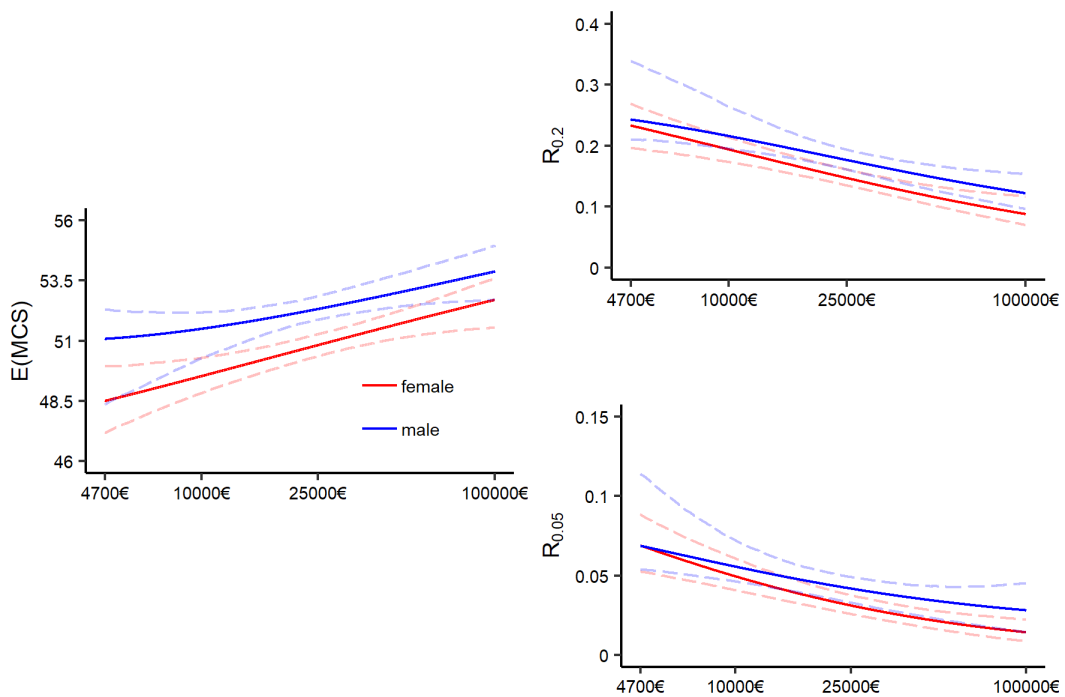

Figure A4. Effect of income on distribution measures for the GB2 model: GB2 ($\mu = \checkmark$, $\sigma = \checkmark$, $\nu = \checkmark$, $\tau = \checkmark$). **Left**: Effect of income on the expectation; **right**: Effect of income on risk of falling below the lowest quintile and lowest vingtile.

Appendix A.5. Bootstrap Confidence Intervals

Bootstrap 0.95 confidence intervals (CI) were calculated within this publication in the following manner: Let θ be the estimate for which a 95%-bootstrap CI should be obtained. θ may be any kind of distributional measure, that is, the expectation or a risk measure, obtained from a prediction based on an estimated GAMLSS model with the training data. Let the training data have N observations.

1. Create a bootstrap sample: Sample with replacement N observations from the training data.
2. Estimate a GAMLSS model with the bootstrap sample.
3. Calculate θ.
4. Repeat steps one to three 2500 times to obtain $B = 2500$ estimates.
5. Calculate the empirical 2.5% and 97.5% quantiles from the distribution of θ resulting from the 2500 bootstrap samples.

One important quantity to discuss is the chosen number of bootstrap samples, B. In general, it holds that the larger the number of bootstrap samples, the more certain one can be to find stable boundaries. The term *stable* hereby means that the boundaries of the confidence intervals do not change remarkably when the number of bootstrap samples is further increased. Figure A5 displays the boundaries of the confidence intervals in dependence of the number of generated bootstrap samples for two exemplarily selected distributional measures resulting from GA models. The graphs on the left of Figure A5 display the boundaries of the confidence intervals for the conditional expectation of average Joe with an annual household net income of 15,000€ (blue) and 30,000€ (red). The graphs on the right display pendants for the risk measure $R_{0.05}$ for average Jane. One can observe that with $B = 1000$ the boundaries have stabilized, and therefore may be sufficient. As a recommendation, $B = 1000$ may serve as rough indication for the minimum amount of bootstrap samples to start with. Since a suitable choice of B is unique to each estimate—and the data and model by which the estimate is produced—bootstrap diagnostics should be applied.
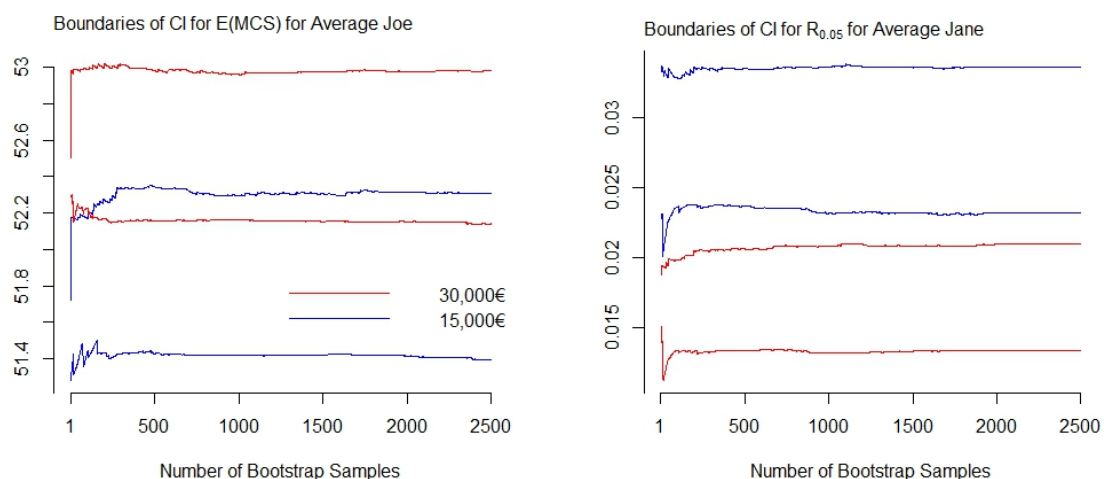

Figure A5. Boundaries of bootstrap confidence intervals for two distributional measures resulting from the GA model with an increasing number of bootstrap samples. **Right**: Expectation of MCS scores; **left**: Risk measure $R_{0.05}$.

References

1. Chetty, R.; Stepner, M.; Abraham, S.; Lin, S.; Scuderi, B.; Turner, N.; Bergeron, A.; Cutler, D. The association between income and life expectancy in the United States, 2001–2014. *JAMA* **2016**, *315*, 1750–1766. [CrossRef] [PubMed]
2. Bird, Y.; Lemstra, M.; Rogers, M.; Moraros, J. The relationship between socioeconomic status/income and prevalence of diabetes and associated conditions: A cross-sectional population-based study in Saskatchewan, Canada. *Int. J. Equity Health* **2015**, *14*, 93. [CrossRef] [PubMed]
3. Kaplan, M.S.; Huguet, N.; Feeny, D.H.; McFarland, B.H. Self-reported hypertension prevalence and income among older adults in Canada and the United States. *Soc. Sci. Med.* **2010**, *70*, 844–849. [CrossRef] [PubMed]
4. Chandola, T.; Jenkinson, C. Validating self-rated health in different ethnic groups. *Ethn. Health* **2000**, *5*, 151–159. [CrossRef]
5. DeSalvo, K.B.; Bloser, N.; Reynolds, K.; He, J.; Muntner, P. Mortality prediction with a single general self-rated health question: A meta-analysis. *J. Gen. Intern. Med.* **2006**, *21*, 267–275. [CrossRef]
6. König, H.H.; Heider, D.; Lehnert, T.; Riedel-Heller, S.G.; Angermeyer, M.C.; Matschinger, H.; Vilagut, G.; Bruffaerts, R.; Haro, J.M.; de Girolamo, G.; et al. Health status of the advanced elderly in six European countries: Results from a representative survey using EQ-5D and SF-12. *Health Qual. Life Outcomes* **2010**, *8*, 143.
7. Ware, J.E., Jr.; Kosinski, M.; Keller, S.D. A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. *Med. Care* **1996**, *34*, 220–233. [CrossRef]
8. Ware, J.E.; Keller, S.D.; Kosinski, M. *SF-36 Physical and Mental Health Summary Scales*; Health Assessment Lab: Boston, MA, USA, 1994.
9. Elgar, F.J.; McKinnon, B.; Torsheim, T.; Schnohr, C.W.; Mazur, J.; Cavallo, F.; Currie, C. Patterns of socioeconomic inequality in adolescent health differ according to the measure of socioeconomic position. *Soc. Indic. Res.* **2016**, *127*, 1169–1180. [CrossRef]
10. Erreygers, G.; van Ourti, T. Measuring socioeconomic Inequality in Health, Health Care and Health Financing by Means of Rank-dependent Indices: A Recipe for Good Practice. *J. Health Econ.* **2011**, *30*, 685–694. [CrossRef]
11. Fleurbaey, M.; Schokkaert, E. Unfair Inequalities in Health and Health Care. *J. Health Econ.* **2009**, *28*, 73–90. [CrossRef]
12. Siegel, M.; Mielck, A.; Maier, W. Individual Income, Area Deprivation, and Health: Do Income-Related Health Inequalities Vary by Small Area Deprivation? *Health Econ.* **2015**, *24*, 1523–1530. [CrossRef] [PubMed]
13. Jacobi, F.; Höfler, M.; Strehle, J.; Mack, S.; Gerschler, A.; Scholl, L.; Busch, M.; Maske, U.; Hapke, U.; Gaebel, W.; et al. Psychische Störungen in der Allgemeinbevölkerung. *Der Nervenarzt* **2014**, *85*, 77–87. [CrossRef] [PubMed]
14. Sareen, J.; Afifi, T.O.; McMillan, K.A.; Asmundson, G.J. Relationship between household income and mental disorders: Findings from a population-based longitudinal study. *Arch. Gen. Psychiatry* **2011**, *68*, 419–427. [CrossRef] [PubMed]
15. Kahneman, D.; Deaton, A. High income improves evaluation of life but not emotional well-being. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16489–16493. [CrossRef]
16. Kushlev, K.; Dunn, E.W.; Lucas, R.E. Higher income is associated with less daily sadness but not more daily happiness. *Soc. Psychol. Personal. Sci.* **2015**, *6*, 483–489. [CrossRef]
17. Erreygers, G.; Kessels, R. socioeconomic Status and Health: A New Approach to the Measurement of Bivariate Inequality. *Int. J. Environ. Res. Public Health* **2017**, *14*, 673. [CrossRef]
18. Koenker, R.; Bassett, G. Regression Quantiles. *Econometrica* **1978**, *46*, 33–50. [CrossRef]
19. Sobotka, F.; Kneib, T. Geoadditive Expectile Regression. *Comput. Stat. Data Anal.* **2012**, *56*, 755–767. [CrossRef]
20. Hothorn, T.; Kneib, T.; Bühlmann, P. Conditional Transformation Models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2014**, *76*, 3–27. [CrossRef]
21. Chernozhukov, V.; Fernandez-Val, I.; Melly, B. Inference on Counterfactual Distributions. *Econometrica* **2013**, *81*, 2205–2268. [CrossRef]
22. Firpo, S.; Fortin, N.M.; Lemieux, T. Unconditional Quantile Regressions. *Econometrica* **2009**, *77*, 953–973. [CrossRef]

23. Stasinopoulos, D.M.; Rigby, R.A. Generalized Additive Models for Location, Scale and Shape (GAMLSS) in R. *J. Stat. Softw.* **2007**, *23*, 1–46. [CrossRef]

24. Klein, N.; Kneib, T.; Lang, S.; Sohn, A. Bayesian Structured Additive Distributional Regression with an Application to Regional Income Inequality in Germany. *Ann. Appl. Stat.* **2015**, *9*, 1024–1052. [CrossRef]

25. Carrieri, V.; Jones, A.M. The Income-Health Relationship 'Beyond the Mean': New Evidence from Biomarkers. *Health Econ.* **2017**, *26*, 937–956. [CrossRef]

26. Heckley, G.; Gerdtham, U.G.; Kjellsson, G. A General Method for Decomposing the Causes of socioeconomic Inequality in Health. *J. Health Econ.* **2016**, *48*, 89–106. [CrossRef]

27. Silbersdorff, A.; Lynch, J.; Klasen, S.; Kneib, T. Reconsidering the Income-Health Relationship using Distributional Regression. *Health Econ.* **2018**, *27*, 1074–1088. [CrossRef]

28. Kessels, R.; Erreygers, G. A Direct Regression Approach to Decomposing socioeconomic Inequality of Health. *Health Econ.* **2019**, *28*, 884–905. [CrossRef]

29. Allen, J.; Balfour, R.; Bell, R.; Marmot, M. Social determinants of mental health. *Int. Rev. Psychiatry* **2014**, *26*, 392–407. [CrossRef]

30. McManus, S.; Meltzer, H.; Brugha, T.; Bebbington, P.; Jenkins, R. *Adult Psychiatric Morbidity in England, 2007: Results of a Household Survey*; The NHS Information Centre for Health and Social Care: Leeds, UK, 2009.

31. Fryers, T.; Melzer, D.; Jenkins, R. Social inequalities and the common mental disorders. *Soc. Psychiatry Psychiatr. Epidemiol.* **2003**, *38*, 229–237. [CrossRef]

32. Lorant, V.; Deliège, D.; Eaton, W.; Robert, A.; Philippot, P.; Ansseau, M. socioeconomic inequalities in depression: A meta-analysis. *Am. J. Epidemiol.* **2003**, *157*, 98–112. [CrossRef]

33. Müters, S.; Hoebel, J.; Lange, C. Diagnose Depression: Unterschiede bei Frauen und Männern. 2013. Available online: https://edoc.rki.de/bitstream/handle/176904/3112/2.pdf?sequence=1&isAllowed=y (accessed on 5 September 2019)

34. Jacobi, F.; Höfler, M.; Siegert, J.; Mack, S.; Gerschler, A.; Scholl, L.; Busch, M.A.; Hapke, U.; Maske, U.; Seiffert, I.; et al. Twelve-month prevalence, comorbidity and correlates of mental disorders in Germany: The Mental Health Module of the German Health Interview and Examination Survey for Adults (DEGS1-MH). *Int. J. Methods Psychiatr. Res.* **2014**, *23*, 304–319. [CrossRef] [PubMed]

35. Jacobi, F.; Höfler, M.; Strehle, J.; Mack, S.; Gerschler, A.; Scholl, L.; Busch, M.A.; Hapke, U.; Maske, U.; Seiffert, I.; et al. Twelve-months prevalence of mental disorders in the German Health Interview and Examination Survey for Adults–Mental Health Module (DEGS1-MH): A methodological addendum and correction. *Int. J. Methods Psychiatr. Res.* **2015**, *24*, 305–313. [CrossRef] [PubMed]

36. Lampert, T.; Kroll, L.E.; Hapke, U.; Jacobi, F. Sozioökonomischer Status und psychische Gesundheit. In *Public Health Forum*; Elsevier: Berlin, Germany, 2014; Volume 22, pp. 6–8.

37. Wirtz, M.A.; Morfeld, M.; Glaesmer, H.; Brähler, E. Normierung des SF-12 Version 2.0 zur Messung der gesundheitsbezogenen Lebensqualität in einer deutschen bevölkerungsrepräsentativen Stichprobe. *Diagnostica* **2018**, *64*, 215–226. [CrossRef]

38. Wood, A.M.; Boyce, C.J.; Moore, S.C.; Brown, G.D. An evolutionary based social rank explanation of why low income predicts mental distress: A 17 year cohort study of 30,000 people. *J. Affect. Disord.* **2012**, *136*, 882–888. [CrossRef]

39. McMillan, K.A.; Enns, M.W.; Asmundson, G.; Sareen, J. The association between income and distress, mental disorders, and suicidal ideation and attempts: Findings from the Collaborative Psychiatric Epidemiology Surveys. *J. Clin. Psychiatry* **2010**, *71*, 1168–1175. [CrossRef]

40. Lee, S.U.; Oh, I.H.; Jeon, H.J.; Roh, S. Suicide rates across income levels: Retrospective cohort data on 1 million participants collected between 2003 and 2013 in South Korea. *J. Epidemiol.* **2017**, *27*, 258–264. [CrossRef]

41. Sareen, J.; Jagdeo, A.; Cox, B.J.; Clara, I.; ten Have, M.; Belik, S.L.; de Graaf, R.; Stein, M.B. Perceived barriers to mental health service utilization in the United States, Ontario, and the Netherlands. *Psychiatr. Serv.* **2007**, *58*, 357–364. [CrossRef]

42. Weich, S.; Lewis, G. Poverty, unemployment, and common mental disorders: Population based cohort study. *BMJ* **1998**, *317*, 115–119. [CrossRef]

43. Lorant, V.; Croux, C.; Weich, S.; Deliège, D.; Mackenbach, J.; Ansseau, M. Depression and socioeconomic risk factors: 7-year longitudinal population study. *Br. J. Psychiatry* **2007**, *190*, 293–298. [CrossRef]

44. Pickett, K.E.; Wilkinson, R.G. Income inequality and health: A causal review. *Soc. Sci. Med.* **2015**, *128*, 316–326. [CrossRef]

45. Ribeiro, W.S.; Bauer, A.; Andrade, M.C.R.; York-Smith, M.; Pan, P.M.; Pingani, L.; Knapp, M.; Coutinho, E.S.F.; Evans-Lacko, S. Income inequality and mental illness-related morbidity and resilience: A systematic review and meta-analysis. *Lancet Psychiatry* **2017**, *4*, 554–562. [CrossRef]

46. Daly, M.C.; Oswald, A.J.; Wilson, D.; Wu, S. Dark contrasts: The paradox of high rates of suicide in happy places. *J. Econ. Behav. Organ.* **2011**, *80*, 435–442. [CrossRef]

47. Lucas, R.E.; Schimmack, U. Income and well-being: How big is the gap between the rich and the poor? *J. Res. Personal.* **2009**, *43*, 75–78. [CrossRef]

48. Matz, S.C.; Gladstone, J.J.; Stillwell, D. Money buys happiness when spending fits our personality. *Psychol. Sci.* **2016**, *27*, 715–725. [CrossRef]

49. Boyce, C.J.; Daly, M.; Hounkpatin, H.O.; Wood, A.M. Money may buy happiness, but often so little that it doesn't matter. *Psychol. Sci.* **2017**, *28*, 544–546. [CrossRef] [PubMed]

50. Westerhof, G.J.; Keyes, C.L. Mental illness and mental health: The two continua model across the lifespan. *J. Adult Dev.* **2010**, *17*, 110–119. [CrossRef] [PubMed]

51. Boyce, C.J.; Brown, G.D.; Moore, S.C. Money and happiness: Rank of income, not income, affects life satisfaction. *Psychol. Sci.* **2010**, *21*, 471–475. [CrossRef]

52. Yu, Z.; Chen, L. Income and well-being: Relative income and absolute income weaken negative emotion, but only relative income improves positive emotion. *Front. Psychol.* **2016**, *7*, 2012. [CrossRef]

53. Sacks, D.W.; Stevenson, B.; Wolfers, J. The new stylized facts about income and subjective well-being. *Emotion* **2012**, *12*, 1181. [CrossRef]

54. Silbersdorff, A. *Analysing Inequalities in Germany*; Springer: Berlin, Germany, 2017.

55. Wagner, G.; Frick, J.; Schupp, J. The German socioeconomic Panel study (SOEP)—Scope, evolution and enhancements. *Schmollers Jahrb.* **2007**, *127*, 139–169.

56. Glemser, A.; Huber, S.; Bohlender, A. *SOEP 2014-TNS Report of SOEP Fieldwork in 2014*; Technical Report, SOEP Survey Papers; SOEP: Berlin, Germany, 2015.

57. Gruebner, O.; Rapp, M.A.; Adli, M.; Kluge, U.; Galea, S.; Heinz, A. Risiko für psychische Erkrankungen in Städten. *Dtsch Arztebl Int.* **2017**, *114*, 121–127. [CrossRef] [PubMed]

58. Paul, K.I.; Moser, K. Unemployment impairs mental health: Meta-analyses. *J. Vocat. Behav.* **2009**, *74*, 264–282. [CrossRef]

59. Paykel, E.; Abbott, R.; Jenkins, R.; Brugha, T.; Meltzer, H. Urban-rural mental health differences in Great Britain: Findings from the National Morbidity Survey. *Psychol. Med.* **2000**, *30*, 269–280. [CrossRef] [PubMed]

60. Peen, J.; Dekker, J.; Schoevers, R.A.; Ten Have, M.; de Graaf, R.; Beekman, A.T. Is the prevalence of psychiatric disorders associated with urbanization? *Soc. Psychiatry Psychiatr. Epidemiol.* **2007**, *42*, 984–989. [CrossRef]

61. Modini, M.; Joyce, S.; Mykletun, A.; Christensen, H.; Bryant, R.A.; Mitchell, P.B.; Harvey, S.B. The mental health benefits of employment: Results of a systematic meta-review. *Australas. Psychiatry* **2016**, *24*, 331–336. [CrossRef]

62. Okulicz-Kozaryn, A. Unhappy metropolis (when American city is too big). *Cities* **2017**, *61*, 144–155. [CrossRef]

63. Andersen, H.H.; Mühlbacher, A.; Nübling, M.; Schupp, J.; Wagner, G.G. Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2. *Schmollers Jahrb.* **2007**, *127*, 171–182.

64. Nübling, M.; Andersen, H.H.; Mühlbacher, A. *Entwicklung eines Verfahrens zur Berechnung der körperlichen und psychischen Summenskalen auf Basis der SOEP-Version des SF 12 (Algorithmus)*; Technical Report, Data Documentation; German Institute for Economic Research: Berlin, Germany, 2006.

65. Cunillera, O.; Tresserras, R.; Rajmil, L.; Vilagut, G.; Brugulat, P.; Herdman, M.; Mompart, A.; Medina, A.; Pardo, Y.; Alonso, J.; et al. Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey. *Qual. Life Res.* **2010**, *19*, 853–864. [CrossRef]

66. Maurischat, C.; Morfeld, M.; Kohlmann, T.; Bullinger, M. *Lebensqualität: Nützlichkeit und Psychometrie des Health Survey SF-36/-12 in der Medizinischen Rehabilitation*; Pabst Science Publ.: Lengerich, Germany, 2004.

67. Fernando, S. *Mental Health, Race and Culture*; Macmillan International Higher Education: London, UK, 2010.

68. Fernando, S. *Mental Health Worldwide: Culture, Globalization and Development*; Springer: Berlin/Heidelberg, Germany, 2014.

69. Heinz, A. *Psychische Gesundheit: Begriff und Konzepte*; Kohlhammer Verlag: Stuttgart, Germany, 2016.

70. Horwitz, A.V. An overview of sociological perspectives on the definitions, causes, and responses to mental health and illness. In *A Handbook for the Study of Mental Health*; Cambridge University Press: New York, NY, USA, 2010; pp. 6–19.

71. Huber, M.; Knottnerus, J.A.; Green, L.; van der Horst, H.; Jadad, A.R.; Kromhout, D.; Leonard, B.; Lorig, K.; Loureiro, M.I.; van der Meer, J.W.; et al. How should we define health? *BMJ* **2011**, *343*, d4163. [CrossRef]

72. Vaillant, G.E. Positive mental health: Is there a cross-cultural definition? *World Psychiatry* **2012**, *11*, 93–99. [CrossRef]

73. WHO. Mental Health: A State of Well-Being. 2014. Available online: http://www.who.int/features/factfiles/mental_health/en/ (accessed on 5th September 2019).

74. Fleishman, J.A.; Lawrence, W.F. Demographic Variation in SF-12 scores: True Differences or Differential Item Functioning? *Med. Care* **2003**, *41*, III75–III86. [CrossRef] [PubMed]

75. Bourion-Bédès, S.; Schwan, R.; Laprevote, V.; Bédès, A.; Bonnet, J.L.; Baumann, C. Differential Item Functioning (DIF) of SF-12 and Q-LES-Q-SF Items among French Substance Users. *Health Qual. Life Outcomes* **2015**, *13*, 172. [CrossRef] [PubMed]

76. Benjamini, Y. Simultaneous and selective inference: Current successes and future challenges. *Biom. J.* **2010**, *52*, 708–721. [CrossRef] [PubMed]

77. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.

78. Silbersdorff, A. *Assessing the Fit of Conditional Distributions Derived by Bayesian Structured Additive Distributional Regression*; ZfS Working Paper, 03/2016; Georg-August-Universität: Göttingen, Germany, 2017.

79. Box, G.E.P. Science and Statistics. *J. Am. Stat. Assoc.* **1976**, *71*, 791–799. [CrossRef]

80. Rigby, R.A.; Stasinopoulos, D.M. Generalized additive models for location, scale and shape. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2005**, *54*, 507–554. [CrossRef]

81. Stasinopoulos, M.D.; Rigby, R.A.; Heller, G.Z.; Voudouris, V.; De Bastiani, F. *Flexible Regression and Smoothing: Using GAMLSS in R*; Chapman and Hall/CRC: Boca Raton, 2017.

82. OECD. *Classifying Educational Programmes: Manual for ISCED-97 Implementation in OECD Countries*; Technical Report; Organisation for Economic Co-operation and Development: Paris, France, 1999.

83. Bach, S.; Corneo, G.; Steiner, V. From Bottom to Top: The Entire Income Distribution in Germany, 1992–2003. *Rev. Income Wealth* **2009**, *55*, 303–330. [CrossRef]

84. Box, G.E.P. Sampling and Bayes' Inference in Scientific Modelling and Robustness. *J. R. Stat. Soc. A* **1980**, *143*, 383–430. [CrossRef]

85. Fenske, N.; Fahrmeir, L.; Rzehak, P.; Höhle, M. *Detection of Risk Factors for Obesity in Early Childhood with Quantile Regression Methods for Longitudinal Data*; Department of Statistics, University of Munich: Munich, Germany, 2008.

86. Thomas, G.; Pereira, A.I.d.A.; Lobos, C.M.V. Analysis of a longitudinal multilevel experiment using GAMLSSs. *arXiv* **2018**, arXiv:1810.03085.

87. Manwell, L.A.; Barbic, S.P.; Roberts, K.; Durisko, Z.; Lee, C.; Ware, E.; McKenzie, K. What is mental health? Evidence towards a new definition from a mixed methods multidisciplinary international survey. *BMJ Open* **2015**, *5*, e007079. [CrossRef]

88. Klein, N.; Kneib, T.; Klasen, S.; Lang, S. Bayesian Structured Additive Distributional Regression for Multivariate Responses. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2015**, *64*, 569–591. [CrossRef]

89. Chalmers, A.F. *Wege der Wissenschaft: Einführung in die Wissenschaftstheorie*; Springer: Berlin/Heidelberg, Germany, 2007

90. Morfeld, M.; Bullinger, M. Der SF-36 Health Survey zur Erhebung und Dokumentation gesundheitsbezogener Lebensqualität. *Phys. Med. Rehabil. Kurortmed.* **2008**, *18*, 250–255. [CrossRef]

91. Bährer-Kohler, S.; Carod-Artal, F.J. *Global Mental Health: Prevention and Promotion*; Springer: Cham, Switzerland, 2017.

92. Diener, E.; Suh, E.M.; Lucas, R.E.; Smith, H.L. Subjective weil-being: Three decades of progress. *Psychol. Bull.* **1999**, *125*, 276–302. [CrossRef]

93. Wittchen, H.U.; Hoyer, J. *Klinische Psychologie & Psychotherapie*; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2011; Volume 2.

94. Hagenaars, A.J.; De Vos, K.; Asghar Zaidi, M. *Poverty Statistics in the Late 1980s: Research Based on Micro-Data*; Office for Official Publications of the European: Luxembourg, 1994.

Int. J. Environ. Res. Public Health **2019**, *16*, 4009

28 of 28

95. Grabka, M.M. *Codebook for the $PEQUIV File 1984–2016: CNEF Variables with Extended Income Information for the SOEP*; Technical Report, Data Documentation; SOEP: Berlin, Germany, 2016.

96. Jacobi, F.; Wittchen, H.U.; Hölting, C.; Höfler, M.; Pfister, H.; Müller, N.; Lieb, R. Prevalence, co-morbidity and correlates of mental disorders in the general population: Results from the German Health Interview and Examination Survey (GHS). *Psychol. Med.* **2004**, *34*, 597–611. [CrossRef] [PubMed]

97. Rigby, R.A.; Stasinopoulos, M.D.; Heller, G.; De Bastiani, F. Distributions for Modelling Location, Scale and Shape: Using GAMLSS in R. 2017. Available online: http://www.gamlss.com (accessed on 5 September 2019).