Journal of Official Statistics, Vol. 38, No. 2, 2022, pp. 599–635, http://dx.doi.org/10.2478/JOS-2022-0027

Iterative Kernel Density Estimation Applied to Grouped Data: Estimating Poverty and Inequality Indicators from the German Microcensus

Paul Walter[1], *Marcus Groß*[1], *Timo Schmid*[2], *and Katja Weimer*[1]

The estimation of poverty and inequality indicators based on survey data is trivial as long as the variable of interest (e.g., income or consumption) is measured on a metric scale. However, estimation is not directly possible, using standard formulas, when the income variable is grouped due to confidentiality constraints or in order to decrease item nonresponse. We propose an iterative kernel density algorithm that generates metric pseudo samples from the grouped variable for the estimation of indicators. The corresponding standard errors are estimated by a non-parametric bootstrap that accounts for the additional uncertainty due to the grouping. The algorithm enables the use of survey weights and household equivalence scales. The proposed method is applied to the German Microcensus for estimating the regional distribution of poverty and inequality in Germany.

Key words: Direct estimation; Interval-censored data; non-parametric estimation; OECD scale; prediction.

1. Introduction

In its Global Risks Report 2017, the World Economic Forum proclaims rising income and wealth disparity as the number one trend in determining global developments, governing the risks of, among others, profound social instability and unemployment (World Economic Forum 2017). Also developed countries, as Germany, are facing an increase in income inequality. Known for stable wages in the 1970s and 1980s (Abraham and Houseman 1995), Germany has faced growing income inequality since its reunification in 1990 (Fuchs-Schiindeln et al. 2010). Yet, the question of how poverty and inequality is defined and can accurately be measured or diagnosed in a society remains debatable, see, for example Hagenaars and Vos (1988) and Lok-Dessallien (1999).

A common way to measure poverty and inequality is the estimation of indicators such as the head count ratio, the poverty gap and the Gini coefficient. Since income information is not easily accessible governments or statistical offices need to conduct surveys or censuses to garner information about personal or household income.

One main difficulty is that, in most societies, income is considered a private topic. In the survey literature, questions about the aspects of income are referred to as "sensitive

[1] Freie Universitat Berlin, Garystraße 21, 14195 Berlin, Germany, Emails: Paul.Walter@fu-berlin.de, Marcus.Gross@inwt-statistics.de and K.Weimer@fu-berlin.de
[2] Otto-Friedrich-Universitat Bamberg, Feldkirchenstraße 21, 96052 Bamberg, Germany. Email: Timo.Schmid@uni-bamberg.de

Acknowledgments: The authors are indebted to the Editor-in-Chief, Associate Editor and the referees for comments that significantly improved the article.

question", therefore item nonresponse is high for these questions. Moore and Welniak (2000) also state that measurement error due to misreporting income is a known issue as bias (usually underreporting) and random error is commonly observed. To counter this, many surveys do not ask for the exact income of their citizens. They ask only for the income group (band) a person or household belongs to, thereby creating a sense of anonymity (Micklewright and Schnepf 2010). Collecting only the grouped information instead of continuous data offers a higher degree of data privacy protection to survey respondents, which lowers response burdens and thus leads to lower item nonresponse rates and higher data quality. On the other hand, the so collected income data is not metric what (a) leads to less information compared to collecting on a metric scale and (b) makes the use of standard formulas for the estimation of poverty and inequality indicators impossible. This problem is also being faced by the Federal Statistical Office of Germany (Destatis) since the German Microcensus, the largest annually carried out household survey in Europe, only collects grouped income data (Statistisches Bundesamt 2018b). Furthermore, there are also censuses in other countries that only collect grouped data, for example, the censuses from New Zealand (Statistics New Zealand 2013), Australia (Australian Bureau of Statistics 2011) and Colombia (Departamento Administrativo Nacional De Estadlstica 2005). Hence, the aim of this article is to provide statistical methodology which enables the direct estimation of statistical indicators from grouped data.

In order to estimate statistical indicators from continuous income data a lot of literature focuses on the parametric estimation of the unobserved distribution. There is a lot of literature focusing on fitting the generalized beta type-II (GB2) distribution (McDonald 1984), its special cases and other distributions, for example, the Singh and Maddala distribution (Singh and Maddala 1976), the Dagum distribution (Dagum 1977), the Weibull distribution (Fréchet 1927) and the generalized gamma distribution (Stacy 1962). Reed and Wu (2008), Kleiber (2008) and Chen (2017) are primarily focusing on the estimation of statistical indicators from grouped data by fitting a parametric distributions to the data. Walter et al. (2021) estimate linear and non-linear indicators for small areas by a nested error regression model when the response variable is grouped.

Kakwani and Podder (2008) argue against the parametric estimation of the income distributions from grouped data due to its lack of precision and present a method that can be utilized to estimate the Lorenz curve directly from the grouped data in order to compute inequality indicators. Another alternative to the parametric estimation is to use non-parametric methods to model income instead. Although most authors do not directly address the topic of grouped data, there is much literature about rounded data, which is obtained from grouped data by substituting the groups with their centers. Hall (1982), Scott and Sheather (1985), and Hall and Wand (1996) study the effects of rounded data on non-parametric kernel density estimation (KDE). They find that censoring affects the bias rather than the variance of the estimate. Additionally, Hall and Wand (1996) present minimum grid sizes for KDE which are needed to achieve a given degree of accuracy. Grid size corresponds to the amount of points and therefore to the amount of groups when the group centers are used on which the density is estimated. Wang and Wertelecki (2013) point out that standard KDE leads to increasingly spiky density estimates at rounded points with a growing sample size. To overcome this issue, they propose a bootstrap-type kernel density estimator and show in a simulation study that the estimator provides better accuracy than the standard KDE in the context of rounded data. Groß et al. (2017) melt the principle of stochastic expectation-maximization algorithms (Nielsen 2000) with KDE to propose a density estimation algorithm for rounded two-dimensional geo-coded data.

From a theoretical perspective, we propose a non-parametric KDE algorithm for grouped data. The proposed method extends the ideas of Groß et al. (2017) from rounded to grouped data. The KDE algorithm enables the estimation of poverty and inequality indicators from grouped data under different censoring schemes and varying group widths (instead of fixed rounding schemes as in Groß et al. 2017). For the estimation of the standard errors of the statistical indicators we propose a non-parametric bootstrap. From an applied perspective, the algorithm allows for the use of equivalence scales, for example, the modified Organisation for Economic Co-operation and Development (OECD) scale, to make income of households of different sizes comparable. Moreover, to obtain representative results, survey or design weights can be used in the estimation process.

The article is structured as follows. In Section 2, the German Microcensus data set and the estimation problem is presented. In Section 3, the KDE algorithm and the proposed non-parametric bootstrap are introduced. In Section 4, the algorithm is then applied to the German Microcensus for estimating the regional distribution of poverty and inequality in Germany. In Section 5, the performance of the KDE algorithm and the bootstrap are evaluated by using Monte Carlo simulation studies under different grouping schemes and different theoretical distributions. A final discussion of the major results, their implications, and an outlook is given in Section 6.

2. The German Microcensus

The German Microcensus is a survey that is conducted annually by the Federal Statistical Office of Germany (Statistisches Bundesamt 2018a). The survey has a long history and was first carried out in 1957 (Statistical Offices of the Federation and the Federal States 2016). The Microcensus is designed as a single-stage cluster sample (Schimpl-Neimanns 2010). The primary sampling units (clusters) are composed of neighboring buildings. From the drawn clusters all households are samples. The clusters are arranged (stratified) by grouping buildings into four different categories (based on size) and by regional characteristics (population size). The sampling of the clusters is conducted as simple random sampling from strata. The post-stratification weights used in the application account for nonresponse. The total sample size is equal to 1% of the German population. This amounts to about 820,000 household members. It is the largest annually conducted household survey in Europe (Statistisches Bundesamt 2018b). The aim of the Microcensus is to provide data on a regular short-term basis on regional level in Germany. Topics covered by the survey are: demographic background, migration, employment, education, poverty, and vocational training (Schwarz 2001). For most questions, answering is compulsory by law, however, there are also questions that are answered on a voluntary basis, such as questions about health status, health insurance, housing situation, and retirement programs (Statistical Offices of the Federation and the Federal States 2016).

Researchers appreciate the Microcensus data set for very low nonresponse rates and high data quality (Schwarz 2001). While low non-response rates are guaranteed by mandatory responses for most questions, high data quality is achieved with face-to-face interviews. Although the Microcensus is valued by many researchers, analyzing the data properly is problematic when the research focuses on income. This is due to the fact that both personal and household income are only observed as an grouped variable. Furthermore, the censoring scheme and the number of groups has changed over time, which makes the longitudinal analyses even more complicated (Boehle 2015). Some researchers even say that because of the grouping of the income variable, the Microcensus is unsuitable for valid research on the topic of income (Stauder and Huning 2004). Therefore, estimation of poverty and inequality indicators is regularly based on alternative surveys, for example, the Socio-Economic Panel (SOEP) or the Income, Receipts, Expenditure survey. In contrast to the Microcensus, participation is voluntary and participants are asked for their exact income (not grouped), which enables the estimation of poverty and inequality indicators using standard formulas. However, the sample sizes of these surveys are smaller leading to higher uncertainty of the estimates compared to the Microcensus.

The data used in this article comes from the Scientific-Use-File (SUF) from 2012, a 70% sample of the German Microcensus (Statistisches Bundesamt 2017). The variable of interest is the monthly household income. All estimation results in the article are based on the monthly income. As previously mentioned, the variable household income is grouped to 24 groups. The distribution is visualized in Figure 1. It is notable that the group width increases with increasing income. While the lowest group has a width of 150 euro the second last group has a width of 8,000 euro. That implies that the distribution is actually highly right skewed which is not clear at the first glance when looking at Figure 1. However, plotting the distribution as histogram or as density plot would also come with caveats because of the grouping of the data and the open-ended upper interval. We therefore choose barplots for presenting the grouped income data in this paper. After data cleaning, we are left with a sample size of $n_{Germany} = 454852$. Since interests also lie in the spatial distribution of poverty and inequality the statistical indicators are estimated for each federal state separately and at the national level.

The sample size for each federal state and its location is given in Table 10 and Figure 8 in the Appendix, Section 7. The sample sizes are very large for each federal state even for Bremen, the state with the smallest sample size $n_{Bremen} = 3356$.

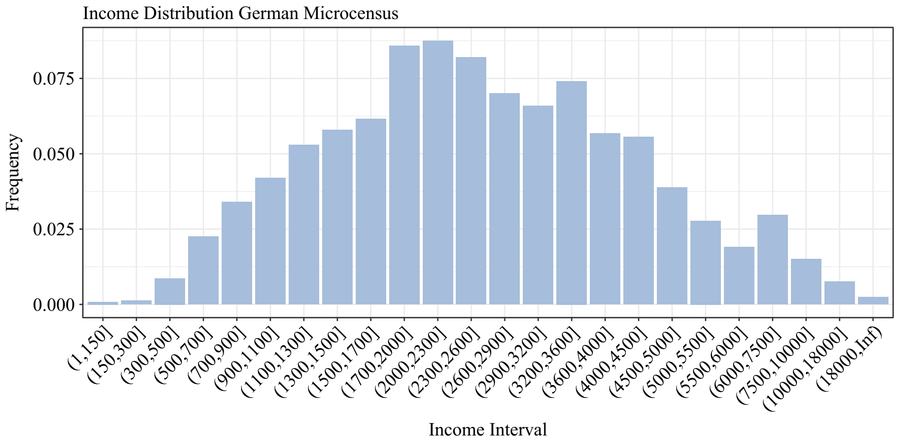


Fig. 1. Distribution of the grouped household income from the German microcensus data set. Numbers are given in euro.

In Section 4 we show that the proposed KDE algorithm enables the estimation of statistical indicator from grouped German Microcensus data. This allows researchers and practitioners to use the German Microcensus for the further and more in-depth investigation of poverty and inequality in Germany.

3. Methodology

In Subsection 3.1, we propose a novel KDE algorithm to generate metric pseudo samples from the observed grouped data. By using the pseudo samples, statistical indicators can be estimated applying standard formulas. The non-parametric bootstrap for the variance estimation of the statistical indicators is introduced in Subsection 3.2. Finally, the incorporation of survey weights and the use of household equivalence scales are discussed in Subsection 3.3.

3.1. Kernel Density Estimation from Grouped Data

Kernel density estimation is one of the most established non-parametric density estimation techniques in the literature and was first introduced by Rosenblatt (1956) and Parzen (1962). It is applied to estimate a continuous density from a random variable with density $f(x)$ directly from its independent and identically distributed observations without making any prior assumptions about its distributional family. Let $X = \{X_1, X_2, . . ., X_n\}$ denote a sample of size n. For $i = 1, . . ., n$ the KDE is defined as

$$\hat{f}_h(x) - \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right), \tag{1}$$

where $k(\cdot)$ is a kernel function and the bandwidth is denoted by $h > 0$. Selection methods for the bandwidth are widely discussed in the literature with the two main categories being plug-in and cross-validation (Jones et al. 1996; Loader 1999; Henderson and Parmeter 2015). The basic idea of the first is to minimize the asymptotic mean integrated squared error whilst substituting the unknown density in the optimization with a pilot estimate, whereas the second method is a more data-driven approach, for example, utilizing leave-one-out cross-validation.

In the KDE (1), it is assumed that observations are taken directly from the continuous distribution that is to be estimated. Often, however, collecting continuous data is not possible due to various restrictions in practice, such as, for example, confidentiality concerns. In these situations we are left with grouped data, where only the grouped information is observed. Thus, only the lower A_{K-1} and upper A_K group bounds (A_{K-1}, A_K) of X_i are observed and its continuous value remains unknown. The continuous scale is divided into n_K groups. The variable $K(1 \leq K \leq n_K)$ indicates which of the groups an observation of X_i falls into. Note that applying KDE (1) to the group midpoints of the grouped data falsely allocates too much probability mass to the midpoints and too little to the unobserved X_i. This leads to spiky estimates, unless the bandwidth is chosen to be very large (Wang and Wertelecki 2013). Increasing the bandwidth cannot be considered as a solution to this problem because this causes additional loss of information about the underlying true distribution.

We aim to estimate the unknown density $f(x)$ from which the sample X is drawn only based on the observed group information K. From Bayes' theorem it follows that the conditional distribution of X given K is:

$$\pi(X|K) \propto \pi(K|X)\pi(X),$$

where $\pi(K|X)$ is defined as a product of Dirac distributions $\pi(K|X) = \prod_{i=1}^{n} \pi(K_i|X_i)$ with

$$\pi(K_i|X_i) = \begin{cases} 1 & \textit{if } A_{K-1} \leq X_i \leq A_K, \\ 0 & \text{otherwise}, \end{cases}$$

for $i = 1, . . ., n$. Using this formulation pseudo samples of the unknown X_i are drawn from the estimated density that enable the estimation of any statistical indicator. Since $\pi(X) = \prod_{i=1}^{n} f(X_i)$ is initially unknown, an initializing estimate $\hat{f}_h(x)$ that is based on the group midpoints, serves as a proxy. After that, the pseudo samples drawn from $\pi(X|K)$ are used to re-estimate $\pi(X)$. The following section focuses on the exact implementation of the proposed algorithm and discusses similarities to the EM algorithm by Dempster et al. (1977) and the stochastic EM (SEM) algorithm by Celeux and Diebolt (1985) and Celeux et al. (1996).

3.1.1. Estimation and Computational Details

To fit the model pseudosamples of X_i are drawn from the conditional distribution

$$\pi(X_i|K_i) \propto \mathbf{I}(A_{K-1} \leq X_i \leq A_K)\, f(X_i),$$

where $\mathbf{I}(\cdot)$ denotes the indicator function. The conditional distribution of X_i given K_i is the product of a uniform distribution and density $f(x)$. As $f(x)$ is unknown it is replaced by $\hat{f}_h(x)$, an estimate that is obtained by the prior defined kernel density estimator. Hence, X_i is iteratively drawn from the known group (A_{K-1}, A_K) with the current density estimate $\hat{f}_h(x)$ used as sampling weight. The steps of the iterative algorithm are described below.

Step 1: Use the midpoints of the groups as pseudo $\tilde{X}_i$ for the unknown X_i. Obtain a pilot estimate of $\hat{f}_h(x)$, by applying KDE with a gaussian kernel. Choose a sufficiently large bandwidth h, such that no rounding spikes occur. We propose to use double the width of the widest finite interval.

Step 2: Evaluate $\hat{f}_h(x)$ on an equal-spaced fine grid $G = \{g_1, . . ., g_j\}$ with j grid points $g_1, . . ., g_j$. The width of the grid is denoted by δ_g. It is given by,

$$\delta_g = \frac{|A_0 - A_{n_K}|}{j-1},$$

and the grid is defined as,

$$G = \{g_1 = A_{0,} g_2 = A_0 + \delta_{g,} g_3 = A_0 + 2\delta_g, . . ., g_{j-1} = A_0 + (j-2)\delta_g, g_j = A_{n_K}\}.$$

Step 3: Sample from $\pi(X\ |K)$ by drawing a pseudo sample $\tilde{X}_i$ randomly from $\{G_K = g_j | g_j \in (A_{K-1}, A_K)\}$ with weights $\hat{f}_h(\tilde{X}_i)$ for $K = 1, . . ., n_K$. The sample size within each group is given by the number of observations within each group.

Step 4: Estimate any statistical indicator of interest $\hat{I}$ using the pseudo $\tilde{X}_i$.

Step 5: Recompute $\hat{f}_h(x)$ with bandwidth h computed by the plug-in estimator (or any other method of choice), using the pseudo samples $\tilde{X}_i$ obtained in iteration Step 3

Step 6: Repeat Steps 2–5, with $B(_{KDE})$ burn-in and $S(_{KDE})$ additional iterations.

Step 7: Discard the $B(_{KDE})$ burn-in iterations and estimate the final $\hat{I}$ by averaging the $\hat{I}$ obtained during the $S(_{KDE})$ iteration steps.

The KDE algorithm estimates the distribution of a grouped variable by only using the group information. An algorithm that is widely used for models that depend on latent variables (in our case the unobserved grouped X) is the EM algorithm (Dempster et al. 1977). In the EM algorithm the expectation of $X|K$ is obtained analytically. However, in the context of kernel density estimation this does not work, that is, producing biased, spiky density estimates, because all values inside a group would be concentrated at one point, the expectation. In a SEM algorithm, the analytical E-step from the EM algorithm is replaced by the drawing of pseudo samples (Celeux and Diebolt 1985; Celeux et al. 1996). In case of the KDE algorithm, it is drawn from the distribution of $\pi(X|K)$. Hence, the proposed KDE algorithm has similarities to a SEM algorithm. In its common form, the EM and SEM algorithm are used for maximum likelihood (ML) estimation with unobserved data. McLachlan and Krishnan (2008) proposed a generalization of the SEM algorithm that can be used with surrogates for the M-step, that is, the maximization of the expected likelihood given the distribution of the unknown true values X. In the KDE algorithm the minimization of the asymptotic mean integrated squared error given the pseudo samples is used as such a surrogate by applying kernel density estimation with plug-in bandwidth on the pseudosamples. To assess convergence of the algorithm, one can do this visually as demonstrated in Figure 2 or by using convergence criteria used in the MCMC literature as the SEM algorithm is a Markov Chain (Nielsen 2000).

We would like to mention that there are various approaches described in the survival analysis literature to estimate the hazard or survival function from interval-censored data using an EM algorithm or smoothing techniques (Betensky et al. 1999; Braun et al. 2005; Li et al. 1997; Pan 2000). However, these papers focus on interval-censored data (overlapping intervals) and not on grouped data. When using interval-censored data the problem of spiky density estimates (Groß and Rendtel 2016) is less relevant because of the much higher number of intervals. Nevertheless, using the proposed KDE algorithm in these situations, instead of an EM algorithm could have an positive effect on the estimation results.

3.2. Variance Estimation

This section introduces a method for the variance estimation of the statistical indicators that are estimated by the KDE algorithm. A common way to estimate the variance, if X is observed on a continuous scale, is linearization. Taylor linearization (Tepping 1968; Woodruff 1971; Wolter 1985; Tille 2001) is a well- known and commonly applied method for the estimation of variance for non-linear indicators, such as ratios or correlations. However, the method cannot be applied for the variance estimation of all non-linear indicators. For the variance estimation of mathematically more complex indicators, for example, the Gini coefficient, Deville (1999) introduced the generalized linearization method. The generalized linearization method is also used by Eurostat for the variance estimation of complex indicators (Osier 2009). Nevertheless, linearization cannot be applied when the variable of interest is observed as an grouped variable (Lenau and Münnich 2016). To still produce variance estimates, resampling methods, such as bootstrapping can be applied as an alternative. In the following, a non-parametric bootstrap introduced by Efron (1979) and Shao and Tu (1995) is used for the variance estimation of the indicators estimated by the KDE algorithm. Also, any confidence interval can be estimated by using the quantiles from the bootstrap results. The non-parametric bootstrap is based on the assumption that the drawn sample is representative of the population. Therefore, the empirical distribution function $\hat{F}$ is a non-parametric estimate of the population distribution F. The desired poverty indicator of interest $\hat{I}$, is the empirical estimate of the true parameter. The bootstrap standard errors are calculated as follows:

Step 1: Draw with replacement a bootstrap sample of the grouped $X_i^{(b)}$ of size n from the sample data set.

Step 2: Apply the KDE algorithm to the bootstrap grouped sample $X_i^{(b)}$ for the estimation of any indicator $\hat{I}^{(b)}$ of interest.

Iterate Steps 1-2, $b = 1, . . ., B$ times and estimate the standard error

$$se(\hat{I}) = \sqrt{\frac{\sum_{b=1}^{B} (\hat{I}^{(b)} - \bar{I}^{(b)})^2}{B}} \text{ with } \bar{I} = \frac{1}{B} \sum_{b=1}^{B} \hat{I}^{(b)}.$$

Please note that the proposed bootstrap does not account for complex survey design. However, survey weights can be included into step 2 of the above procedure as described in the following subsection. For complex designs there exist different approaches in the literature. For an overview we refer to Mashreghi et al. (2016).

3.3. Using Survey Weights and Household Equivalence Scales in the Estimation Process

In the following we assume a superpopulation model (Dorfman and Valliant 2005) and thus our sample is assumed to be drawn from an infinite population. We emphasize here, that a kernel density estimate on a given sample X is not necessarily an unbiased estimate of the population density as the sampling design may over- or undersample certain parts of the population. However, it can be useful to recover a *sample* population density. The KDE algorithm and the bootstrap can be extended to enable the estimation of weighted statistical indicators and its standard errors from grouped data and to allow for the usages of household equivalence scales.

In order to estimate weighted indicators the KDE algorithm draws, as described before, pseudo samples from the conditional distribution of $\pi(X_i|K_i)$. However, differently then stated in Subsection 3.1, pseudo $\tilde{X}_i$ are drawn together with its corresponding survey weight w_i, for $i = 1, . . ., n$. Hence, a sample of $(\tilde{X}_i, w_i)$ is obtained in each of the $B(_{KDE}) + S(_{KDE})$ iteration steps of the KDE algorithm. Using the $\tilde{X}_i$ one is actually estimating the so-called sample density function as described in Pfeffermann and Sverchkov (1999). To get a point estimate for population measures one can weight the sample observations by the inverse sample selection probabilities, that is, survey weights w_i (Pfeffermann et al. 1998). Thus, by using this formulation of the KDE algorithm any weighted statistical indicator $\hat{I}$ can be estimated in step 4, based on $(\tilde{X}_i, w_\mathrm{i})$. The final weighted indicator $\hat{I}$ is then computed by averaging the obtained $S(_{KDE})$ estimates.

Regarding the estimation of the variance the bootstrap described in Subsection 3.2 needs to be adjusted. In step 2 of the bootstrap algorithm the weighted statistical indicator has to be estimated by the KDE algorithm taking the survey weights from the bootstrap sample into account. Whenever there is not much variation in the survey weights this naive non-parametric bootstrap provides reasonable results (Alfons and Templ 2013). However, in cases with large variation in the survey weights a calibrated bootstrap might be preferable. Moreover, when complex survey designs are present as in the German Microcensus example (cf. beginning of Section 2), the proposed bootstrapping method is not optimal. Nevertheless, other bootstrapping methods accounting for sampling designs should be straightforward to implement. We refer to for example, Field and Welsh (2007) for bootstrapping on clustered data. An alternative to the proposed scheme for incorporation of survey weights would be to apply the weights when estimating the kernel density as in Buskirk and Lohr (2005). The results in our application were virtually the same.

When working with household income the size of the household (in terms of people belonging to the household) needs to be considered for reasonable inference. To make household income comparable between households of different sizes, household equivalence scales, for example, the OECD scale can be used to estimate equivalised household income. The household incomes need to be equivalised prior to the application of the KDE algorithm. Therefore each household's group bounds are divided by its corresponding equivalence scale weight. For instance, a household within grouped (1,000; 2,000] with a OECD weight of 2 has equivalence group bounds of (500; 1,000]. This procedure can lead to overlapping group bounds (interval-censored data). A straightforward generalization to the proposed algorithm is necessary by defining lower and upper bounds A_{K_-} and A_{k_+} separately for each K replacing the bounds A_{K-1} and A_K. Thus, overlapping of the groups does not impede the use of the KDE algorithm since it simply draws from each unique group with a sample size equal to the number of observation (households) within each group. Using survey weights in the KDE algorithm after the estimation of the equivalised income is possible as described above.

4. Estimating Poverty and Inequality Indicators Using German Microcensus Data

In this section poverty and inequality indicators are estimated at the federal state level in Germany using grouped household income data from the German Microcensus we described in Section 2. Before applying the KDE algorithm to the grouped data we make income of households of different sizes comparable to each other. The distribution of the OECD household size is given in Table 1. The median household has a size of 1.5 and the largest household a size of 12.2. The KDE algorithm is applied to the equivalised grouped household income data as described in Subsection 3.3. Furthermore, for representative results the survey weights are used for the estimation of weighted statistical indicators as described in Section 3.3. The distribution of the survey weights is shown in Table 2 with

Table 1. Distributions of the OECD household size.

Min.	$Q_{.25}$	Median	Mean	$Q_{.75}$	Max.
1.0	1.5	1.5	1.7	2.1	12.2

Table 2. Distributions of the extrapolations factors.

$Q_{.10}$	$Q_{.25}$	Median	Mean	$Q_{.75}$	$Q_{.90}$
139.8	151.6	166.0	168.7	182.5	199.0

50% of the survey weights having a value between 151.6 and 182.5. Since there is not much variation in the survey weights the standard errors of the statistical indicators are estimated with the non-parametric bootstrap as described in Subsection 3.2. The number of bootstrap samples is set to 100. This number gives stable results as shown in the simulation study in Section 5.

The following statistical indicators, which are frequently used by statistical institutes in Germany to measure poverty and inequality, are estimated based on $(\tilde{X}_i, w_i)$. Where $\tilde{X}_i$ is the metric pseudo data generated by the KDE algorithm and w_i is the observed corresponding survey weight. The weighted mean and the weighted quantiles (10%, 25%, 50%, 75%, 90%) are given by

$$\hat{I}_{Mean} = \frac{\sum_{i=1}^{n} w_i \tilde{X}_i}{\sum_{i=1}^{n} w_i}, \tag{2}$$

$$\hat{I}_{\mathrm{Q}(p)} = \begin{cases} \frac{1}{2}\left(\tilde{X}_i + \tilde{X}_{i+1}\right) & \text{if } \sum_{j=1}^{i} w_j = p \sum_{j=1}^{n} w_j, \\ \tilde{X}_{i+1} & \text{if } \sum_{j=1}^{i} w_j \leq p \sum_{j=1}^{n} w_j \leq \sum_{j=1}^{i+1} w_j, \end{cases} \tag{3}$$

where p denotes the quantile $p \in (0, 1)$. The weighted poverty measures head count ratio (HCR) and poverty gap (PGap) (Foster et al. 1984) are given by

$$\hat{I}_{\mathrm{HCR}} = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \mathrm{I}\left(\tilde{X}_i \leq z\right), \tag{4}$$

$$\hat{I}_{\mathrm{PGap}} = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \left(\frac{z - \tilde{X}_i}{z}\right) \mathbf{I}\left(\tilde{X}_i \leq z\right), \tag{5}$$

where $\mathbf{I}(\cdot)$ denotes the indicator function. The HCR and PGap include a threshold z that is known as the poverty line. For the application a regional relative poverty line, defined as 60% of the median of the equivalised grouped household income is chosen. This corresponds to the EU definition (Eurostat 2014). The HCR is a measure of the percentage of observations (individuals or households) below the poverty line, whereas the PGap measures the average distance of those observations from the poverty line. Inequality is commonly measured by the Gini coefficient (Gini 1912) and the quintile share ratio (QSR). The weighted indicators are estimated by

$$\hat{I}_{\mathrm{Gini}} = \left[\frac{2\sum_{i=1}^{n}\left(w_i x_i \sum_{j=1}^{i} w_j\right) - \sum_{i=1}^{n} w_i^2 \tilde{X}_i}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i \tilde{X}_i} - 1\right], \tag{6}$$

$$\hat{I}_{\mathrm{QSR}} = \frac{\sum_{i=1}^{n} \mathrm{I}\left(X_i \geq \hat{Q}_{0.8}\right) w_i \tilde{X}_i}{\sum_{i=1}^{n} \mathrm{I}\left(X_i \leq \hat{Q}_{0.2}\right) w_i \tilde{X}_i}. \tag{7}$$

The range of the Gini coefficient lies between 0 and 1. The higher its value, the higher the inequality. If the Gini coefficient is equal to 0 there is perfect equality in the data, whereas a Gini coefficient of 1 indicates perfect inequality. The QSR is the ratio of observations richer than 20% of the richest observations to the 20% of the poorest observations. Higher values of the QSR indicate higher inequality.

The KDE algorithm is applied with $B(_{KDE}) = 80$ burn-in iterations and $S(_{KDE}) = 400$ additional iteration. The number of $B(_{KDE})$ and $S(_{KDE})$ is sufficiently large as is seen in the convergence plot in Figure 2. Both indicators have converged after 480 iterations. While indicators that are dependent on the whole distribution converge more slowly (for example, the Gini coefficient), indicators that depend only on the unknown distribution of the data within one group (for example, the HCR) converge faster. Also, all other indicators are graphically checked for convergence. The number of grid points is set to $j = 4{,}000$. Simulation results in Section 5 show that choosing this number leads to reliable estimates when working with income type data.

Furthermore, the KDE algorithm cannot handle open-ended groups. Lower bounds equal to -∞ or upper bounds equal to $+\infty$ have to be replaced by a finite number. The chosen value affects the performance of the KDE algorithm. However, not all poverty and inequality indicators depend on the outer groups. Indicators that depend on the outer groups are indicators that depend, by their definition, on the whole distribution for example, the mean or the Gini coefficient. These indicators are always influenced by the way in which open-ended groups are handled, whereas other indicators, such as the median, are only affected if they fall into one of the open-ended outer groups. In the application we replace $+\infty$ of the upper group with a value of three times the value of the upper groups lower bound. Hence, the upper bound of the upper income group of the German Microcensus (18,000, + ∞) is replaced with the value of $18{,}000 * 3 = 54{,}000$, resulting in the group (18,000; 54,000] which is then used in the estimation process by the KDE algorithm. In an application the practitioner should choose the group bounds for open-ended groups with regard to content and to the censoring scheme. However, our experiences running several simulations (see Section 5) indicate that a value of three times the value of the lower bound serves as a good proxy when working with grouped income data. A further hybrid possibility that requires

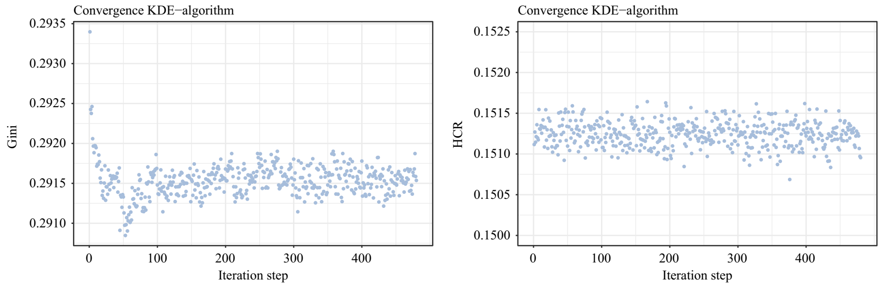


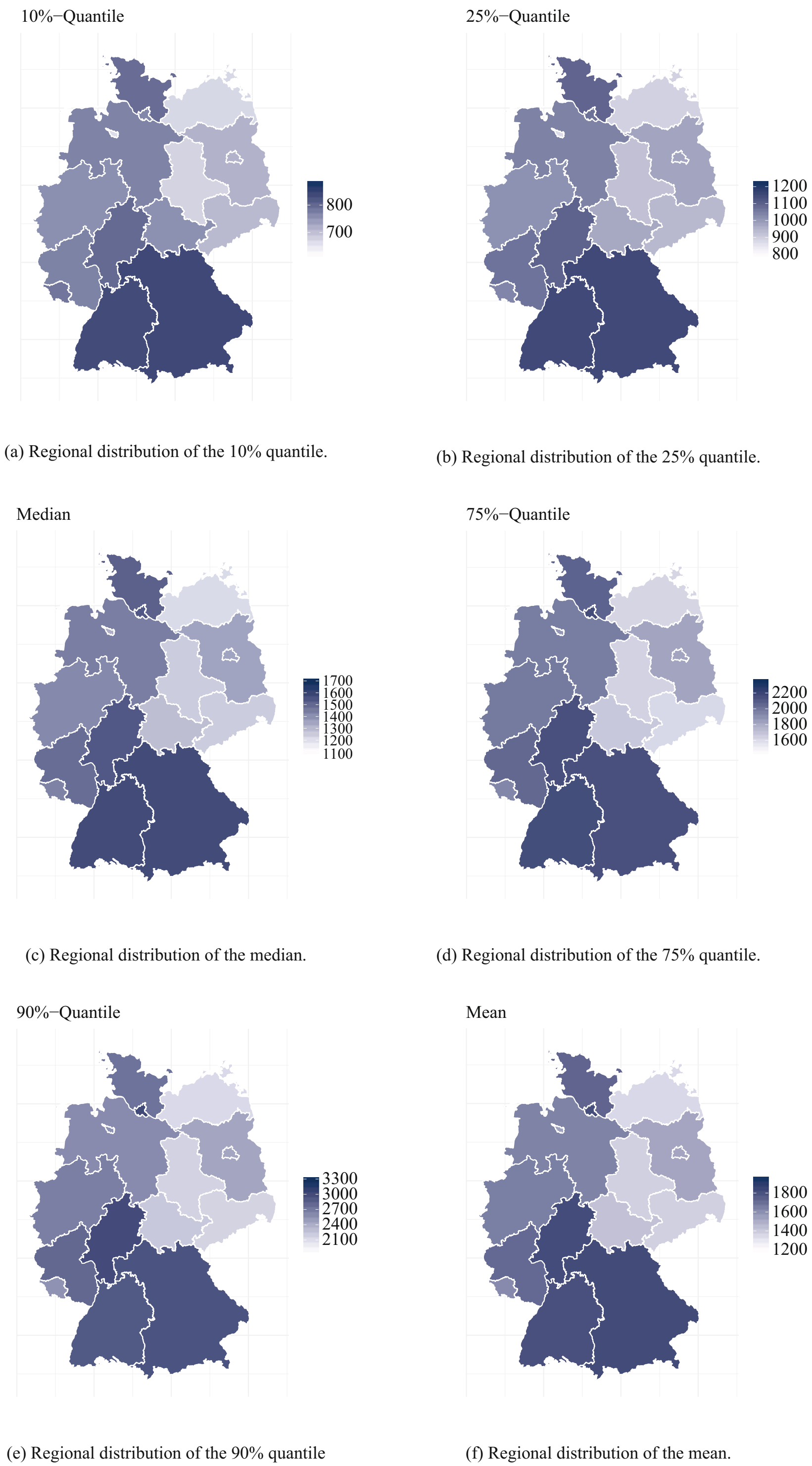
Fig. 2. Convergence of the KDE algorithm for the Gini coefficient and the HCR.

parametric assumptions about the underlying distribution would be to model open-ended groups with a parametric distribution. We did not consider that approach because we want to stick with a –flexible– purely non-parametric algorithm.

All estimated indicators are presented in Figure 3 and 4 and the exact values and the estimated standard errors are given in Appendix (Section 7) in Table 11. The indicator that is frequently used by the European Union to describe poverty is the HCR, while inequality is most frequently measured by the Gini and the QSR. At the national level the HCR $= 0.15$, the Gini $= 0.29$ and the QSR $= 4.31$. To put these numbers in perspective we can compare Germany to other countries from the European Union. According to OECD data, the country with the highest estimated Gini coefficient in the European Union is Lithuania with a Gini of 0.38 and the country with the lowest Gini coefficient is Slovak Republic with an estimated Gini coefficient of 0.24 (OECD 2018). This leaves Germany with an average Gini coefficient. Regarding the estimated QSR the same holds true. The country with the highest QSR is Lithuania with a QSR of 7.5 and the country with the lowest QSR is Slovenia with 3.6, leaving Germany with an average QSR (OECD 2018). The estimated HCR is slightly lower then the European Union average, which was 16.8 in 2012 (Eurostat 2018). Hence, based on the estimated indicators Germany shows average strength of poverty and inequality compared to other countries from the European Union.

Owing to the large sample size, reliable estimates for smaller geographical areas (federal state level) can be produced to evaluate the regional distribution of poverty and inequality in Germany in detail. The quantiles and the mean indicate that the East (formerly German Democratic Republic, GDR) is poorer than the West. This result is commonly known in Germany and is not very surprising. Nevertheless, Brandenburg and Berlin have higher incomes than the rest of East Germany (Mecklenburg-Vorpommern, Saxony, Saxony-Anhalt and Thuringia) due to the economically growing city Berlin. However, there are also federal states in the West with comparable low incomes. For instance Bremen, shows low income for the 10% and 25% quantile in comparison to the rest of West Germany, while for the higher quantiles Bremen shows similar results to the rest of West Germany. The poorest states with a median of 1,211.29 euro and 1,247.05 euro are Mecklenburg-Vorpommern and Saxony-Anhalt and the richest ones with a median of 1,580.43 euro and 1,580.35 euro are Baden-Wurttemberg and Bavaria. For the estimation of the HCR and PGap, a regional poverty line (defined as 60% of the median) is used. The HCR indicates that in the East fewer people live under the regional poverty line than in the West. Also, the people living under the poverty line live closer to it in the East, as shown by the PGap. When looking at the QSR and the Gini coefficient, the East-West trend is less striving. Nevertheless, the states in the East have lesser income inequality. The most unequal states with a Gini coefficient of 0.32 and 0.31 are Hamburg and Bremen and the most equal ones with a Gini coefficient of 0.25 and 0.25 are Saxony and Thuringia. The estimated standard errors of the indicators on state areas are quite small. Therefore, estimating indicators for smaller geographical areas would also be desirable, in order to get an even closer look at the geographical distribution of poverty and inequality. However, the PUF does not include regional identifiers below the federal state level.

The application demonstrates how the KDE algorithm enables the estimation of poverty and inequality indicators from grouped data. The precise estimates obtained by the KDE algorithm enable statisticians and statistical offices to report a variety of poverty and



10%−Quantile

(a) Regional distribution of the 10% quantile.

25%−Quantile

(b) Regional distribution of the 25% quantile.

Median

(c) Regional distribution of the median.

75%−Quantile

(d) Regional distribution of the 75% quantile.

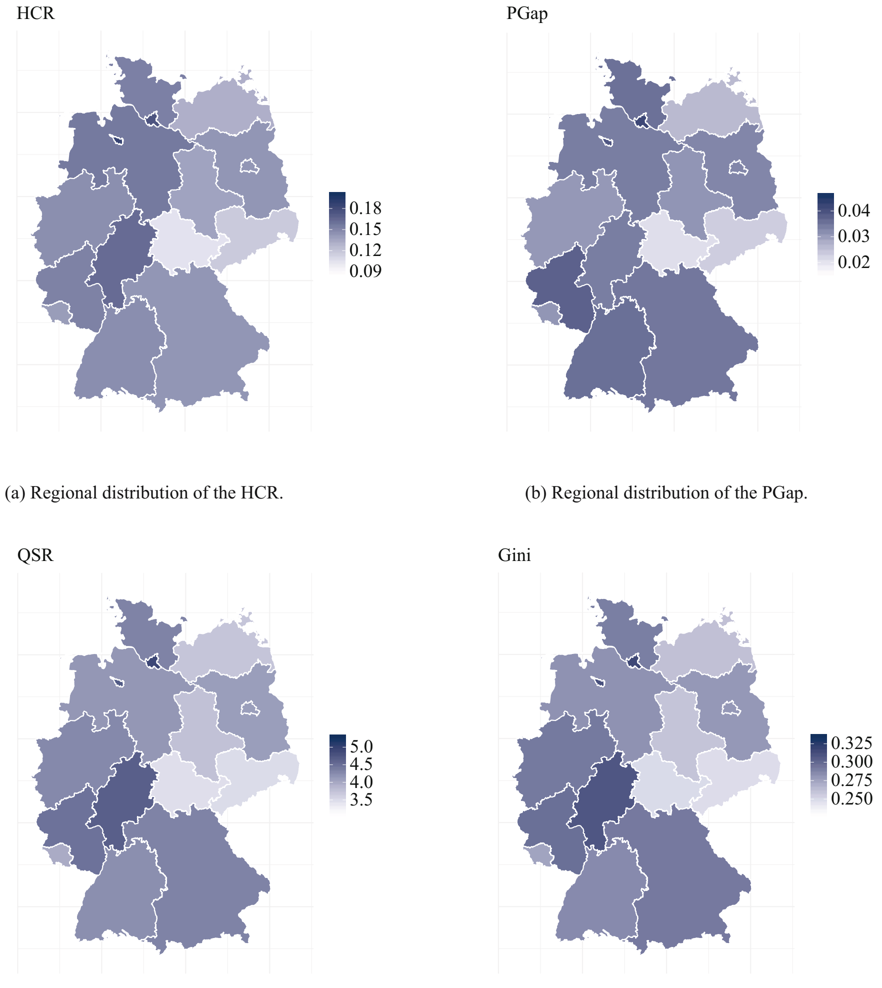
90%−Quantile

(e) Regional distribution of the 90% quantile

Mean

(f) Regional distribution of the mean.

Fig. 3. Regional distribution of different statistical indicators in Germany.



(a) Regional distribution of the HCR.

(b) Regional distribution of the PGap.

(c) Regional distribution of the QSR.

(d) Regional distribution of the Gini coefficient.

Fig. 4. Regional distribution of different statistical indicators in Germany.

inequality indicators using the most valuable data source in Germany, the German Microcensus. The regional estimates will help to identify regions with lower income and higher inequality to target political activities more accurately for those in need.

5. Simulation Results

This section presents model-based simulation results in order to evaluate the performance of the KDE algorithm in the context of estimating poverty and inequality from grouped income data. The code of the simulation study, as well as the simulated data used in the simulation study are available as online supplementary material. Due to the high computational complexity of the simulation study the code needs to run in a parallel computing environment, for example a high performance cluster. Additionally, the KDE algorithm is also included in the open source R package smicd (Walter 2021). This enables practitioners to easily apply the KDE algorithm to the estimation of poverty and inequality indicators on a local R environment.

The simulation study is set up with the following specifications. From a theoretical distribution $M = 500$ samples of simulated monthly income data are generated. The samples are grouped. The sample size for each sample is $n = 10{,}000$. The KDE algorithm is evaluated for large samples because grouped income data is common for surveys (like the German Microcensus) and for censuses which, in general, have very large sample sizes. From the simulated grouped income data different poverty and inequality indicators are estimated: the mean, the quantiles $(10\%, 25\%, 50\%, 75\%, 90\%)$, the HCR, the PGap, the Gini coefficient and the QSR. The formulas are given in Equations (2) – (7). In the simulation study sampling weights are not included, because the scope of the study is to evaluate the performance of the KDE algorithm. Therefore, $w_i = 1$ for $i = 1, \ldots, 10,000$ in the simulation study.

The indicators are estimated by the proposed KDE algorithm. The number of burn-in iterations of the algorithm is set to $B_{(KDE)} = 80$, the number of additional iterations $S_{(KDE)} = 400$. Our experiences running several simulations show that 480 iterations are usually enough to ensure convergence. Nevertheless, we check the convergence plots from randomly chosen simulation runs to assure that the indicators in the presented simulations converge. The number of grid points is set to $j = 4,000$. In general, a higher number of grid points leads to more precise estimation results, because the number of grid points determines how many unique values the pseudo samples of the grouped variable can consist of. However, the estimation time increases with the increasing number of grid points. The presented poverty and inequality indicators are not only estimated by the KDE algorithm (KDE). For comparison, the indicators are also estimated by linear interpolation. This method is used by the Federal Statistical Office of Germany for the estimation of poverty and inequality indicators from the grouped income variable of the German Microcensus (Information und Technik (NRW) 2009). This approach gives the same results as assuming a uniform distribution within the income classes (Uni). Furthermore, the statistical indicators are estimated by using the midpoints (Mid) of the groups as a proxy for the unobserved values within the income group. The estimated indicators of the continuous un-grouped data (True) are treated as a benchmark because they are not affected by the censoring.

Furthermore, we included a parametric estimation approach based on the GB2 distribution (Para). This methods works by replacing the kernel density estimate in step 1 and 5 of the algorithm by a parametric ML- estimate using the pseudo samples assuming a GB2 distribution. We expect the method to work well for our simulation scenarios, because we use the GB2 distribution and special cases of it for generating the data in the simulation study. The results of the parametric estimator are discussed in Subsection 5.4.

The results of a generic estimator $\hat{I}$, for example the KDE alg. or Uni, for a generic indicator I, for example the HCR and the Gini, are evaluated by the relative bias (rB),

$$rB(\hat{I}) = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{\hat{I}_m - I}{I}\right) \times 100,$$

and the empirical standard errors (se.emp),

$$se.emp(\hat{I}) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{I}_m - \bar{I}\right)^2},$$

with

$$\bar{I} = \frac{1}{M}\sum_{m=1}^{M}\hat{I}_m,$$

where the true value I is calculated from the theoretical distribution. The proposed non-parametric bootstrap for the estimation of the standard errors is evaluated by comparing the average over the 500 simulation runs of the estimated standard errors to the average over the 500 simulation runs of the empirical standard errors. The bootstrap is run with $B = 100$. This number shows it is sufficient to obtain valid approximations of the standard errors.

As mentioned before, the KDE algorithm cannot be applied to open-ended groups. We therefore replace the upper open-ended group bound by a value of three times its lower bound, as described in Section 4. The replacement value used for open-ended upper group also has an impact on the performance of the methods Uni and Mid. For these two methods we also replace the open-ended upper bound by three times its lower bound. We expect the impact of the replacement of the upper bound for the method Uni and Mid to be more sever then for the KDE algorithm. Adding distributional assumptions for the open-ended interval could improve these traditional methods. For the KDE algorithm the above described replacement rule gives very good results in our simulation study.

The simulation study is divided into four subsections. In Subsection 5.1, the influence of different numbers of groups on the performance of the KDE algorithm is evaluated. In Subsection 5.2, different underlying distributions are evaluated and, in Subsection 5.3, the effect of equal versus ascending group width is studied. In Subsection 5.4 the results of the KDE algorithm are compared with a parametric approach and in Subsection 5.5 the main results are summarized.

5.1. Different Grouping Scenarios

In this section, the influence of the number of groups on the performance of the KDE algorithm is studied. As theoretical distribution the four-parameter GB2 distribution that is often used to model income is used. Its parameters are specified such that the GB2 distribution well approximates the empirical German income distribution (Graf and Nedyalkova 2014). The chosen parameters are given in Table 5. The drawn samples are grouped using three different censoring scenarios. In Scenario 1, the data is divided into 24 groups as in the German Microcensus (Statistisches Bundesamt 2017) that is used in the application in Section 4. The group widths are chosen such that the frequencies within each group of the theoretical distribution match the frequencies within each group in the empirical distribution of the grouped household income in the German Microcensus. This is visualized in Figure 5 in the left panel. The middle and the right panel show the GB2 distribution divided into 16 groups (Scenario 2) and eight groups (Scenario 3). The performance of the algorithm with the lower number of classes is studied because surveys and censuses from other countries censor the

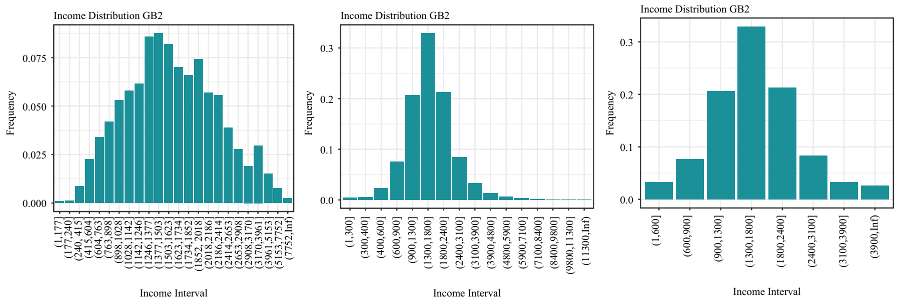


Fig. 5. GB2 distribution divided into 24 (left), 16 (middle) and eight groups (right).

income variable to fewer than 24 groups. For example, in the census from New Zealand the income variable is divided into 16 groups (Statistics New Zealand 2013), in the Australian census the data is divided into 12 groups (Australian Bureau of Statistics 2011), and in the Colombian census the income variable is divided into only nine groups (Departamento Administrativo Nacional De Estadlstica 2005).

The results of the point estimates are given in Table 3. Using the continuous un-grouped data for the estimation of the poverty and inequality indicators leads to unbiased results. This is not surprising as the sample size ($n = 10{,}000$) is very large. Using only the group information, the KDE algorithm outperforms the other approaches (Mid and Uni) in all three scenarios. The out-performance is especially stronger for indicators that rely on the whole shape of the distribution (Gini coefficient, mean), for the more extreme quantiles (10% quantile and 90 % quantile), and for indicators that rely on more extreme quantiles (QSR). As the number of groups decreases, the performance of the KDE algorithm worsens. Nevertheless, the bias is still under 1% for all indicators, except for the QSR, PGap and the Gini coefficient. The QSR shows a bias of -1.1%, the PGap a bias of 2.3%, and the Gini coefficient shows a bias of -1.9% in the eight-group scenario. The estimated indicators using the other approaches (Mid and Uni) exhibit far larger biases as the number of groups decreases. For example, in the eight-group scenario the PGap has a bias of 22% and 20% and the Gini coefficient of 14% and 24% for the estimation approaches Uni and Mid, respectively.

The precision of the KDE algorithm, measured by the empirical standard error (se.emp), is for all three scenarios close to the estimation results using the un-grouped data. This is the case because the estimated indicators rely on the metric pseudo samples from the KDE algorithm. However, the pseudo samples can – in rare circumstances – include very extreme values that lead to a higher variance when statistical indicators are estimated that rely on the whole distribution. This is, for example, the case for the mean in the 24-group scenario. The KDE algorithm almost loses no precision for a lower number of groups. The methods Uni and Mid lead to less precise estimation results, especially with fewer groups. For some of the estimated quantiles the empirical standard error of the Mid approach is 0. This is due to the fact that the Mid approach estimates the indicators on the midpoints of the groups. This leads to only 24, 16 or eight unique values, respectively. With a sample size of ($n = 10{,}000$) the estimated quantiles are likely to fall on the same midpoint for each of the 500 Monte Carlo iterations.

Table 3. Relative bias (rB) and the empirical standard error (se.emp) for the different estimation methods estimated for a selection of statistical indicators.

		$Q_{0.1}$	$Q_{0.25}$	Median	$Q_{0.75}$	$Q_{0.9}$	Mean	HCR	QSR	PGap	Gini
Measure	Method	GB2: 24 groups									
rB	True	0.053	0.036	0.008	-0.003	0.017	0.023	-0.087	-0.005	-0.163	-0.005
	KDE	-0.102	-0.059	-0.033	-0.045	0.121	0.002	-0.141	0.720	0.181	-0.036
	Uni	-0.366	-0.086	0.065	0.080	0.171	1.104	1.087	3.751	2.628	3.374
	Mid	-4.654	0.003	-0.313	1.501	1.848	2.218	-11.962	35.517	1.529	6.161
	Para	-0.078	-0.038	-0.03	-0.046	0.109	0.052	-0.128	0.531	0.013	-0.203
se.emp	True	87.600	72.172	71.259	109.180	222.019	95.973	0.00322	0.0492	0.00105	0.00280
	KDE	84.944	68.284	69.756	112.048	227.883	121.231	0.00316	0.0668	0.00107	0.00415
	Uni	96.181	69.987	70.633	119.183	240.357	111.912	0.00317	0.0596	0.00109	0.00347
	Mid	83.717	0.000	0.000	738.583	1092.148	137.517	0.00306	0.3512	0.00107	0.00463
	Para	83.555	67.637	69.221	111.531	222.439	93.430	0.00301	0.0483	0.00107	0.00265
		GB2: 16 groups									
rB	True	-0.007	0.012	0.022	0.021	0.014	-0.020	-0.030	-0.077	0.109	-0.102
	KDE	0.323	-0.021	0.260	0.190	-0.051	-0.018	0.478	0.699	0.034	-0.401
	Uni	-0.991	-1.832	0.823	3.492	3.543	1.154	4.522	5.113	7.699	3.691
	Mid	-14.210	-8.097	-1.200	3.499	3.098	1.536	-12.619	92.185	6.194	0.835
	Para	0.326	0.107	0.132	0.052	0.037	0.034	0.324	0.764	-0.425	-0.264
se.emp	True	90.029	72.505	78.428	113.178	232.863	101.242	0.00337	0.0484	0.00111	0.00272
	KDE	88.476	72.731	73.944	119.657	229.199	101.652	0.00327	0.0489	0.00110	0.00268
	Uni	120.142	84.036	81.005	131.425	248.381	110.794	0.00336	0.0552	0.00115	0.00307
	Mid	221.137	0.000	0.000	0.000	0.000	121.311	0.00311	0.3210	0.00113	0.00384
	Para	87.649	71.521	77.137	113.925	215.498	102.116	0.00322	0.0490	0.00112	0.0027

Table 3. Continued

		GB2: 8 groups									
rB	True	0.076	0.006	-0.016	0.021	0.017	-0.006	-0.103	-0.051	-0.131	-0.037
	KDE	0.106	-0.173	0.252	0.145	-0.141	-0.685	0.119	-1.151	2.329	-1.871
	Uni	-0.980	-1.850	0.820	3.519	3.587	4.190	4.323	17.586	21.758	13.522
	Mid	-13.972	-8.012	-1.155	3.582	3.092	10.187	-12.555	164.261	20.273	24.256
	Para	0.087	-0.022	0.08	0.048	-0.057	-0.358	-0.188	-0.908	-0.61	-1.25
se.emp	True	92.276	75.720	71.976	111.044	240.443	100.286	0.00346	0.0505	0.00109	0.00288
	KDE	88.373	74.822	70.126	113.115	231.700	126.809	0.00338	0.0714	0.00119	0.00443
	Uni	120.998	86.888	73.876	128.360	253.586	132.150	0.00346	0.0752	0.00135	0.00374
	Mid	220.916	0.000	0.000	0.000	0.000	183.278	0.00321	0.4810	0.00131	0.00511
	Para	87.805	72.731	72.891	109.482	217.465	96.147	0.00330	0.0479	0.00116	0.00255

In Table 4, the proposed bootstrap for the estimation of the standard errors is evaluated for the three different censoring scenarios. The standard errors estimated by the non-parametric bootstrap (se.est) offer a good approximation of the empirical standard errors (se.emp). This underlines the reliability of the proposed bootstrap method.

5.2. Different Theoretical Distributions

While the previous section evaluates the performance of the KDE algorithm using different censoring schemes, this section focuses on the evaluation of the performance using different theoretical distributions. According to several authors, among others McDonald (1984), McDonald and Xu (1995) and Bandourian et al. (2002), the GB2 distribution is well-suited for modelling income and it is superior to other parametric distributions. Nevertheless, two special cases of the GB2 distribution are used – in addition to the GB2 scenario discussed in Subsection 5.1 – for evaluations in order to illustrate the flexibility of the KDE algorithm: the Dagum (Dagum 1977) distribution and the Singh-Maddala (Singh and Maddala 1976) distribution. The choice of parameters follows Bandourian et al. (2002) (see Table 5) in order to closely approximate empirical income distributions. In Bandourian et al. (2002) it is shown that the Dagum and Singh-Maddala distribution specified with the parameters given in Table 5 have approximated the German income distribution well in the past. The data is divided into eight groups and the group width is chosen such that the relative frequency within each group is similar to the eight-group GB2 scenario from the previous section (see Figure 5 and 6). The eight-group scenario is chosen to evaluate the KDE algorithm under extreme scenarios. By keeping the relative frequencies equal within each group the effect of different distributions (GB2, Dagum, and Singh-Maddala) on the estimation results is isolatedly evaluated.

The estimation results of the point estimates are given in Table 6 (Dagum and Singh-Maddala) and Table 3 (GB2). As expected, using the un-grouped data (True) leads to unbiased estimation results. Also, the KDE algorithm that only uses the group information yields unbiased results for all indicators under the different scenarios. Hence, the performance of the KDE algorithm is not impaired by the underlying theoretical distribution. The benchmark methods (Uni and Mid) give biased estimation results, especially for indicators that depend on the whole distribution. For example, the QSR has a bias of 16.5% (Uni) and 210% (Mid) for the Dagum scenario and 18.5% (Uni) and 200% (Mid) for the Singh-Maddala scenario. Regarding the precision, the conclusions from the previous section are transferable.

As given in Table 7, the estimated standard errors offer a good approximation of the empirical standard errors for the different scenarios.

5.3. Equal and Ascending Group Width

While the German Microcensus (Statistisches Bundesamt 2017), the Australian (Australian Bureau of Statistics 2011), the Colombian (Departamento Administrativo Nacional De Estadística 2005), and the census from New Zealand (Statistics New Zealand 2013) use ascending class width, previous research shows that the performance of alternative estimation methods depends on the group width (Lenau and Münnich 2016).

Table 4. Empirical and estimated standard error for the selected statistical indicators.

		$Q_{0.1}$	$Q_{0.25}$	Median	$Q_{0.75}$	$Q_{0.9}$	Mean	HCR	QSR	PGap	Gini
Measure	Method					GB2: 24 groups					
se.emp	KDE	84.944	68.284	69.756	112.048	227.883	121.231	0.00316	0.0668	0.00107	0.00415
se.est		84.945	71.525	72.437	110.804	234.200	120.855	0.00318	0.0675	0.00107	0.00419
						GB2: 16 groups					
se.emp	KDE	88.476	72.731	73.944	119.657	229.199	101.652	0.00327	0.0489	0.0011	0.00268
se.est		87.972	70.564	68.708	110.969	224.122	96.000	0.00323	0.0503	0.0011	0.00278
						GB2: 8 groups					
se.emp	KDE	88.373	74.822	70.126	113.115	231.700	126.809	0.00338	0.0714	0.00119	0.00443
se.est		85.036	71.131	68.217	109.751	229.160	132.415	0.00323	0.0762	0.00117	0.0048

Table 5. Specified parameters for the theoretical distributions.

Distribution	Parameter			
	a	b	p	q
GB2	7.481	16351	0.4	0.468
Dagum	4.413	94075	0.337	–
Singh-Maddala	1.771	500000	25.12	–

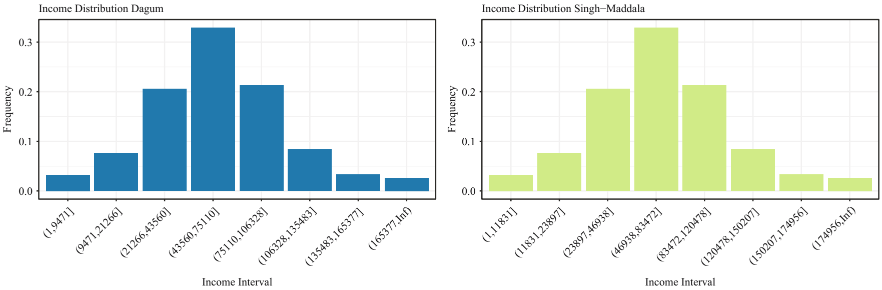


Fig. 6. Dagum and Singh-Maddala distribution divided into eight groups.

More precisely, performance depends on whether the data is grouped to equal width or ascending width. Therefore, the GB2 distribution from Table 5 is now grouped to eight groups with equal group width (except the last group, which has an open-ended upper group bound). In all previous simulation scenarios ascending group width is used. Figure 7 shows the grouped GB2 distribution. The theoretical distribution is kept fixed in order to evaluate the influence of the censoring on the performance.

The results of the point estimates are given in Table 8. As before, using the un-grouped data leads to unbiased estimates. The estimates obtained by the KDE algorithm are unbiased except for the QSR, PGap, and Gini coefficient. These estimates exhibit a small

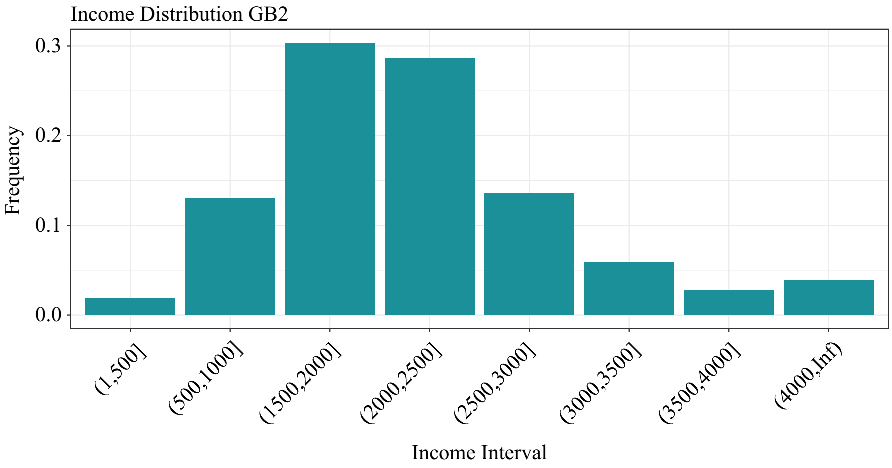


Fig. 7. GB2 distribution grouped to equally sized groups (except the last – open-ended – group).

Table 6. Relative bias (rB) and the empirical standard error (se.emp) for the different estimation methods estimated for a selection of statistical indicators.

		$Q_{0.1}$	$Q_{0.25}$	Median	$Q_{0.75}$	$Q_{0.9}$	Mean	HCR	QSR	PGap	Gini
Measure	Method					Dagum: 8 groups					
rB	True	0.041	-0.014	0.020	0.003	0.005	0.015	0.032	0.072	0.036	0.028
	KDE	0.192	0.088	-0.146	0.225	0.038	-0.396	-0.126	-0.770	-0.084	-0.851
	Uni	-0.977	-1.719	0.675	3.150	2.883	5.454	2.579	16.532	4.163	9.840
	Mid	-23.304	-12.787	-2.552	3.227	2.420	12.042	29.23	209.641	-2.171	16.251
	Para	0.169	0.061	-0.229	0.075	0.127	-0.243	-0.08	-0.42	-0.222	-0.486
se.emp	True	399.449	437.44	455.249	584.052	988.153	442.182	0.00433	0.1276	0.0022	0.0028
	KDE	382.632	422.677	440.771	567.565	964.208	479.943	0.00412	0.1349	0.00222	0.00315
	Uni	459.406	461.163	456.904	645.016	1052.903	613.491	0.00422	0.1706	0.00216	0.00374
	Mid	0	0	0	0	0	826.842	0.00457	1.0238	0.00208	0.00506
	Para	380.622	404.966	440.349	555.477	936.435	451.409	0.00394	0.1268	0.00218	0.00269
						Singh-Maddala: 8 groups					
rB	True	-0.070	0.001	0.035	0.014	-0.015	0.003	0.023	0.017	0.041	-0.006
	KDE	0.270	0.014	0.042	-0.039	-0.031	0.093	-0.039	0.714	0.085	0.213
	Uni	-1.031	-1.21	1.652	2.963	2.039	6.269	1.800	18.504	4.321	11.024
	Mid	-21.083	-11.797	-1.789	3.039	1.636	12.618	27.516	199.584	-1.651	17.009
	Para	0.117	0.251	0.207	-0.353	-0.201	0.318	-0.406	1.123	-0.21	0.537
se.emp	True	416.957	486.609	555.653	731.369	1049.186	443.818	0.00449	0.0994	0.00205	0.00213
	KDE	389.926	447.684	546.007	698.835	998.289	462.384	0.00422	0.1056	0.00211	0.00239
	Uni	467.696	502.097	547.127	784.601	1072.791	598.248	0.00444	0.1451	0.00209	0.00334
	Mid	784.213	0	0	0	0	784.707	0.00481	0.9298	0.00199	0.00443
	Para	397.6629	429.378	537.419	673.410	935.624	460.042	0.00405	0.1038	0.00207	0.00233

Table 7. Empirical and estimated standard error for the selected statistical indicators.

		$Q_{0.1}$	$Q_{0.25}$	Median	$Q_{0.75}$	$Q_{0.9}$	Mean	HCR	QSR	PGap	Gini
Measure	Method					Dagum: 8 groups					
se.emp	KDE	382.632	422.677	440.771	567.565	964.208	479.943	0.00412	0.135	0.00222	0.00315
se.est		385.340	420.523	445.765	573.573	953.225	468.896	0.00409	0.134	0.00221	0.00317
						Singh-Maddala: 8 groups					
se.emp	KDE	389.926	447.684	546.007	698.835	998.289	462.384	0.00422	0.106	0.00211	0.00239
se.est		386.539	430.594	523.137	691.090	983.671	460.726	0.00405	0.110	0.00211	0.00245

Table 8. *Relative bias (rB) and the empirical standard error (se.emp) for the different estimation methods estimated for a selection of statistical indicators.*

		$Q_{0.1}$	$Q_{0.25}$	Median	$Q_{0.75}$	$Q_{0.9}$	Mean	HCR	QSR	PGap	Gini
Quality Measure	Estimation Method	GB2: 8 groups (equally sized)									
rB	True	0.079	0.035	0.013	-0.026	-0.092	-0.024	-0.127	-0.144	-0.248	-0.110
	KDE	-0.005	-0.422	0.238	-0.066	0.050	-0.840	0.290	-1.706	1.370	-2.181
	Uni	-7.074	-2.388	0.909	0.560	1.704	4.648	7.351	21.365	30.052	16.251
	Mid	-14.151	4.640	11.598	10.730	3.174	12.498	19.720	73.226	28.594	30.467
	Para	0.012	-0.091	0.072	-0.032	-0.013	-0.566	0.324	-1.332	-0.325	-1.66
se.emp	True	88.841	75.061	72.038	111.139	233.943	95.398	0.00334	0.0513	0.00107	0.00292
	KDE	86.255	70.621	76.142	109.506	223.898	128.012	0.00337	0.0756	0.00114	0.00484
	Uni	116.469	70.955	88.391	156.503	260.393	130.810	0.00349	0.0763	0.00130	0.00364
	Mid	0	0	0	544.426	0	180.793	0.00367	0.2806	0.00117	0.00481
	Para	81.864	74.955	72.680	111.250	213.929	89.753	0.00324	0.0470	0.00114	0.00254

Table 9. Empirical and estimated standard error for the selected statistical indicators.

		$Q_{0.1}$	$Q_{0.25}$	Median	$Q_{0.75}$	$Q_{0.9}$	Mean	HCR	QSR	PGap	Gini
Measure	Method	GB2: 8 groups (equally sized)									
se.emp	KDE	86.255	70.621	76.142	109.506	223.898	128.012	0.00337	0.0756	0.00114	0.00484
se.est		84.456	67.507	75.134	108.778	224.587	138.326	0.00326	0.0794	0.00114	0.00506

bias of -1.7%, 1.4% and -2.2%. However, the results are comparable to the estimation results from the GB2 scenario with eight groups with ascending group width. Hence, the KDE algorithm does not seem to be affected by the censoring scheme. The benchmark indicators Uni and Mid show, as before, large biases especially for indicators that rely on the whole shape of the distribution. With regard to precision, the results and interpretation are the same as before. The proposed bootstrap also gives valid results with equal-sized groups (see Table 9).

5.4. Parametric Comparison

Comparable to the KDE algorithm the parametric approach gives almost unbiased results for all of the investigated scenarios (Table 3, 6, 8). For the very extreme scenarios, with only eight groups, the parametric SEM-algorithm even slightly outperforms the non-parametric KDE algorithm. However, in the simulation study the data is generated based on a GB2 distribution and special cases of it. If the true data generating process is different from a GB2 distribution the parametric approach will give biased results. However, for the non-parametric KDE- algorithm we do not expect a sharp decrease in performance. Hence, due to its flexibility and unbiasedness the KDE algorithm remains preferable for applications.

5.5. Conclusion and Final Remarks

The simulation results show that the proposed KDE algorithm outperforms currently used approaches by statistical offices (Uni and Mid) in terms of bias in the investigated scenarios. The KDE algorithm gives almost unbiased results under different censoring schemes and for different underlying theoretical distributions. The relative bias increases slightly whenever the number of groups decreases. However, also in extreme censoring scenarios (with only eight groups), the results are precise. The relative bias is under 1% for almost all indicators. The KDE algorithm shows comparable results in terms of precision to the direct estimation of the indicators from the continuous un-grouped data. Additionally, it is superior to other commonly used approaches (Mid and Uni) that show worse precision for most indicators. Due to its easy usage, its ability to adapt to different underlying theoretical distributions and different censoring schemes and its precision practitioners should prefer the KDE algorithm to the other approaches (Uni and Mid).

6. Discussion, Limitations and Future Research

In numerous surveys and censuses for example, the German Microcensus or the Australian census, the variable household or personal income is not observed on a continuous scale, but is rather divided into specific groups. This is due to confidentiality constraints or to reduce item non-response. Estimating poverty and inequality indicators from these kinds of data requires more sophisticated statistical methods. As an estimation method we propose a new iterative KDE algorithm that enables the precise estimation of statistical indicators from grouped data. The proposed KDE algorithm has similarities to SEM algorithms that are commonly used for the estimation of models that depend on latent unobserved variables (in our case the grouped income). However, instead of maximizing the likelihood as is common for SEM algorithms, the asymptotic mean integrated squared error of the KDE is minimized. For the estimation of the standard errors of the statistical indicators a non-parametric bootstrap is proposed. The KDE algorithm and the bootstrap work for different censoring scenarios and different underlying true distributions. The methodology is available in the R package `smicd` from the Comprehensive R Archive Network (Walter 2021). Our simulation results demonstrate that the estimated poverty and inequality indicators outperform other estimation techniques (linear interpolation or the use of the midpoints of the groups) in terms of bias. Also, the standard errors of the estimates are close to the standard errors from the estimates that were obtained with the un-grouped data, supporting the precision of the algorithm. Furthermore, the KDE algorithm has the advantage of adapting to different grouped theoretical distributions. Therefore, it is applicable for the estimation of poverty and inequality indicators from grouped income data. We demonstrate the usefulness by estimating regional poverty and inequality indicators using German Microcensus data. To get representative results the estimation is carried out by taking the OECD scale and the survey weights of the Microcensus into account. The estimated regional indicators are plotted on maps that visualize the magnitude of poverty and inequality in Germany. With the help of the KDE algorithm statistical indicators can be precisely estimated from grouped data in order to tackle the increasing problem of rising poverty and inequality in societies all over the world.

In our article we did not focus on complex survey designs. As already mentioned in Subsection 3.3 the algorithm could be altered such that survey weights are including into the estimation of the density in each step of the KDE algorithm (Buskirk and Lohr 2005). For the considered examples, we could not find any meaningful differences compared to our proposed weighting method. However, for variance estimates this is a more serious issue. We think that more research is needed here, especially in context of complex survey designs. Some surveys also oversample certain parts of the population, for example, very wealthy people to assure that 'rare outcomes' are include in the survey. In particular, oversampling very wealthy people is done by the German Socio-Economic panel. Under these settings it could be beneficial to examine the effects of the weighting method. Deriving theoretical properties and generating empirical evidence for the KDE algorithm under complex survey designs can be part of future research. In this article we have empirically evaluated the naive bootstrap. However, theoretical deriving the efficiency of the bootstrap or considering more complex bootstrap methods for specific sampling designs could also be part of future work.

The algorithm can also be extended to situations in which every observation has its own unique group bounds. This can lead to overlapping groups and to gaps between different groups. However, this does not impede the use of the KDE algorithm, because a properly adjusted KDE algorithm simply draws from the unique groups. Also, situations in which only some observations are grouped and others are observed on a continuous scale can be handled by the proposed KDE algorithm. In this scenario, the KDE algorithm only draws pseudo samples for the grouped observations and the continuous observations stay constant (since they are known) during the iterations of the KDE algorithm. Further research will also focus on convergence criteria for the KDE algorithm that make the manual choice of the number of iteration obsolete and on the derivation of analytic standard error of the estimated statistical indicators.

7. Appendix

Table 10. Sample size for Germany and each of the 16 federal states.

State	Sample size	Number in map
Germany	454852	
Schleswig-Holstein	15302	1
Hamburg	8630	2
Lower Saxony	45828	3
Bremen	3356	4
North Rhine-Westphalia	90778	5
Hesse	35730	6
Rhineland-Palatinate	21229	7
Baden-Württemberg	58685	8
Bavaria	75244	9
Saarland	5688	10
Berlin	19311	11
Brandenburg	15400	12
Mecklenburg-Vorpommern	8706	13
Saxony	24609	14
Saxony-Anhalt	13495	15
Thuringia	12861	16

Table 11. Estimated statistical indicators for Germany and the 16 federal states. Standard errors are given in parentheses.

	Quant0.1	Quant0.25	Median	Quant0.75	Quant0.9	Mean	HCR	QSR	PGap	Gini
Germany	770.16	1040.23	1445.53	1998.96	2714.63	1675.88	0.15	4.31	0.03	0.29
	(0.00)	(3.28)	(4.93)	(2.59)	(3.69)	(1.85)	(0.00)	(0.01)	(0.00)	(0.00)
Schleswig-Holstein	794.21	1092.99	1512.96	2071.11	2743.79	1736.25	0.15	4.33	0.04	0.29
	(6.04)	(5.83)	(7.39)	(7.8)	(20.01)	(9.35)	(0.00)	(0.05)	(0.00)	(0.00)
Hamburg	765.68	1069.24	1540.2	2166.83	3002.79	1815.45	0.17	4.92	0.04	0.32
	(7.16)	(9.21)	(10.47)	(13.44)	(29.75)	(14.14)	(0.00)	(0.09)	(0.00)	(0.00)
Lower Saxony	770.08	1040.25	1445.04	1970.84	2603.04	1636.36	0.16	4.16	0.03	0.28
	(4.1)	(2.53)	(5.44)	(7.11)	(13)	(5.79)	(0.00)	(0.04)	(0.00)	(0.00)
Bremen	665.44	876.91	1328.23	1879.66	2564.1	1540.03	0.18	4.72	0.04	0.31
	(10.82)	(9.93)	(16.76)	(25.28)	(47.95)	(19.19)	(0.01)	(0.12)	(0.00)	(0.01)
North Rhine-Westphalia	756.85	1013.41	1418.5	1985.64	2674.27	1649.22	0.15	4.29	0.03	0.29
	(0.72)	(0.01)	(3.01)	(4.7)	(9.05)	(3.69)	(0.00)	(0.02)	(0.00)	(0.00)
Hesse	798.23	1094.75	1540.36	2149.61	2997.03	1825.06	0.16	4.66	0.03	0.31
	(4.56)	(4.88)	(6.03)	(7.49)	(14.23)	(7.54)	(0.00)	(0.04)	(0.00)	(0.00)
Rhineland-Palatinate	770.65	1067.23	1485.95	2052.43	2810.09	1720.3	0.15	4.49	0.04	0.3
	(5.39)	(5.36)	(6.36)	(8.97)	(17.17)	(8.4)	(0.00)	(0.06)	(0.00)	(0.00)
Baden-Wütrtemberg	837.76	1148.33	1580.43	2160.84	2900.98	1806.4	0.15	4.24	0.04	0.29
	(2.23)	(4.1)	(4.08)	(6.62)	(10.5)	(5.52)	(0.00)	(0.03)	(0.00)	(0.00)
Bavaria	841.94	1148.28	1580.35	2147.52	2944.04	1826.62	0.14	4.33	0.03	0.29
	(5.37)	(4.62)	(4.99)	(5.19)	(8.81)	(4.75)	(0.00)	(0.03)	(0.00)	(0.00)
Saarland	784.7	1035.41	1434.2	1938.86	2559.91	1615.95	0.14	3.98	0.03	0.27
	(8.77)	(9.73)	(10.55)	(14.7)	(32.54)	(13.4)	(0.01)	(0.07)	(0.00)	(0.00)
Berlin	730.96	912.14	1328.5	1867.05	2552.45	1547.47	0.15	4.15	0.02	0.29
	(5.38)	(5.19)	(8.41)	(11.2)	(15.87)	(8.77)	(0.01)	(0.05)	(0.00)	(0.00)
Brandenburg	716.8	979.24	1351.02	1823.73	2446.72	1528.25	0.14	4.1	0.03	0.28
	(5.93)	(6.78)	(6.43)	(10.03)	(17.36)	(8.08)	(0.00)	(0.05)	(0.00)	(0.00)

Table 11. Continued

	Quant0.1	Quant0.25	Median	Quant0.75	Quant0.9	Mean	HCR	QSR	PGap	Gini
Mecklenburg-Vorpommern	671.3	895.52	1211.29	1629.61	2120.56	1355.61	0.13	3.74	0.03	0.26
	(4.92)	(5.77)	(7.51)	(9.78)	(20.77)	(11.96)	(0.00)	(0.08)	(0.00)	(0.01)
Saxony	709.57	945.88	1247.4	1622.64	2155.08	1383.2	0.12	3.52	0.02	0.25
	(4.62)	(4.00)	(4.31	(5.48	(10.93	(5.09	(0.00)	(0.03	(0.00)	(0.00)
Saxony-Anhalt	675.7	928.47	1247.05	1643.78	2161.33	1382.23	0.14	3.78	0.03	0.26
	(5.25)	(5.85)	(5.9)	(7.68)	(17.09)	(6.24)	(0.00)	(0.05)	(0.00)	(0.00)
Thuringia	755.17	973.23	1283.8	1683.44	2226.4	1435.52	0.11	3.5	0.02	0.25
	(5.32)	(4.23)	(5.50)	(7.11)	(16.00)	(7.45)	(0.00)	(0.04)	(0.00)	(0.00)

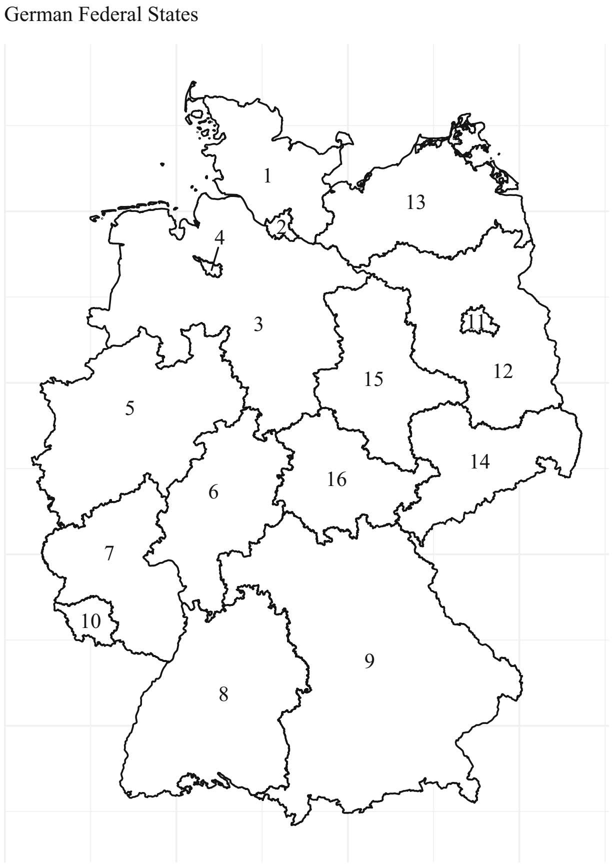


Fig. 8. German federal states, the names of the corresponding numbers are given in table 10.

8. References

Abraham, K., and S. Houseman. 1995. "Earnings inequality in Germany." *Differences and Changes in Wage Structures*, edited by R.B. Freeman and L.F. Katz: 371–404. Chicago: Nber Comparative Labor Markets.

Alfons, A., and M. Templ. 2013. "Estimation of social exclusion indicators from complex surveys: the R package laeken." *Journal of Statistical Software* 54(15): 1–25. DOI: https://doi.org/10.18637/jss.v054.i15.

Australian Bureau of Statistics. 2011. *Census household form*. DOI: https://unstats.un.org/unsd/demographic/sources/census/quest/AUS2 011en.pdf (accessed April 2018).

Bandourian, R., J. McDonald, and R.S. Turley. 2002. *A comparison of parametric models of income distribution across countries and over time*. Technical report, Luxembourg Income Study. Available at: http://www.lisdatacenter.org/wps/liswps/305.pdf.

Betensky, R.A., J. Lindsey, L. Ryan, and M. Wand. 1999. "Local EM estimation of the hazard function for interval-censored data." *Biometrics* 55: 238–245. DOI: https://doi.org/10.1111Zj.0006-341X.1999.00238.x.

Boehle, M. 2015. *Armutsmessung mit dem Mikrozensus: Methodische Aspekte und Umsetzung für Querschnitts- und Trendanalysen*. Technical report, Gesis Leibniz-Institut fur Sozialwissenschaften. Available at: https://www.ssoar.info/ssoar/handle/-document/45724.2.

Braun, J., T. Duchesne, and J. Stafford. 2005. "Local likelihood density estimation for interval censored data." *Canadian Journal of Statistics* 33: 39–60. DOI: https://doi.org/10.1002/cjs.5540 330104.

Buskirk, T. and S.L. Lohr. 2005. "Asymptotic properties of kernel density estimation with complex survey data." *Journal of Statistical Planning and Inference* 128: 165–190. DOI: https://doi.org/10.1016/j.jspi.2003.09.036.

Celeux, G., D. Chauveau, and J. Diebolt. 1996. "Stochastic versions of the EM algorithm: an experimental study in the mixture case." *Journal of Statistical Computation and Simulation* 55(4): 287–314. DOI: https://doi.org/10.1080/00949659608811772.

Celeux, G. and J. Diebolt. 1985. "The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem." *Computational Statistics Quarterly* 2: 73–82. Available at. https://www.researchgate.net/publication/229100768_The_SEM_.

Chen, Y.T. 2017. "A unified approach to estimating and testing income distributions with grouped data." *Journal of Business & Economic Statistics* 36(3): 1–18. DOI: https://doi.org/10.1080/07350015.2016.1194762.

Dagum, C. 1977. "A new model of personal income distribution: specification and estimation." *Economie Appliquee* 30: 413–437. Available at: https://ideas.repec.org/h/spr/esichp/978-0-387-72796-7_1.html.

Dempster, A., N. Laird, and D. Rubin. 1977. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society. Series B* 39(1): 1–38. DOI: https://doi.org/10.1111/j.2517-6161.197 7.tb01600.x.

Departamento Administrativo Nacional De Estadistica. 2005. *Censo general* 2005. DOI: https://www.dane.gov.co/files/censos/libroCenso2005nacional.pdf? (accessed April 2018).

Deville, J. 1999. "Variance estimation for complex statistics and estimators: linearization and residual techniques." *Survey Methodology* 25(2): 193-203.

Dorfman, A.H., and R. Valliant. 2005. "Superpopulation models in survey sampling." *Encyclopedia of Biostatistics* 8. DOI: https://doi.org/10.1002/0470011815.b2a16076.

Efron, B. 1979. "Bootstrap methods: another look at the jackknife." *The Annals of Statistics* 7(1): 1–26. DOI: https://doi.org/10.1214/aos/1176344552.

Eurostat. 2014. *Statistics explained: at-risk-of-poverty rate.* DOI: http://ec.europa.eu/eurosta/statistics-explained/index.php/Glossary:At-risk-of-poverty_rate. Accessed: 2018-05-30.

Eurostat. 2018. *At-risk-of-poverty rate by poverty therreshold.* DOI: http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do. Accessed: 2018-12-30.

Field, C.A., and A.H. Welsh. 2007. "Bootstrapping clustered data." *Journal of the Royal Statistical Society:* 69(3): 369–390. DOI: https://doi.org/10.1111/j.1467-9868.2007.00593.x.

Foster, J., J. Greer, and E. Thorbecke .1984. A class of decomposable poverty measures. *Econometrica* 52(3): 761–766. DOI: https://doi.org/10.2307/1913475.

Fréchet, M. 1927. “Sur la loi de probabilité de l’écart maximum.” *Annales de la Societe Polonaise de Mathe-matique* 6: 92–116.

Fuchs-Schündeln, N., D. Krueger, and M. Sommer. 2010. “Inequality trends for Germany in the last two decades: a tale of two countries.” *Review of Economic Dynamics* 13(1): 103–132. DOI: https://doi.org/10.1016/j.red.2009.09.004.

Gini, C. 1912. *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*. Studi economico-giuridici pubblicati per cura della facoltà di Giurisprudenza della R. Università di Cagliari. Bologna: Tipogr. di P. Cuppini.

Graf, M., and D. Nedyalkova. 2014. “Modeling of income and indicators of poverty and social exclusion using the generalized beta distribution of the second kind.” *Review of Income and Wealth* 60(4): 821–842. DOI: https://doi.org/10.1111/roiw.12031.

Groß, M., and U. Rendtel. 2016. “Kernel density estimation for heaped data.” *Journal of Survey Statistics and Methodology* 4(3): 339–361. DOI: https://doi.org/10.1093/jssam/smw011.

Groß, M., U. Rendtel, T. Schmid, S. Schmon, and N. Tzavidis. 2017. “Estimating the density of ethnic minorities and aged people in Berlin: multivariate kernel density estimation applied to sensitive georeferenced administrative data protected via measurement error.” *Journal of the Royal Statistical Society* 180(1): 161–183. DOI: https://doi.org/10.1111/rssa.12179.

Hagenaars, A., and K.D. Vos. 1988. “The definition and measurement of poverty.” *Journal of Human Resources* 23(2): 211–221. DOI: https://doi.org/10.2307/145776.

Hall, P. 1982. “The influence of rounding errors on some nonparametric estimators of a density and its derivatives.” *SIAM Journal on Applied Mathematics* 42(2): 390–399. DOI: https://doi.org/10.1137/0142030.

Hall, P., and M.P. Wand. 1996. “On the accuracy of binned kernel density estimators.” *Journal of Multivariate Analysis* 56(2): 165–184. DOI: https://doi.org/10.1006/jmva.1996.0009.

Henderson, D.J. and C.F. Parmeter. 2015. *Applied Nonparametric Econometrics*. New York: Cambridge University Press.

Information und Technik (NRW). 2009. *Berechnung von Armutsgefährdungsquoten auf Basis des Mikrozensus* DOI: http://www.amtliche-sozialberichterstattung.de/pdf/Berechnung%20von%20Armutsgefaehrd ungsquoten_090518.pdf. (accessed April 2018).

Jones, M.C., J.S. Marron, and S.J. Sheather. 1996. “A brief survey of bandwidth selection for density estimation.” *Journal of the American Statistical Association* 91(433): 401–407. DOI: https://doi.org/10.1080/01621459.1996.10476701.

Kakwani, N.C., and N. Podder. 2008. “Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations Lorenz curve and associated inequality measures from grouped observations.” In *Modeling Income Distributions and Lorenz Curves*, edited by D. Chotikapanich: 57–70. New York: Springer.

Kleiber, C. 2008. “A guide to the Dagum distributions Lorenz curve and associated inequality measures from grouped observations. In *Modelig Income Distributions and Lorenz Curves*, edited by D. Chotikapanich: 97–117. New York: Springer.

Lenau, S., and R. Münnich. 2016. *Estimating income poverty and inequality from income classes*. Technical report, InGRID Integrating Expertise in Inclusive Growth: Case Studies.

Li, L., T. Watkins, and Q. Yu. 1997. "An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data." *Scandinavian Journal of Statistics* 24: 531–542. DOI: https://doi.org/10.1111/1467-9469.0007 9.

Loader, C.R. 1999. "Bandwidth selection: classical or plug-in?" *Annals of Statistics* 27(2): 415–438. DOI: https://doi.org/10.1214/aos/1018031201.

Lok-Dessallien, R. 1999. *Review of poverty concepts and indicators*. Technical report, United Nations Development Programme. Available at: http://mirror.unpad.ac.id/orar-i/library/library-ref-ind/ref-ind-1/application/poverty-.

Mashreghi, Z., D. Haziza, and C. Leger. 2016. "A survey of bootstrap methods in finite population sampling." *Statistics Surveys* 10: 1–52. DOI: https://doi.org/10.1214/16-SS113.

McDonald, J.B. 1984. "Some generalized functions for the size distribution of income." *Econometrica* 52(3): 647–663. DOI: https://doi.org/10.2307/1913469.

McDonald, J.B., and Y.J. Xu. 1995. "A generalization of the beta distribution with applications." *Journal of Econometrics* 66(1): 133–152. DOI: https://doi.org/10.1016/0304-4076(94)01612-4.

McLachlan, G., and T. Krishnan. 2008. *The EM Algorithm and Extensions*. New York: Wiley.

Micklewright, J., and S. Schnepf. 2010. "How reliable are income data collected with a single question?" *Journal of the Royal Statistical Society:* 173(2): 409–429. DOI: https://doi.org/10.1111/j.1467-98 5X.2009.00632.x.

Moore, J.C., and E.J. Welniak. 2000. "Income Measurement Error in Surveys: a Review." *Journal of Official Statistics* 16(4): 331. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/income-measurement-error-in-surveys-a-review.pdf (accessed March 2022).

Nielsen, S.F. 2000. "The stochastic EM algorithm: estimation and asymptotic results." *Bernoulli* 6(3): 457–489. DOI: https://doi.org/10.2307/3318 671.

OECD. 2018. *Oecd data, income inequality*. Available at: DOI: https://data.oecd.org/i-nequality/income-inequality.htm (accessed December 2018).

Osier, G. 2009. "Variance estimation for complex indicators of poverty and inequality using linearization techniques." *Survey Research Methods* 3(3): 167–195. DOI: https://doi.org/10.18148/srm/2009.v3i3.369.

Pan, W. 2000. "Smooth estimation of the survival function for interval censored data." *Statistics in Medicine* 19: 2611–2624. DOI: https://doi.org/10.1002/1097-0258 (20001015)19:19 < 2 611:aid-sim538 > 3.0.co;2-o.

Parzen, E. 1962. "On estimation of a probability density function and mode." *The Annals of Mathematical Statistics* 33(3): 1065–1076. DOI: https://doi.org/10.1214/aoms/1177704472.

Pfeffermann, D., A.M. Krieger, and Y. Rinott. 1998. "Parametric distributions of complex survey data under informative probability sampling." *Statistica Sinica* 8(4): 1087–1114. Available at: https://pluto.huji.ac.il/~rinott/publications/PfKRR.pdf.

Pfeffermann, D., and M. Sverchkov. 1999. "Parametric and semi-parametric estimation of regression models fitted to survey data." *Sankhya: The Indian Journal of Statistics*, 61(1): 166–186. Available at: https://www.jstor.org/stable/25053074.

Reed, W.J., and F. Wu. 2008. "New four- and five-parameter models for income distributions." In *Modeling Income Distributions and Lorenz Curves*, edited by D. Chotikapanich: 211–224. New York: Springer.

Rosenblatt, M. 1956. "Remarks on some nonparametric estimates of a density function." *The Annals of Mathematical Statistics* 27(3): 832–837. DOI: https://doi.org/10.1214/aoms/1177728190.

Schimpl-Neimanns, B. 2010. *Varianzschaetzung fuer Mikrozensus Scientific Use Files ab 2005*, GESIS-Technical Reports 3. Mannheim: GESIS-Leibniz-Institut fuer Sozialwissenschaften. Available at: https://pluto.huji.ac.il/~rinott/publications/PfKRR.pdf https://www.jstor.org/stable/25053074.

Schwarz, N. 2001. "The German Microcensus." *Schmollers Jahrbuch* 132(1): 1–26. DOI: https://doi.org/10.3790/schm.132.1.1.

Scott, D.W., and S.J. Sheather. 1985. "Kernel density estimation with binned data." *Communications in Statistics – Theory and Methods* 14(6): 1353–1359. DOI: https://doi.org/10.1080/0361092 8508 828 980.

Shao, J., and D. Tu. 1995. *The Jackknife and Bootstrap*. New York: Springer.

Singh, S., and G. Maddala. 1976. "A function for the size distribution of incomes." *Econometrica* 44(5): 963–970. DOI: https://doi.org/10.2307/1911538.

Stacy, E. 1962. "A generalization of the gamma distribution." *The Annals of Mathematical Statistics* 33: 1187–1192. DOI: https://doi.org/10.1214/aoms/1177704481.

Statistical Offices of the Federation and the Federal States. 2016. *Data supply: Microcensus*. Available at: http://www.forschungsdatenzentrum.de/en/database/microcensus/index.asp. (accessed June 2018).

Statistics New Zealand. 2013. *New Zealand census of population and dwellings*. Available at: DOI: https://unstats.un.org/unsd/demographic/sources/census/quest/NZL2 013enIn.pdf (accessed May 2018).

Statistisches Bundesamt. 2017. *Datenhandbuch zum Mikrozensus Scientific-Use-File 2012*. Available at: http://www.forschungsdatenzentrum.de/bestand/mikrozensus/suf/2012/fdz_mz_suf_2012_schluesselverzeichnis.pdf. (accessed: July 2017).

Statistisches Bundesamt. 2018a. *Der Mikrozensus stellt sich vor*. Available at: DOI: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Mikrozensus.html. (accessed September 2018).

Statistisches Bundesamt. 2018b. *Microcensus*. Available at: DOI: https://www.destatis.de/EN/FactsFigures/SocietyState/Population/HouseholdsFamilies/Methods/Microcensus.html. (accessed June 2018).

Stauder, J., and W. Hüning. 2004. *Die Messung von Äquivalenzeinkommen und Armutsquoten auf der Basis des Mikrozensus*. Technical report, Statistische Analysen und Studien NRW. Available at: https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche.

Tepping, B. 1968. "Variance estimation in complex surveys." Proceedings of the American Statistical Association Social Statistics Section: 11–18. Available at:

http://www.asasrms.org/Proceedings/y1968/Variance%20Estimation%20In%20Complex%20Surveys.

Tille, Y. 2001. *Theorie des sondages: Echantillonnage et estimation en populations finies*. Paris: Dunod.

Walter, P. 2021. “The R package smicd: Statistical methods for interval- censored data”. *The R Journal* 13(1): 396–412. DOI: https://doi.org/10.32614/RJ-2 021-045.

Walter, P., M. Groß, T. Schmid, and N. Tzavidis. 2021. “Domain prediction with grouped income data.” *Journal of the Royal Statistical Society* 184(4): 1501–1523. DOI: https://doi.org/10.1111/rssa.12736.

Wang, B., and M. Wertelecki. 2013. “Density estimation for data with rounding errors.” *Computational Statistics & Data Analysis* 65: 4–12. DOI: https://doi.org/10.1016Zj.csda.2012.02.016.

Wolter, K. 1985. *Introduction to Variance Estimation*. New York: Springer.

Woodruff, R.S. 1971. “A simple method for approximating the variance of a complicated estimate.” *Journal of the American Statistical Association* 66(334): 411–414. DOI: https://doi.org/10.1080/01621459.1971.10482279.

World Economic Forum. 2017. *Global risks 2017*. Available at: http://reports.weforum.org/global-risks-2 017/part-1-global-risks-2017/ (accessed September 2017).

Received June 2020
Revised June 2021
Accepted February 2022