ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 051/26

Using distributional random forests for the analysis of the income distribution

Martin Biewen, Stefan Glaisner, Simon Zeller

www.rfberlin.com FEBRUARY 2026

Using distributional random forests for the analysis of the income distribution

Authors

Martin Biewen, Stefan Glaisner, Simon Zeller

Reference

JEL Codes: D31, I32

Keywords: inequality, poverty, small-area estimation, grouped income data

Recommended Citation: Martin Biewen, Stefan Glaisner, Simon Zeller (2026): Using distributional random forests for the analysis of the income distribution. RFBerlin Discussion Paper No. 051/26

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Working Paper, February 2026, 1–33
doi: n.a.

Using distributional random forests for the analysis of the income distribution

Martin Biewen[1,2,3*] **Stefan Glaisner**[1] **and Simon Zeller**[1]

[1]School of Business and Economics, University of Tübingen, Germany, [2]Institute for the Study of Labor (IZA), Bonn, Germany and [3]ROCKWOOL, Berlin, Germany

[*]Correspondence: School of Business and Economics, University of Tübingen, Mohlstr. 36, 72074 Tübingen, Germany, martin.biewen@uni-tuebingen.de

Abstract

This paper explores distributional random forests as a flexible machine learning method for analysing income distributions. Distributional random forests avoid parametric assumptions, capture complex interactions among covariates, and, once trained, provide full estimates of conditional income distributions. From these, any type of distributional index such as measures of location, inequality and poverty risk can be readily computed. They can also efficiently process grouped income data and be used as inputs for distributional decomposition methods. We consider four types of applications: (i) estimating income distributions for granular population subgroups, (ii) analysing distributional change over time, (iii) small-area estimation of income distributions, and (iv) purging spatial income distributions of differences in spatial characteristics. Our application based on the German Microcensus provides new results on the socio-economic and spatial structure of the German income distribution.

Key words: inequality, poverty, small-area estimation, grouped income data

1. Introduction

Measuring distributions of income and wealth is a central concern of both statistics and the social sciences. A large number of statistical techniques have been developed to estimate such distributions and to investigate their structure (e.g., Jenkins and Van Kerm, 2009; Fortin et al., 2011; Chernozhukov et al., 2013; Cowell and Flachaire, 2015; Chotikapanich et al., 2018; Molina et al., 2022). In many cases, these techniques involve strong parametric assumptions about distributional shapes or the structure of regression models that describe how distributions depend on covariates. A key advantage of modern machine learning methods such as random forests (Breiman, 2001; Athey et al., 2019) is their ability to avoid such assumptions. As a recursive partitioning algorithm, the random forest is based on sequentially splitting the covariate space into cells of observations that are similar with respect to a target criterion, and by aggregating independent repetitions of this procedure. It has a non-parametric structure, accomodates complex interactions and potentially non-smooth relationships, and implicitly addresses model-selection problems. Random forests have demonstrated remarkable success across a wide range of applications (Biau and Scornet, 2016).

Breiman (2001)'s original random forest was designed for non-parametric mean estimation. Subsequent extensions included survival analysis (Hothorn et al., 2004), conditional quantile estimation (Meinshausen, 2006), and estimators defined by local moment conditions (Athey et al., 2019). More recently, Cevid et al. (2022) and Näf et al. (2023) proposed a highly general variant of the random forest aimed at estimating full conditonal distributions (distributional random forest, DRF).

As pointed out by Cevid et al. (2022), building random forests for full distributions – rather than for individual target objects such as means, quantiles or other distributional indices – has a number of advantages. These advantages are particularly relevant to analyses of the income distribution. First, forest building needs to be carried out only once to obtain estimates for arbitrarily many targets. For example, if one is interested in distributional indices such as the median income, the at-risk-of-poverty rate, quantile ratios, or the Gini index for small population subgroups, one has to fit the random forest only once and then obtain estimates of these targets from the conditional distribution. Second, since the estimates for different targets are obtained from the same forest, they are mutually compatible. This is not necessarily the case if a new forest is fit for each target. For example, it is well known that conditional quantiles may cross if they are estimated separately. Similarly, fitting separate forests may produce values of the at-risk-of-poverty rate and the Gini coefficient for individual subgroups that are difficult to reconcile. Third, fitting separate forests for different target objects requires suitable target-specific splitting criteria. For many targets these are unknown or could be difficult to derive. By contrast, the DRF directly uses a powerful distributional criterion for splitting, the maximum mean discrepancy (MMD) statistic (Gretton et al., 2007).

As a statistical method, the distributional random forest follows the same estimation goal as a number of other estimators of conditional distributions. These alternatives typically have a parametric or semi-parametric structure, see, e.g., Donald et al. (2000), Biewen and Jenkins (2005), Rigby and Stasinopoulos (2005),

Hothorn et al. (2013). Conditional quantile models (Koenker, 2005) and binary models for distributional thresholds (Chernozhukov et al., 2013) can also be used to construct conditional distribution functions, but they require fitting a large number of quantiles or threshold models. However, in all of these models, it is not easy to accomodate issues such as non-smooth dependencies, complex interaction effects and automatic variable selection, features that are automatically handled by the random forest. Before the development of the fully non-parametric distributional random forest, Schlosser et al. (2019) and Hothorn and Zeileis (2021) proposed parametric variants based on fitting predefined distributional forms. Such specifications can be attractive if the number of training observations is limited. By contrast, this paper uses Cevid et al. (2022)'s non-parametric version of the distributional random forest as a fully flexible device to estimate the relationship between outcome distributions and covariates using a large data set.

The purpose of this paper is to apply distributional random forests to several estimation problems in the analysis of the income distribution. We consider four applications: (i) estimating income distributions for granular population subgroups, (ii) analysing distributional change over time, (iii) small-area estimation of income distributions, and (iv) purging spatial income distributions of differences in spatial characteristics. Application (i) is commonly used by governments and statistical agencies to monitor the well-being of population subgroups and to inform policy measures (e.g., poverty alleviation). Task (ii) decomposes changes in the aggregate distribution over time by separating changes attributable to shifts in population composition from changes due to income dynamics within population subgroups. Application (iii) is also a common task of governments and statistical agencies aimed at constructing maps of statistical information on quantities such as median income, at-risk-of-poverty indices or income inequality across geographical areas with potentially sparse observations. This question has been addressed by a large literature on small-area estimation, see Tzavidis et al. (2018) and Molina et al. (2022) for overviews. Given the inherent smoothing properties of random forests (Lin and Jeon, 2006), this method appears well-suited for small-area estimation. Indeed, Krenmair and Schmid (2022) recently incorporated a random forest component into a small-area mixed effects model for estimating area-level means. In this paper, we use the DRF to estimate area-level distributions with the goal of constructing area-level statistical indices (means, at-risk-of-poverty rates, inequality indices). In a final application (iv), we consider the problem of purging spatial income distributions of differences in spatial characteristics. This isolates the 'pure' spatial structure of income levels and inequality, independent of variation in age, employment, education, and other characteristics across spatial units. To the best of our knowledge, this application is novel in the literature

Our empirical analysis is based on the German Microcensus, an annual survey conducted by the Federal Statistical Office of Germany (Federal Statistical Office, 2024). The Microcensus is the largest sample survey in Germany and in Europe. Despite its large sample size and exceptional representativeness, it has rarely been used for income distribution analysis because it reports income in grouped form. While grouped income information necessarily limits the information content, we

demonstrate in this paper how the distributional random forest can effectively deal with this issue.

In addition to demonstrating the usefulness of the distributional random forest approach for analysing the income distribution, this paper contributes a number of substantive results on the German income distribution based on Microcensus data for the years 2005 and 2019. Specifically, we provide new evidence on the incomes and poverty risk of granular population subgroups and analyse distributional change over time. We show that inequality and poverty risk increased between 2005 and 2019, but that this was the result of changes in the composition of the population rather than by income changes within population subgroups. Finally, we provide distributional maps of household income and inequality for Germany at a much higher geographical resolution than previous analyses (Immel and Peichl, 2020; Walter et al., 2022).

The remainder of this paper is structured as follows. Section 2 outlines the method of distributional random forests due to Cevid et al. (2022) and Näf et al. (2023). Section 3 provides basic information on the data. Sections 4 to 7 present our empirical applications, section 8 presents simulation evidence, and section 9 concludes.

2. Distributional random forest

We outline the main properties of the distributional random forest (DRF) as introduced by Cevid et al. (2022) and Näf et al. (2023). Let $\mathbf{Y} = (Y_1, \ldots, Y_d) \in \mathbb{R}^d$ denote a potentially multivariate outcome vector and $\mathbf{X} = (X_1, \ldots, X_p) \in \mathbb{R}^p$ a vector of covariates. The goal of the DRF is to estimate the conditional distribution $\mathbb{P}(\mathbf{Y}|\mathbf{X} = \mathbf{x})$ based on a random sample $(\mathbf{y}_i, \mathbf{x}_i), i = 1, \ldots n$.

The DRF produces an estimate $\hat{\mathbb{P}}(\mathbf{Y}|\mathbf{X} = \mathbf{x})$ of the conditional distribution by repeating a recursive partitioning algorithm (= tree building) $k = 1, \ldots, N$ times on random pertubations of the data and by averaging the results (= random forest). For each tree k, the sample is successively partitioned into groups of observations (= leaves). The partitioning proceeds greedily by splitting a parent node P into two child nodes $C_L = \{X_j \leq l\}$ and $C_R = \{X_j > l\}$ based on candidate splitting variables X_j that are chosen randomly (see below). The split is chosen to maximize the difference between C_L and C_R with respect to a specified objective function.

In Breiman (2001)'s original random forest for mean outcomes, splits were performed so that (in the case of an univariate outcome), the resulting mean outcomes in C_L and C_R differed as much as possible, i.e.,

$$\max_{l: C_L = \{X_j \leq l\}, C_R = \{X_j > l\}} \frac{n_L n_R}{n_P^2} \left(\frac{1}{n_L} \sum_{i \in C_L} y_i - \frac{1}{n_R} \sum_{i \in C_R} y_i \right)^2, \tag{1}$$

where n_L, n_R and n_P are the number of observations in the child and parent nodes, respectively.

In the DRF, splits are performed to maximize *distributional differences* between the resulting child nodes C_L and C_R. Distributional differences are measured by the Maximum Mean Discrepancy (MMD) statistic (Gretton et al., 2007). The MMD statistic builds on the theory of distribution embeddings in Reproducing Kernel Hilbert Spaces (RKHS) (Muandet et al., 2017).

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ denote a RKHS of real valued functions on $\mathbb{R}^d$ induced by a positive-definite kernel $k(\cdot,\cdot)$ with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$, norm $\|\cdot\|_{\mathcal{H}}$, and implicit feature map $\varphi : \mathbb{R}^d \to \mathcal{H}$ satisfying $k(\mathbf{y}, \mathbf{y}') = \langle\varphi(\mathbf{y}), \varphi(\mathbf{y}')\rangle_{\mathcal{H}}$. The feature map $\varphi(\mathbf{y})$ can be interpreted as a (possibly infinite-dimensional) collection of aspects of $\mathbf{y}$. The term $k(\mathbf{y}, \mathbf{y}')$ is then a measure of similarity between points $\mathbf{y}$ and $\mathbf{y}'$ in terms of all their aspects captured by the feature map. This similarity measure is linear in the feature space, but may be very nonlinear in the original space $\mathbb{R}^d$, depending on the richness of the feature map ('kernel trick').

Let $\mathcal{P}$ be a distribution and define

$$\mu(\mathcal{P}) = \mathbb{E}_{\mathbf{Y}\sim\mathcal{P}}\left[\varphi(\mathbf{Y})\right] \tag{2}$$

as its mean embedding in the Hilbert space $\mathcal{H}$ (i.e., every distribution $\mathcal{P}$ is represented as an element of $\mathcal{H}$). For certain choices of the kernel (i.e., characteristic kernels), this mapping is one-to-one, so that each distribution is uniquely represented by one element in the RKHS. Differences between two distributions $\mathcal{P}$ and $\mathcal{Q}$ can thus be measured by the distance function in the corresponding Hilbert space, i.e., $d(\mathcal{P}, \mathcal{Q}) = \|\mu(\mathcal{P}) - \mu(\mathcal{Q})\|^2_{\mathcal{H}}$ (i.e., the distance between their mean embeddings in the Hilbert space).

The distributional random forest uses this distance measure between the distributions of outcomes in two child nodes C_L and C_R to find splits that make distributions in C_L and C_R as different as possible. In this case, the MMD statistic is

$$\begin{aligned}\mathcal{D}_{\text{MMD}}(C_L, C_R) &= \|\mu(\mathcal{P}_{C_L}) - \mu(\mathcal{P}_{C_R})\|^2_{\mathcal{H}} \\ &= \left\| \frac{1}{n_L} \sum_{i\in C_L} \varphi(\mathbf{y}_i) - \frac{1}{n_R} \sum_{i\in C_R} \varphi(\mathbf{y}_i) \right\|^2_{\mathcal{H}}.\end{aligned} \tag{3}$$

Note the similarity to Breiman (2001)'s original splitting criterion (1), which results when the feature map consists only of the value $\mathbf{y}$ itself (i.e., $\varphi(\mathbf{y}) = \mathbf{y}$). In this case, the statistic only measures average differences in the *level* of $\mathbf{y}$. By contrast, if the feature map is richer, it measures average differences *in all features* encoded by the feature map $\varphi(\cdot)$ (see (3)). For example, if $\varphi(\mathbf{y})$ includes higher-order terms of $\mathbf{y}$, it will not only measure differences in means between C_L and C_R but in higher-order moments. It can be shown that the implicit feature maps of characteristic kernels is infinite-dimensional and powerful enough to detect *any* differences between distributions (Gretton et al., 2007).[1] The MMD statistic can be equivalently written as

$$\begin{aligned}\mathcal{D}_{\text{MMD}}(C_L, C_R) = &\frac{1}{n_L^2} \sum_{i,j\in C_L} k(\mathbf{y}_i, \mathbf{y}_j) \; + \frac{1}{n_R^2} \sum_{i,j\in C_R} k(\mathbf{y}_i, \mathbf{y}_j) \\ &- \frac{2}{n_L n_R} \sum_{i\in C_L} \sum_{j\in C_R} k(\mathbf{y}_i, \mathbf{y}_j).\end{aligned} \tag{4}$$

[1] In our empirical application, we use the Gaussian kernel as in Cevid et al. (2022).

This formulation provides an intuitive interpretation: the statistic measures how similar – as described by the kernel – observations are *within* each sample, as compared to *between* the two samples.

As in Breiman (2001)'s original splitting criterion (1), the distributional random forest uses a version of $\mathcal{D}_{\text{MMD}}$ that is rescaled by the factor $n_L n_R / n_P^2$. In order for the forest to be consistent for the true conditional distribution $\mathbb{P}(\mathbf{Y}|\mathbf{X} = \mathbf{x})$, forest construction has to comply with a number of rules (Athey et al., 2019; Cevid et al., 2022):

1. *Honesty:* Splits are determined on one half of the data, distributional predictions are computed on the other half of the data.
2. *Random-split:* The probability that the split occurs along feature X_j is bounded from below by π/p for some $\pi > 0$ (p is the number of covariates).
3. *Symmetry:* The tree output does not depend on the ordering of the training samples.
4. *Regularity:* Each child contains at least a fraction $\alpha \leq 0.2$ of the parent node. Trees are grown until each leaf contains between κ and $2\kappa - 1$ observations.
5. *Subsampling:* Trees are grown on subsamples of size $s_n = n^\beta$ drawn from of the original n sample observations, where β must lie within particular bounds depending on p, π and α (Cevid et al., 2022).

The distributional random forest $\mathcal{F}$ is based on N trees $\mathcal{T}_1, \ldots, \mathcal{T}_N$ grown according to the rules above. Define $\mathcal{L}_k(\mathbf{x})$ as the set of training data observations that end up in the same leaf as $\mathbf{x}$ in tree $k = 1, \ldots, N$. The main output of the distributional random forest is a set of observation- and test-point-specific weights

$$\hat{w}_i(\mathbf{x}) = \frac{1}{N} \sum_{k=1}^{N} \frac{1(\mathbf{x}_i \in \mathcal{L}_k(\mathbf{x}))}{|\mathcal{L}_k(\mathbf{x})|}, \tag{5}$$

measuring the frequency with which training observation $i = 1, \ldots, n$ ended up in the same leaf as a test point with $\mathbf{X} = \mathbf{x}$. The weights quantify the importance of each training data point $(\mathbf{y}_i, \mathbf{x}_i), i = 1, \ldots, n$ for predicting the conditional distribution of $\mathbf{Y}$ at test-point $\mathbf{X} = \mathbf{x}$. Formally, the resulting estimate is given by

$$\hat{\mathbb{P}}(\mathbf{Y}|\mathbf{X} = \mathbf{x}) = \sum_{i=1}^{n} \hat{w}_i(\mathbf{x}) \cdot \delta_{\mathbf{y}_i}, \tag{6}$$

where $\delta_{\mathbf{y}_i}$ denotes the point mass at $\mathbf{y}_i$. The weights (5) characterize the distributional random forest as a locally adaptive nearest-neighbour method that smooths observations across the covariate space (Lin and Jeon, 2006).

Cevid et al. (2022) show that (6) is consistent in the sense that the estimated conditional distribution function converges in probability to the true conditional distribution function. This, in turn, implies that smooth functionals of the conditional distribution also converge to their population counterparts. In practice, this means that the random forest weights $\hat{w}_i(\mathbf{x})$ can be used to compute any statistic of interest $I(\hat{\mathbb{P}}(\mathbf{Y}|\mathbf{X} = \mathbf{x}))$ (e.g., the mean, the Gini index, the at-risk-of-poverty rate) based on the plug-in principle. This approach yields estimates for small subgroups

with $\mathbf{X} = \mathbf{x}$ even when these subgroups are only weakly represented – or not represented at all – in the sample, by borrowing information from training observations with covariates that are most similar to $\mathbf{x}$.

Näf et al. (2023) show that, under suitable conditions, the mean embeddings of the distributional random forest estimators are asymptotically normal, which implies that sufficiently smooth functionals based on the random forest weights are also asymptotically normal. Moreover, these sampling distributions can be practically simulated by a bootstrap half-sampling procedure. To this end, $b = 1, \ldots, B$ half-samples $\mathcal{S}_b$ are drawn from the original training observations. For each half-sample, L trees are grown to build 'mini forests' $\mathcal{F}_b, b = 1, \ldots, B$. The weights $\hat{w}_i^b(\mathbf{x})$ of these mini forests then serve as bootstrap replications of the original weights $\hat{w}_i(\mathbf{x})$ to compute bootstrap versions of the statistics of interest. The procedure can be efficiently used to construct the overall forest consisting of $N = L \cdot B$ trees by pooling the L mini forests to form the total forest. In our empirical analysis, we use the procedures of Cevid et al. (2022) and Näf et al. (2023) to compute point estimates and confidence intervals for the statistics of interest.

The following algorithm summarizes the steps of the estimation procedure:

Algorithm 1: Pseudocode for computation of DRF point estimate and standard errors for inequality statistic $I(\mathbf{x}) = I(\hat{\mathbb{P}}(\mathbf{Y}|\mathbf{X} = \mathbf{x}))$ at test point $\mathbf{x}$

procedure BUILDFOREST($\mathbf{Y}, \mathbf{X}$) `// training data` $(\mathbf{Y}, \mathbf{X})$

  for $b \leftarrow 1$ **to** B **do**
    $\mathcal{S}_b \leftarrow$ draw a half-sample from original sample
    $\mathcal{F}_b \leftarrow \text{DRF}(\mathcal{S}_b, L)$ `// mini forest number` b
  return $\{\mathcal{F}_1, \ldots, \mathcal{F}_B\}$ `// mini forests`

procedure FORESTWEIGHTS($\mathbf{x}$) `// for test point` $\mathbf{x}$

  for $b \leftarrow 1$ **to** B **do**
    $w_b(\mathbf{x}) \leftarrow$ get weights from mini forest $\mathcal{F}_b$ at test point $\mathbf{x}$
  $w(\mathbf{x}) \leftarrow \frac{1}{B} \sum_{b=1}^{B} w_b(\mathbf{x})$
  return $w(\mathbf{x})$ `// weights for overall forest at` $\mathbf{x}$
  return $w_1(\mathbf{x}), \ldots, w_B(\mathbf{x})$ `// weights for bootstrap replications`

procedure ESTIMATESANDSTDERRORS($\mathbf{x}$) `// for test point` $\mathbf{x}$

  $I(\mathbf{x}) \leftarrow I\big(w(\mathbf{x}), \mathbf{Y}, \mathbf{X}\big)$ `// inequality point estimate at` $\mathbf{X} = \mathbf{x}$
  for $b \leftarrow 1$ **to** B **do**
    $I_b(\mathbf{x}) \leftarrow I\big(w_b(\mathbf{x}), \mathbf{Y}, \mathbf{X}\big)$ `// bootstrap replications`
  $\text{Stderr}_{I(\mathbf{x})} \leftarrow \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \Big(I_b(\mathbf{x}) - \frac{1}{B} \sum_{j=1}^{B} I_j(\mathbf{x})\Big)^2}$
  return $I(\mathbf{x})$ `// inequality point estimate at` $\mathbf{X} = \mathbf{x}$
  return $\text{Stderr}(I(\mathbf{x}))$ `// bootstrap standard error`

3. Data

Our analysis is based on the German Microcensus for the years 2005 and 2019 (Federal Statistical Office, 2024). The Microcensus is conducted annually and provides a 1% random sample of the German population, including information on income and socio-economic characteristics for all persons in the surveyed households. It is the largest sample survey in Germany and in Europe. Data quality is high, and non-response is low due to mandatory participation. Most parts of our analysis rely on the Scientific Use File (SUF) of the Microcensus (Federal Statistical Office, 2024). For analyses requiring local identifiers at the municipality level, we use a restricted version of the Microcensus, accessible only on site at the Research Data Centers (RDC) of the Federal Statistical Offices.

Although the Microcensus is the largest and most representative sample survey for Germany, it has rarely been used for income distribution analysis (Boehle, 2015; Hochgürtel, 2019; Walter et al., 2022). One reason is that income is recorded in grouped form. In the two survey years considered here, respondents were asked to provide information on monthly household net income in income brackets of increasing width. The income brackets used for grouped income data are given in table 1. Note that the top bracket is open-ended.

Table 1. Income brackets for monthly household net income (euros)

(0; 150]	(150; 300]	(300; 500]	(500; 700]	(700; 900]
(900; 1,100]	(1,100; 1,300]	(1,300; 1,500]	(1,500; 1,700]	(1,700; 2,000]
(2,000; 2,300]	(2,300; 2,600]	(2,600; 2,900]	(2,900; 3,200]	(3,200; 3,600]
(3,600; 4,000]	(4,000; 4,500]	(4,500; 5,000]	(5,000; 5,500]	(5,500; 6,000]
(6,000; 7,500]	(7,500; 10,000]	(10,000; 18,000]	(18,000; ∞)	

Source: German Microcensus, 2005 and 2019.

Following standard practice, we adjust household income using the OECD equivalence scale. This scale assigns a weight of one to the first household member, a weight of 0.5 to each additional member aged 15 and over, and a weight of 0.3 to each additional member aged 14 and under. For example, if a household's net income falls within the interval (4,000; 4,500], the equivalised income for a household with two adults and two children (equivalence weight $= 1 + 0.5 + 0.3 + 0.3 = 2.1$) lies in the interval (1,905; 2,143]. As the distributional random forest can handle multivariate outcomes, we define the lower and upper bounds of the (equivalised) income interval as the dependent outcome, i.e., $\mathbf{Y} = (y_{\text{lower}}, y_{\text{upper}})$.

To ensure applicability across all income groups, we impose an upper bound on the highest income bracket, which is open-ended in the data. Following Walter et al. (2022), we define this limit as $3 \cdot 18,000 = 54,000$, resulting in a top interval of (18,000; 54,000]. While Walter et al. (2022) did not provide a formal justification for their choice, a reasonable rationale is that household incomes in the Microcensus follow an approximate Pareto tail with tail parameter $a = 2$, implying that the midpoint of the interval (18,000; 54,000] coincides with the expected income of this group, i.e.,

Table 2. Covariates for the distributional random forest

	2005		2019	
Variable	Mean	Std. dev.	Mean	Std. dev.
# adults in hh	2.032	0.802	1.981	0.792
# adults 18-29 years	0.380	0.668	0.340	0.645
# adults 30-49 years	0.866	0.869	0.717	0.842
# adults 50-64 years	0.434	0.717	0.536	0.766
# adults 65+ years	0.351	0.668	0.386	0.698
# children in hh	0.691	1.028	0.642	0.642
# children 0-3 years	0.123	0.382	0.136	0.403
# children 4-6 years	0.304	0.651	0.403	0.637
# children 7-17 years	0.262	0.581	0.217	0.217
# adults foreign nationality	0.159	0.556	0.239	0.652
Share foreign adults > 0.5	0.090	0.286	0.134	0.341
# adults male	0.996	0.601	0.975	0.587
# adults female	1.035	0.496	1.006	0.508
0 FT[1], 0 PT, 0 MPT	0.294	0.454	0.253	0.408
0 FT, 0 PT, ≥ 1 MPT	0.028	0.164	0.029	0.167
0 FT, ≥ 1 PT, ≥ 0 MPT	0.046	0.207	0.072	0.258
1 FT, 0 PT, 0 MPT	0.240	0.426	0.202	0.401
1 FT, 0 PT, ≥ 1 MPT	0.059	0.235	0.049	0.215
1 FT, ≥ 1 PT, ≥ 0 MPT	0.117	0.320	0.171	0.376
≥ 2 FT, ≥ 0 PT, ≥ 0 MPT	0.212	0.408	0.221	0.414
# registered unemployed in hh	0.153	0.426	0.059	0.280
# unemployment benefits in hh	0.136	0.404	0.106	0.528
# adults tertiary education[2]	0.253	0.559	0.511	0.728
# adults higher secondary	0.174	0.452	0.196	0.476
# adults vocational training	1.097	0.894	0.919	0.865
# adults low education	0.507	0.775	0.775	0.689
East Germany	0.218	0.413	0.194	0.395
Indicators for 16 federal states		(details omitted)		
# individuals		440,268		506,615
# households		211,833		255,164

Source: Microcensus, 2005 and 2019. [1] FT = Full-time, PT = Part-time, MPT = Marginal part-time. [2] Highest educational qualification.

$\mathbb{E}(\text{household income}|\text{household income} > 18{,}000) = a/(a-1) \cdot 18{,}000 = 36{,}000$ (Blanchet and Piketty, 2022, p. 275). This approach is consistent with standard practice for grouped data, where interval midpoints are commonly used as approximations

to group means (see below). We find that our results are fairly robust to different choices of the upper limit, as only a small fraction of observations fall into the top income interval (0.21% in 2005 and 0.45% in 2019), see supplementary appendix for more details.

The equivalisation procedure leads to overlapping income intervals across observations, which does not pose a problem for the distributional random forest. To construct a proper distribution function $F^e(y \mid \mathbf{x})$ for equivalised income y conditional on characteristics $\mathbf{X} = \mathbf{x}$, we aggregate probability mass over the upper bounds of the income intervals, following standard practice in grouped-data analysis (an income bracket contributes fully to the probability mass only once its upper interval bound is reached). Concretely, we compute the share of observations with equivalised income less than or equal to A_1, less than or equal to A_2, and so on, where $A_1, A_2, \ldots$ denote the ordered distinct upper interval limits observed in the data. This yields the conditional cumulative distribution function of equivalised income and makes full use of the information contained in the grouped income variable.

We use the resulting income groups $(A_0; A_1], (A_1; A_2], \ldots, (A_{J-1}; A_J]$ and their implied frequencies f_j to compute statistics of interest (quantiles, means, and Gini coefficients) using the grouped-data formulas developed by Tille and Langel (2012). These formulas rely on linear interpolation within intervals for quantiles and the CDF, while the Gini formula is based on a quadratic interpolation of the Lorenz curve with an additional correction term for within-interval inequality. Let $l_j = A_j - A_{j-1}$ and $c_j = (A_j + A_{j-1})/2$ denote interval widths and midpoints, respectively. The grouped-data formulas for the mean, CDF, quantiles, and the Gini coefficient are given as follows (we omit conditioning on $\mathbf{X} = \mathbf{x}$):

$$\overline{x} = \sum_{j=1}^{J} f_j c_j \tag{7}$$

$$F^e(y) = \sum_{k<j(y)} f_k + f_{j(y)} \frac{y - A_{j(y)-1}}{l_{j(y)}},\ j(y) : A_{j-1} \leq y < A_j \tag{8}$$

$$Q(p) = A_{j(p)-1} + l_{j(p)} \frac{p - F^e(A_{j(p)-1})}{f_{j(p)}},\ j(p) : F^e(A_{j-1}) \leq p < F^e(A_j) \tag{9}$$

$$G_2 = \frac{1}{2\overline{x}} \sum_{j=1}^{J} \sum_{k=1}^{J} f_j f_k \left| c_j - c_k \right| + \sum_{j=1}^{J} \frac{f_j^2}{6} l_j \tag{10}$$

Our simulation exercise below suggests that interval censoring has only a minor influence on the values of these statistics in our application (relative to the uncensored case), because the number of intervals is relatively large.

We account for the Microcensus sampling (= grossing-up) weights when computing and aggregating distributions from the fitted random forest. Specifically, we incorporate these weights in the computation of f_j for results conditional on $\mathbf{X} = \mathbf{x}$, and we use them whenever we aggregate such results across $\mathbf{x}$, for example as in (11) below. By contrast, it is currently not possible to fully incorporate sampling weights into the training of the random forest. We do not expect this limitation to materially affect our estimates, because the variation in Microcensus weights is limited. As a

sensitivity check, we re-estimated selected specifications using a reweighted sample constructed from the original sampling weights. This led to results that were very close to those from the original sample, see supplementary appendix for more details.

4. Estimating income distributions for granular population subgroups

Our first goal is to estimate distributions of equivalised net income for narrowly defined population subgroups. This is an important task for monitoring the well-being of specific groups, especially those at risk of poverty or social exclusion. To define these subgroups, we leverage the rich set of socio-economic characteristics at the individual and household level provided by the Microcensus. Since equivalised income is based on the assumption of income pooling within households, all covariates are constructed at the household level (for the equivalised income of a given individual, it matters which household she lives in). A summary of the covariates $\mathbf{X}$ used in our analysis is shown in table 2.

Table 3. Tuning parameters of the distributional random forest

Tuning parameter	Range	Description
mtry	2, 3, 8, 12, 15, 20, 30	# variables tried at each split
min.node.size	2, 5, 10, 15, 20, 25	Target minimal leaf size
sample.fraction	0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1	Subsampling fraction
alpha	0, 0.01, 0.05, 0.1, 0.25	Maximum imbalance of split ($=\alpha$)
imbalance.penalty	0, 0.05, 0.1	Penalty for split imbalance

Note: See Cevid et al. (2022) for details.

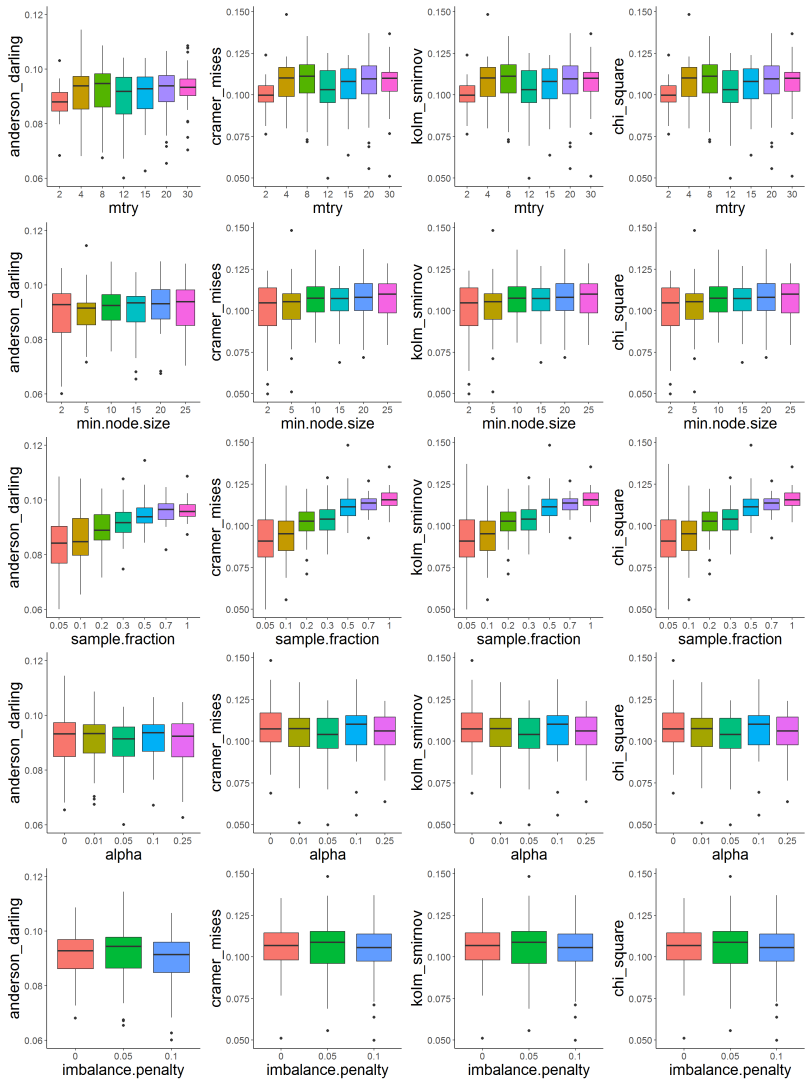
Our first step is to fit and tune the distributional random forest. Cevid et al. (2022) do not discuss tuning of the distributional random forest. To arrive at a practically feasible procedure, we proceed as follows. We use a training sample (40% of the original 2019 sample) and a test sample (30% of the original 2019 sample). The steps are as follows:

1. *Random parameter selection:* We randomly varied the tuning parameters within the ranges given in Table 3, generating 300 parameter combinations.
2. *Training:* For each parameter combination, we fitted the distributional random forest in the training sample and computed conditional cdf's $F^e(y|\mathbf{x})$ for all observations in the test sample.
3. *Testing:* We then computed the model-implied aggregate distribution of equivalised income,

$$F^e(y) = \int_{\mathbf{x}} F^e(y|\mathbf{x})\, dF_{\mathbf{X}}(\mathbf{x}) \tag{11}$$

on the test sample and compared it with the observed empirical distribution $F^{\text{emp}}(y)$ of equivalised incomes in the test sample.

Fig. 1: Tuning of the distributional random forest



Note: The figure shows the distribution of goodness-of-fit indicators across 300 specification variants for tuning parameters evaluated on a test sample (lower values indicate better fit).

4. *Evaluation:* We assessed goodness-of-fit using several statistical distance measures between $F^{e}(y)$ and $F^{\text{emp}}(y)$, including Anderson-Darling, Cramer–von Mises, Kolmogorov-Smirnov, and Chi-square statistics.

Table 4. Ten examples of population subgroups.

Label	Description	n_{2019}
Two pensioners	2 adults aged 65+ years (m+f), 2 vocational training, no employment, North-Rhine Westfalia	2,923
Single mother	1 female 18-29 years, PT, low education, 2 children (0-3, 4-11 years), Berlin	< 3
DINK	m+f 30-49 years, 2 FT, 2 tertiary education, Hamburg	66
5-person family BW	m+f 30-49 years, 1 FT, 1 PT, 1 tertiary, 1 voc. tr., 3 children (4-11, 2×12-17 years), Baden-Württemberg	11
Single immigrant	male 18-29 years, MPT, low education, foreign nationality, Lower Saxony	< 3
Single unemployed	male 30-49 years, registered unemployed, unemployment benefits, low education, Saxony	26
Immigrant family	m+f 30-49 years, 1 FT, 1 voc. training, 3 children (0-3, 2×4-11 years), foreign, North-Rhine Westfalia	7
Young professional	male 18-29 years, 1 FT, tertiary education, Berlin	84
Elderly widow	female 65+ years, no employment, Saxony-Anhalt	339
5-person family MW	m+f 30-49 years, 1 FT, 1 PT, 1 tertiary, 1 voc. tr., 3 children (4-11, 2×12-17 years), Meck-Westpomerania	< 3

Notes: m = male, f = female, FT = full-time, PT = part-time, MPT = marginal part-time, DINK = Double Income No Kids. The last column shows the number of cases in the 2019 sample.

The results of this exercise are shown in figure 1. By minimizing the discrepancy statistics between the predicted and observed income distributions within reasonable parameter ranges, we chose the final tuning parameters as mtry=12, min.node.size=5, sample.fraction=0.1, alpha=0.05, imbalance.penalty=0.1. We found that our random forest results typically did not vary much across different specifications of tuning parameters, which is reflected in the small differences in goodness-of-fit between alternative choices (figure 1). Using these tuning parameters, we fitted our final random forest on the full sample based on $L \cdot B = 250 \cdot 50 = 12{,}500$ trees. Our final model produced an aggregate income distribution function that was practically indistinguishable from the empirical distribution in the test set. This was generally true even for suboptimal tuning parameters.

Table 4 defines ten examples of narrowly defined population subgroups for which we estimate equivalised net income distributions using DRFs for 2005 and 2019. These subgroups consist of individuals in households with specific socio-economic characteristics, as summarized in table 2. In most cases, the number of observations in a given subgroup would be much too low to estimate meaningful income statistics directly. Like in regression analysis, the random forest estimates group-specific outcomes by leveraging observations with similar characteristics (i.e., observations from other regions and households that are similar in terms of age, education and employment behaviours of their members).

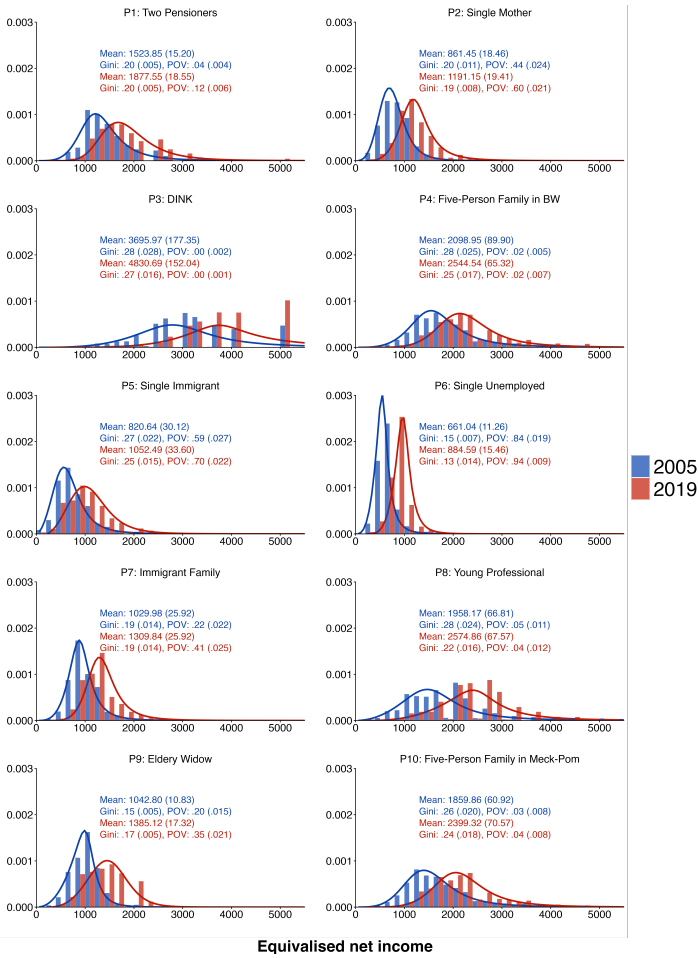
Figure 2 presents the results for different subgroups. For illustration, the graphs include a GB2 density fit. This is done solely for visualization purposes to highlight distributional shapes. The generalized beta distribution of the second kind (GB2) has been shown to provide a good fit to aggregate income distributions (Chotikapanich et al., 2018), but it fits less well for some of our subgroup-specific distributions. For each subgroup, we compute key statistics such as mean equivalised income, the at-risk-of-poverty rate, and the Gini coefficient using the grouped-data formulas in Tille and Langel (2012).[2]

The results shown in figure 2 exhibit highly plausible patterns. There is overall income growth from 2005 to 2019, but gains in mean income are heterogeneous. They range from 21.2% for individuals from the five-person family in BW to 38.3% for individuals in the single-mother household with two children, as defined in table 4. As expected, the at-risk-of-poverty rate varies considerably across household types, from zero percent in double-income-no-kids households to around 90% in the single unemployed household. The characteristics of the latter were intentionally chosen to be unfavorable in order to illustrate an extreme case. Finally, there are large differences in within-group inequality across subgroups as measured by the Gini coefficient. Some groups are extremely homogeneous (individuals in single unemployed households with a Gini coefficient of around 0.14), while others are highly heterogenous even within the narrow type definitions considered here (individuals in the double-income-no-kids household with a Gini of around 0.28).We also observe a general trend of

[2] Following standard practice in European countries, the at-risk-of-poverty rate is defined as the proportion of the population with equivalised income below 60% of the population median income.

decreasing within-group inequality for most types between 2005 and 2019 (an exception are elderly widows). In summary, once fitted, the distributional random forest enables policymakers and statistical agencies to flexibly monitor multiple aspects of economic welfare for finely defined population subgroups.

Fig. 2: Distribution of equivalised net incomes in narrow population subgroups



Source: Microcensus 2005 and 2019. Own computations.
Notes: See table 4 for the definition of types. Blue = 2005, red = 2019. POV = At-risk-of poverty rate (share of incomes below 60 % of population median). The figures include a GB2 density fit for illustration. Bootstrap standard errors in parentheses ($B = 50$ mini forests).

5. Analysis of distributional change over time

The distributional random forest captures information on income distributions for finely defined population subgroups. It can therefore be used to decompose changes in the aggregate distribution into compositional and structural components (Fortin et al., 2011). To this end, consider the counterfactual distribution

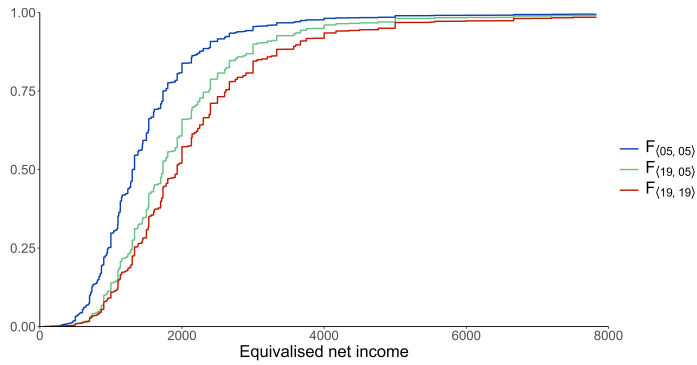
$$F^c_{\langle 19,05\rangle}(y) = \int_{\mathbf{x}} F^e_{2019}(y|\mathbf{x})\, dF_{\mathbf{X},2005}(\mathbf{x}), \tag{12}$$

which is the distribution of equivalised incomes that would have prevailed in 2019, if the distribution of household characteristics $F_{\mathbf{X}}(\mathbf{x})$ remained as in 2005. This gives rise to the decomposition

$$F_{\langle 19,19\rangle}(y) - F_{\langle 05,05\rangle}(y) = \underbrace{F_{\langle 19,19\rangle}(y) - F^c_{\langle 19,05\rangle}(y)}_{\text{composition effect}} + \underbrace{F^c_{\langle 19,05\rangle}(y) - F_{\langle 05,05\rangle}(y)}_{\text{structural effect}}, \tag{13}$$

i.e., changes in the distribution of equivalised incomes between 2005 and 2019 are decomposed into effects attributable to changes in population composition $F_{\mathbf{X}}$, and by changes in income structures $F^e(y|\mathbf{x})$ as captured by the distributional random forest.

Fig. 3: Aggregate decomposition, 2005-2019



Source: Microcensus 2005 and 2019. Cumulative distribution functions.

Results for this decomposition are presented in table 5 and figure 3. Between 2005 and 2019, we observe overall income growth, but also an increase in inequality and poverty risk: mean equivalised income increased from 1,556 to 2,298 euros, the median rose from 1,307 to 1,928, while the Gini coefficient increased from 0.304 to 0.321 and the at-risk-of-poverty rate rose from 0.136 to 0.171. Inequality increased primarily in the lower half of the distribution: the P90/P10-ratio rose from 3.384 to 3.616, but this was entirely driven by an increase of the P50/P10-ratio from 1.845 to 1.955, while the P90/P50-ratio only changed very little from 1.833 to 1.876.

The counterfactual results in the middle column of table 5 suggest that holding the population composition fixed at its 2005 level while updating income structures

Table 5. Decomposition of distributional change, 2005-2019

	2005	Counterfactual[1]	2019
Mean	1,555.731	2,022.35	2,298.12
	(3.81)	(3.91)	(4.54)
Gini	0.304	0.297	0.321
	(0.001)	(0.001)	(0.001)
P10	708.07	915.97	1,000
	(2.61)	(8.76)	(0.01)
P50	1,306.66	1,731.65	1928.11
	(11.21)	(0.07)	(0.43)
P90	2,396.40	3,034.72	3,616.20
	(0.33)	(15.70)	(3.50)
P90/P10	3.384	3.313	3.616
	(0.012)	(0.038)	(0.004)
P90/P50	1.833	1.753	1.876
	(0.015)	(0.009)	(0.002)
P50/P10	1.845	1.891	1.955
	(0.015)	(0.009)	(0.002)
At-risk-of-poverty rate	0.136	0.140	0.171
	(0.001)	(0.001)	(0.001)

Source: Microcensus 2005 and 2019. Bootstrap standard errors in parentheses. Differences between 2005 and 2019 are statistically significant at 1% level.

[1]Population composition from 2005, income structure from 2019.

to their 2019 levels accounts for most of the observed income growth between 2005 and 2019. However, this shift has little impact on inequality and poverty levels. Indeed, when only income structures $F^e(y|\mathbf{x})$ are updated (while keeping composition constant), inequality as measured by the Gini coefficient and the P90/P10 ratio even slightly declines (from 0.304 to 0.297, and from 3.384 to 3.313, rows 2 and 6 of table 5). Updating income structures alone also leads to small increases in the at-risk-of-poverty rate and the P50/P10 ratio (from 0.136 to 0.140, and from 1.845 to 1.890, respectively), and to slight inequality reductions in the upper half of the distribution, as indicated by the counterfactual fall of the P90/P50-ratio from 1.833 to 1.753 – but these effects are modest.

In contrast, updating the population composition to its 2019 level implies large increases in inequality and poverty risk (middle vs. last column of table 5). These compositional shifts largely account for the observed increase in inequality and poverty between 2005 and 2019, suggesting that the rise in inequality over this period can be fully explained by compositional changes in the population.

How did the composition of the population change between 2005 and 2019? Table 2 summarizes these shifts. We observe population ageing, an increasing share of households with foreign nationals, greater heterogeneity in employment outcomes, and growing polarization in educational qualifications. Together, these changes increased

population heterogeneity, which in turn amplified income inequality. In some cases, the observed shifts also increased the proportion of low-income households, thereby raising the aggregate at-risk-of-poverty rate.

Figure 3 provides a graphical summary of the decomposition. Updating income structures shifts the distribution markedly upward, with little apparent change in inequality. Updating population composition in addition yields only modest further income growth but stretches the distribution to the right, consistent with higher inequality.

6. Small-area estimation of income distributions

In this section, we leverage the smoothing properties of the distributional random forest to estimate local income distributions. To this end, we use detailed geographical identifiers from the Microcensus down to the municipality level, along with additional municipality-level predictors (INKAR, 2024) that are commonly used in small-area estimation (Fabrizi et al., 2020; Gardini et al., 2022; Molina et al., 2022; De Nicolo et al., 2024). Germany has approximately 10,000 municipalities, including around 2,000 towns and cities and roughly 8,000 smaller administrative entities that combine multiple geographical units.[3]

To estimate local distributions of equivalised net income, we fit a distributional random forest using both the latitude and longitude of geographical units as well as area-specific predictors $\mathbf{m}$. In particular, we estimate

$$F^e(y|(\text{latitude}, \text{longitude}), \mathbf{m}), \tag{14}$$

where (latitude, longitude) refer to the center of a geographical unit. The vector $\mathbf{m}$ collects the following area-specific information obtained from INKAR (2024):

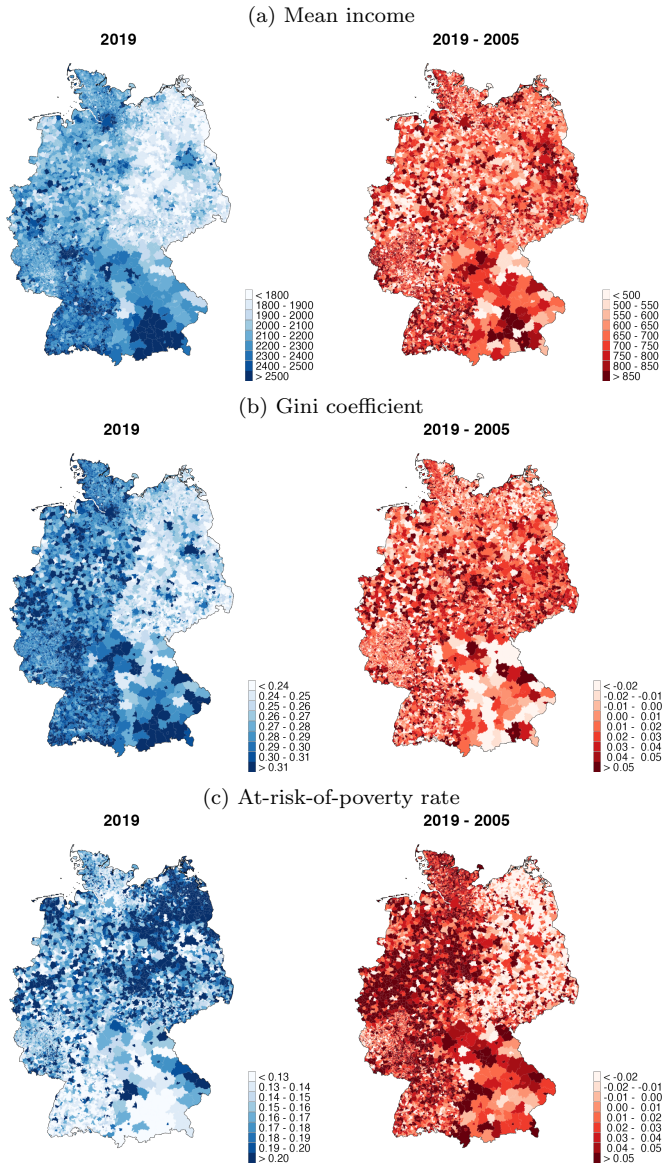
Population size
Unemployment rate
Tax revenue per inhabitant
Population share of age group 0-18
Population share of age group 18-30
Population share of age group 30-65
Population share of age group 65 years or over
Population density (= population/area)

Our approach combines smoothing across area-specific predictors $\mathbf{m}$ (as in classical small-area estimation) with direct smoothing across geographical space, which is absent from most small-area methods.[4] The random forest is a natural tool to smooth over both area-specific predictors and geographical locations as it is based

[3] Due to stricter data protection rules, geographical identifiers for Bavaria are available only at the county level, the next administrative tier above municipalities. We therefore use county-level data for Bavaria while retaining municipality-level data for all other regions.

[4] An exception is Sugasawa et al. (2020), who smooth local income distributions using a latent spatial correlation structure but do not incorporate additional area-specific predictors.

Fig. 4: Distributional indices and their change between 2005 and 2019



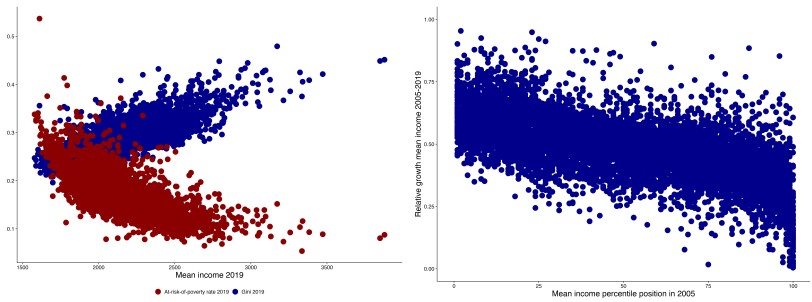
Source: Microcensus 2005 and 2019. Computations are based on estimated distributions of equivalised net incomes at the municipality level (county level for Bavaria).

on grouping observations that are 'similar' both in terms of the area-specific predictors and geographical location. We obtain estimates of local income distributions

$F^e(y|(\text{latitude}, \text{longitude}), \mathbf{m})$ for spatial unit $[(\text{latitude}, \text{longitude}), \mathbf{m}]$ from which we compute location and inequality measures as in the previous sections.

Figure 4 presents maps of distributional indices for Germany. To the best of our knowledge, these are the first municipality-level maps for Germany reporting distributional indices for household net incomes. Net incomes are widely regarded as the most informative indicators for personal financial well-being as they represent net disposable incomes after government transfers, taxes and social security deductions. Frieden et al. (2023) and Garbasevschi et al. (2023) provide municipality-level maps for pre-tax incomes, while Immel and Peichl (2020) and Walter et al. (2022) analyse regional differences in household net incomes at a more aggregated spatial level than municipalities. Schluter and Trede (2024) present a spatial analysis of wage incomes across regional labour markets, which are also defined at a higher level than municipalities.

Fig. 5: Associations between distributional aspects across geographical units



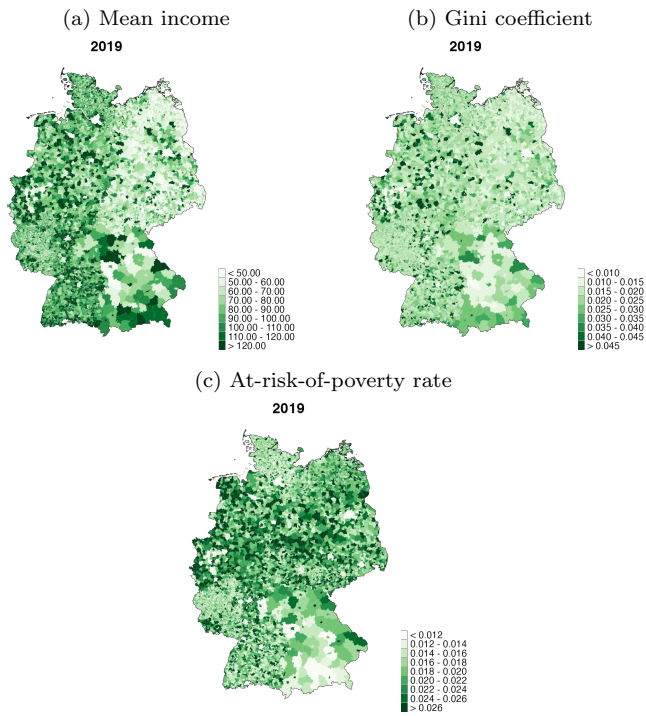
Source: Microcensus 2005 and 2019. The left panel plots the Gini coefficient and the at-risk-of-poverty rate against mean equivalised income for 2019. The right panel plots the growth rate of mean income from 2005 to 2019 against the percentile rank of mean income in the 2005 distribution across geographical units.

The local distributional indices presented in the maps have several important applications. First, they allow statistical agencies and policymakers to monitor local levels of well-being and to identify areas with high or low income and inequality. Second, the high degree of spatial heterogeneity is interesting in its own right, providing useful variation for studying relationships between different aspects of the distribution. For example, the left-hand graph in figure 5 plots the Gini coefficient and the at-risk-of-poverty rate against the mean income of geographical units. Mean income and inequality as measured by the Gini coefficient are positively related, i.e., geographical units with high mean equivalised incomes also tend to exhibit higher income inequality. In contrast, there is a weakly negative relationship between mean income and the at-risk-of-poverty rate, which likely reflect that the poverty threshold is defined at the national level (60 % of national median income). The right-hand graph of figure 5 relates the growth rate of mean income in a geographical unit to the original relative position of the unit in the base year 2005. The results indicate that units

with relatively low mean income in 2005 experienced higher relative income growth than those with a higher initial income level, suggesting convergence of mean incomes across regions. However, growth rates exhibit considerable variation, indicating that this relationship is only approximate.

An important additional application of the data in figure 4 is its potential as explanatory variables in microeconomic or spatial analyses. Local measures of income, inequality, or poverty can serve as covariates in studies of individual behaviour (e.g., the effect of local inequality on individual consumption behaviour), or local outcomes (e.g., the impact of local poverty rates on local election outcomes). In order to support such applications, we make our estimates of distributional indices for the roughly 9,000 geographical units considered in this paper available on request.

Fig. 6: Standard errors for distributional indices



Source: Microcensus 2019. Bootstrap standard errors ($B = 50$ mini forests).

As described in section 2, the distributional random forest includes a bootstrap procedure that allows one to compute standard errors for the spatial estimates shown in figure 4. We report these estimates in figure 6. The estimates for the standard errors are based on $B = 50$ mini forests with $L = 200$ trees each, leading to $N = B \cdot L = 50 \times 200 = 10{,}000$ trees for the overall forest. The magnitude of the standard

errors indicates reasonable estimation precision compared to the typical magnitude of the different statistics.[5]

To put our results into context, we compare them to estimates from a standard Fay-Herriot (FH) small-area model (Fay and Herriot, 1979). The FH model is based on regressing the direct inequality estimate $\widehat{I}_d$ for spatial unit d computed from the raw data on the unit's characteristics $\mathbf{m}_d$ (see above), i.e.

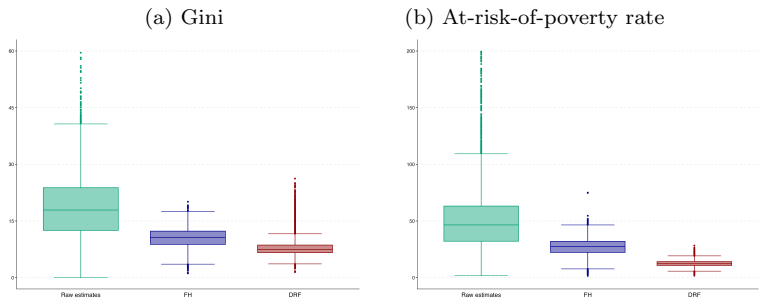
$$\widehat{I}_d = \mathbf{m}_d'\beta + u_d + e_d \tag{15}$$

where u_d with $\sigma_u^2 = \text{var}(u_d)$ is the error variance of the regression model, and e_d with $\sigma_{e_d}^2 = \text{var}(e_d) = \text{var}(\widehat{I}_d)$ denotes the sampling error of the raw estimate $\widehat{I}_d$. The Fay-Herriot predictions for spatial unit d result as a weighted average of the raw estimate and the regression prediction, i.e.

$$\widehat{I}_d^{FH} = \widehat{\gamma}_d \widehat{I}_d + (1 - \widehat{\gamma}_d)\mathbf{m}_d'\widehat{\beta} \quad \text{with} \quad \widehat{\gamma}_d = \widehat{\sigma}_u^2/(\widehat{\sigma}_u^2 + \widehat{\sigma}_{e_d}^2) \tag{16}$$

using estimated $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_{e_d}^2$. We compute the unit-specific $\widehat{I}_d$ on the raw data using the grouped-data formulas (7) to (10) and compute their sampling variances $\widehat{\text{var}}(e_d)$ by bootstrapping. Our Fay-Herriot predictions are based on the implementation of Halbmeier et al. (2019) for the Gini coefficient and the at-risk-of-poverty rate, using the arcsine transformation (because both statistics take values in the unit interval).[6]

Fig. 7: Box plots of the coefficients of variation



Source: Microcensus 2019. FH = Fay-Herriot, DRF = Distributional Random Forest. Only spatial units with at least three Microcensus observations included.

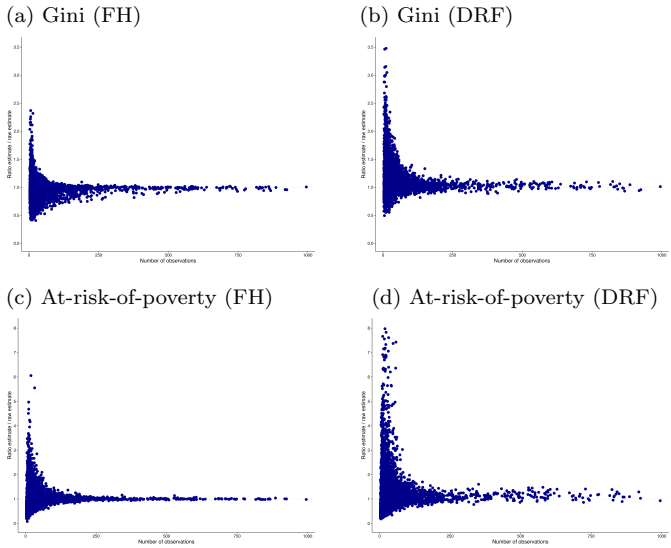
To gauge differences in precision across different methods, figure 7 presents the box plots for the coefficient of variation for the direct (raw) estimates (based on the grouped-data formulas and their bootstrap standard errors), the FH and the DRF estimates. The coefficient of variation for spatial unit d is defined as $\text{CV}_d =$

[5] We acknowledge that using a higher number of mini forests would be advantageous for the bootstrap, but substantial computational constraints at the research data centers where the restricted Microcensus data must be processed prevent us from increasing B while still maintaining a sufficiently large number of trees per mini forest.

[6] The results do not differ much when we do not use the arcsine transformation.

Stderr$(\widehat{I}_d)/\widehat{I}_d$ and summarizes its relative precision. The comparison only refers to in-sample-units, i.e. municipalities that are represented in the Microcensus.[7] The results suggest that both the FH and the DRF model improve upon the direct estimates, and that the DRF estimates compare favourably to the FH ones in terms of precision.

Fig. 8: Ratio of model estimates to direct estimates



Source: Microcensus 2019. FH = Fay-Herriot, DRF = Distributional Random Forest. Only spatial units with at least three Microcensus observations are included.

To gain more insights, figure 8 presents the ratio of FH/DRF model estimates to the direct (raw) sample estimates plotted against the sample size of the underlying spatial units. For both FH and DRF estimates, this ratio clusters more tightly around one for larger underlying sample sizes. Note that, for the FH model, this is mechanically true by construction, see equation (16). For the DRF, this is not mechanically true and thus serves as a plausibility check. It is well-known that Gini estimates from small samples tend to be downward biased (De Nicolo et al., 2024). This is also true of the at-risk-of-poverty rate as the extreme left tail of the distribution is typically

[7] In addition, we are not allowed to report information on raw estimates for municipalities with fewer than three observations in the Microcensus due to data protection rules. From the 8,857 German municipalities considered by us (counties for Bavaria), 5,086 are represented by at least three observations in the Microcensus and are thus reported in figures 7 and 8. A further 125 are represented by only 1 or 2 observations, while 3,646 municipalities have no observations in the Microcensus at all. The municipalities not covered (or covered by only 1 or 2 observations) are very small in size (median population = 510) and together account for only 3.2 percent of the German population, see supplementary appendix for more information.

underrepresented in small samples. The DRF appears to correct part of this small-sample downward bias (ratios above one at low sample sizes), whereas the FH model does not. More generally, note that a ratio of one is not necessarily the appropriate benchmark as both the FH and the DRF model exploit *additional* information (FH: $\mathbf{m}$, DRF: $(\text{latitude}, \text{longitude}), \mathbf{m}$)) for their local predictions that is not contained in the random sample for given spatial unit. FH and DRF estimates may thus be much closer to the ground truth (= true population values) than sample observations.

As a further validation check, we aggregated our municipality-level estimates of mean income and the at-risk-of-poverty rate to the county level and compared them with the corresponding direct estimates at that level. The latter are more precisely estimated because county-level sample sizes are sufficiently large. Note that the Gini coefficient cannot be aggregated from municipalities to counties because it is not decomposable into within-municipality and between-municipality components. The results presented in the supplementary appendix show that both the FH- and the DRF-estimates aggregate up similarly well to the empirical county values. Again, perfect agreement with the county-level direct estimates is not necessarily a reasonable target, because both FH and DRF incorporate external information not contained in the county-level samples (in particular, information from municipalities with similar characteristics; DRF additionally exploits spatial proximity via latitude and longitude).

7. Purging spatial income distributions of differences in spatial characteristics

As our final application, we address the problem of purging spatial income distributions of differences in observed characteristics in order to recover a *pure* spatial income structure—one that is independent of the fact that individuals and municipalities in different regions tend to differ in socio-economic composition. To achieve this, we fit a distributional random forest conditional on both location $(\text{latitude}, \text{longitude})$ and municipality/household characteristics $(\mathbf{m}, \mathbf{x})$, i.e.,

$$F^e(y|(\text{latitude}, \text{longitude}), \mathbf{m}, \mathbf{x}). \tag{17}$$

Here, $(\text{latitude}, \text{longitude})$ represent the coordinates of a geographical unit as before. The vector $\mathbf{m}$ contains the municipality variables described above, while $\mathbf{x}$ includes all household characteristics listed in table 2 (excluding regional indicators, whose information is now captured by latitude and longitude).

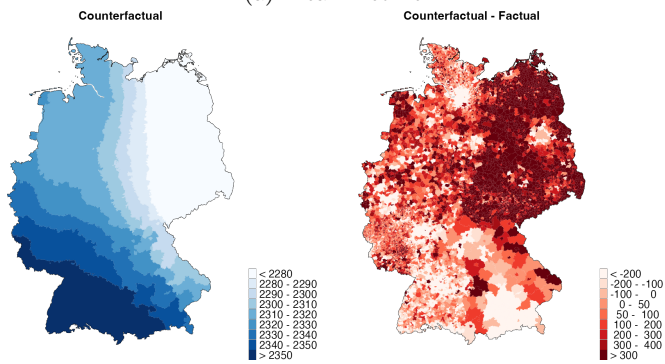
To construct local income distributions that do not depend on local composition in municipality and household characteristics, we consider

$$\begin{aligned} &F^c(y|(\text{latitude,longitude})) \\ &\qquad = \int_{\mathbf{m},\mathbf{x}} F^e(y|(\text{latitude}, \text{longitude}), \mathbf{m}, \mathbf{x}))\, dF_{\mathbf{m},\mathbf{x},\text{Germany}}(\mathbf{m}, \mathbf{x})), \end{aligned} \tag{18}$$

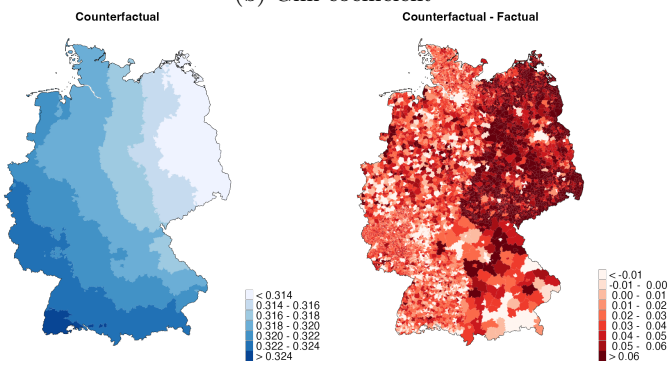
i.e., the local income distribution that would prevail in region $(\text{latitude}, \text{longitude})$ if municipality and household characteristics were distributed as in Germany as a whole.

Fig. 9: Purged spatial income distributions

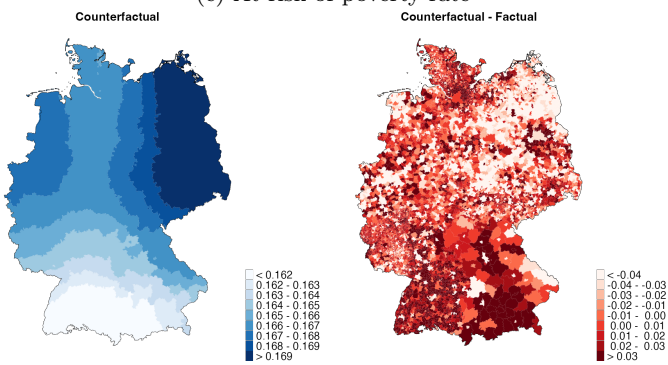
(a) Mean income



(b) Gini coefficient



(c) At-risk-of-poverty rate



Source: Microcensus 2005 and 2019. The maps on the right show differences between counterfactual and factual maps.

This results in informative maps, shown in figure 9. The results reveal a divide in mean income, inequality and poverty risk between East and West Germany, as well as between North and South. Under equalized composition across regions, mean

income and inequality are lower in the East than in the West, while poverty risk is higher in the East. However, these regional differences are substantially smaller than in the factual maps where regional characteristics are allowed to differ (Figure 4). This suggests that variations in characteristics $(\mathbf{m}, \mathbf{x})$ across regions explain disparities in income, inequality, and poverty risk to a very large extent.

8. Simulation evidence

In this section, we provide basic Monte Carlo evidence on the performance of the distributional random forest for estimating inequality and poverty indices for population subgroups. For our simulation, we use the following data generating process (DGP):

$$\begin{aligned}
&\mathbf{X} = (X_1, \ldots, X_5) \sim \text{Unif}([0,1])^5, \quad G_1 = 1(0.00 \leq X_1 < 0.02) \\
&G_2 = 1(0.02 \leq X_1 < 0.37), \quad G_3 = 1(0.37 \leq X_1 < 0.47) \\
&G_4 = 1(0.47 \leq X_1 < 0.52), \quad G_5 = 1(0.52 \leq X_1 \leq 1.00) \\
&Y \mid G_j = 1 \ \sim \ \Gamma(k_j, \theta_j), \qquad j = 1, \ldots, 5,
\end{aligned}$$

where $\Gamma(k, \theta)$ denotes the Gamma distribution with shape parameter k and scale parameter θ. An advantage of using the Gamma distribution for subgroups is that means, Ginis and other statistics can be easily computed using closed-form expressions. For example, if $Y \sim \Gamma(k, \theta)$, then $\text{mean}(Y) = k\theta$, $\text{skewness}(Y) = \frac{2}{\sqrt{k}}$ and $\text{Gini}(Y) = \Gamma(2k+1)/(2^{2k}\,\Gamma(k+1)^2)$ (McDonald and Jensen, 1979). The DGP is chosen such that the resulting unconditional mixture distribution resembled the Microcensus sample income distribution (in fact, its cdf lies between the 2005 and 2019 distributions shown in figure 3).

In order to assess the ability of the DRF to cope with interval-censored data, we carried out three different variants of simulations:

- *Variant 1:* The simulated outcome variable is artificially interval-censored using the income brackets in table 1, and the statistics of interest are estimated using grouped-data formulas. The corresponding 'true' values are also computed using the grouped-data formulas (7) to (10). Specifically, we use the true subgroup and unconditional CDFs to compute group masses f_j, which – together with interval widths and midpoints – enter the grouped formulas (7) to (10).
- *Variant 2:* The simulated outcome variable is artificially interval-censored using the income brackets in table 1, and the statistics of interest are estimated using grouped-data formulas. However, the 'true' values are computed from the continuous (un-grouped) DGP described above.
- *Variant 3:* The simulated outcome variable is continuous (i.e., not interval-censored). The statistics of interest are estimated using standard non-parametric formulas for the mean, the Gini coefficient, and the at-risk-of-poverty rate. The 'true' values are computed from the continuous (un-grouped) DGP described above.

Variant 1 is the most relevant, since the DRF can only target the true grouped values if it uses grouped-data formulas for estimation. However, using grouped versus un-grouped formulas turns out to make only small differences, both for the estimated statistics and for the corresponding 'true' values. This is already apparent in table 6 which reports both true values computed from grouped-data formulas (upper panel) and from un-grouped formulas (lower panel). We therefore report only *Variant 1* in the main text. The supplementary appendix further shows that grouped-data formulas also target the un-grouped true values well *(Variant 2)*. Observing continuous (i.e., not interval-censored) incomes would yield additional improvements *(Variant 3)*. This, however, is only a theoretical benchmark as long as the survey questionnaire elicits income in interval-censored form. Interval censoring imposes an information limit that cannot be overcome by statistical methods.

Table 6. True values for the data generating process

Group	Fraction	DGP $(Y \mid G_j)$	Mean	Median	Gini	POV
Based on Tille-Langel grouped-data formulas applied to DGP						
G_1	0.02	$\Gamma(2.5,\ 480)$	1202.1561	1046.7456	0.341974	0.436578
G_2	0.35	$\Gamma(3.0,\ 500)$	1502.5936	1338.7015	0.314629	0.290125
G_3	0.10	$\Gamma(6.0,\ 250)$	1502.4939	1419.6643	0.228572	0.179762
G_4	0.05	$\Gamma(5.0,\ 440)$	2204.1486	2057.8114	0.247994	0.066840
G_5	0.48	$\Gamma(4.0,\ 475)$	1903.4204	1746.8084	0.275370	0.139679
TOTAL	1.00	$\sum_{j=1}^{5} \pi_j\, \Gamma(k_j, \theta_j)$	1724.0495	1564.0870	0.293123	0.198639
Based on continuous distributions of DGP						
G_1	0.02	$\Gamma(2.5,\ 480)$	1200.0	1044.3504	0.339531	0.436869
G_2	0.35	$\Gamma(3.0,\ 500)$	1500.0	1337.0302	0.312500	0.289411
G_3	0.10	$\Gamma(6.0,\ 250)$	1500.0	1417.5403	0.225586	0.176983
G_4	0.05	$\Gamma(5.0,\ 440)$	2200.0	2055.1999	0.246094	0.065160
G_5	0.48	$\Gamma(4.0,\ 475)$	1900.0	1744.2289	0.273438	0.138073
TOTAL	1.00	$\sum_{j=1}^{5} \pi_j\, \Gamma(k_j, \theta_j)$	1721.0	1562.1164	0.291132	0.197263

Notes: TOTAL = Unconditional mixture distribution. POV = At-risk-of-poverty rate, defined as the share below 0.6 times the unconditional median. Entries in the upper panel are computed using the Tille-Langel grouped-data formulas (7) to (10) from bin probabilities f_j implied by the (true) subgroup and mixture CDFs over the prescribed brackets, together with class midpoints and widths. For the last class, probabilities use $(18{,}000, \infty)$, while midpoints and widths use $(18{,}000, 36{,}000)$.

We use $R = 1{,}000$ simulation runs for sample sizes $n \in \{10{,}000; 20{,}000; 50{,}000\}$.[8] For each simulation run, we fit a distributional random forest to estimate $F(Y|\mathbf{X}, G_1, G_2, G_3, G_4, G_5)$ and compute predictions at test points representing typical observations from group G_1 to G_5: $TP_j = (\overline{X}_{1j}, 0.2, 0.4, 0.6, 0.8, 1(G_1 =$

[8] The DRF is computationally intensive, so larger sample sizes were infeasible. Simulations for $n = 50{,}000$ observations already took several weeks to run on current server CPUs despite ten-fold parallelization.

Table 7. Monte Carlo summaries: grouped-data/estimation + grouped true values

Group	Bias	Variance	MAD	Coverage	Coverage_db
Number of observations $n = 10{,}000$					
Estimand: Gini					
G_1	-0.012912238	0.000163251	0.015292994	0.874	0.979
G_2	-0.001670334	9.48388e-05	0.007877185	0.980	0.982
G_3	0.012352705	0.000106787	0.013443234	0.863	0.973
G_4	0.013837017	9.38539e-05	0.014489594	0.835	0.984
G_5	-0.000597319	6.74549e-05	0.006596148	0.984	0.985
Estimand: POV					
G_1	-0.079102975	0.000977658	0.079263222	0.413	0.971
G_2	0.000479923	0.000401681	0.016021839	0.989	0.988
G_3	0.007038427	0.000465803	0.017684365	0.966	0.966
G_4	0.033065227	0.000204331	0.033099450	0.576	0.986
G_5	-0.001090276	0.000222033	0.011900292	0.985	0.987
Number of observations $n = 20{,}000$					
Estimand: Gini					
G_1	-0.008586657	0.000148229	0.012229024	0.926	0.980
G_2	-0.000533487	7.62437e-05	0.006942900	0.989	0.989
G_3	0.002284648	6.98827e-05	0.006818871	0.982	0.983
G_4	0.008717290	8.15121e-05	0.010319721	0.950	0.983
G_5	-0.000631842	5.83551e-05	0.006110186	0.989	0.989
Estimand: POV					
G_1	-0.058262690	0.001013685	0.059366700	0.628	0.979
G_2	0.001276598	0.000335516	0.014691413	0.990	0.989
G_3	0.002316055	0.000342308	0.014977382	0.975	0.973
G_4	0.020483056	0.000173558	0.021040500	0.882	0.974
G_5	-0.000739240	0.000183427	0.010836350	0.985	0.986
Number of observations $n = 50{,}000$					
Estimand: Gini					
G_1	-0.005939976	0.000113526	0.009954624	0.954	0.988
G_2	3.39493e-05	6.62390e-05	0.006453961	0.992	0.991
G_3	0.001389395	5.05173e-05	0.005801351	0.985	0.982
G_4	0.006883450	6.16488e-05	0.008547985	0.972	0.991
G_5	-0.000450358	4.21232e-05	0.005259946	0.999	0.999
Estimand: POV					
G_1	-0.044854370	0.000703256	0.046296276	0.713	0.974
G_2	0.001542728	0.000294849	0.013912465	0.996	0.994
G_3	0.001568460	0.000271013	0.012985097	0.984	0.983
G_4	0.015712105	0.000121902	0.016287574	0.933	0.988
G_5	-0.000514975	0.000137749	0.009450191	0.995	0.995

POV = At-risk-of-poverty rate, MAD = Mean Absolute Deviation, Coverage = fraction of cases in which 95% confidence interval covers true value, Coverage_db = coverage when bias is subtracted from confidence limits. $R = 1{,}000$ simulation runs.

$1), 1(G_2 = 1), 1(G_3 = 1), 1(G_4 = 1), 1(G_5 = 1))$, where $\overline{X}_{1j}$ is the mid-point of the X_1-interval defining $G_j = 1$ in the DGP. For example, the test point for a group $G_1 = 1$ observation is $TP_1 = (0.01, 0.2, 0.4, 0.6, 0.8, 1, 0, 0, 0, 0)$. For each simulation run, we fit a distributional random forest with the following paramaters: $N = B \cdot L = 50 \cdot 200 = 10,000$ trees, mtry=10, min.node.size = 5, sample.fraction = 0.5, alpha = 0.05, imbalance.penalty = 0.1. From the fitted DRF, we compute the mean, the Gini, and the at-risk-of-poverty rate using either the grouped-data formulas *(Variants 1 and 2)* or the usual non-parametric formulas *(Variant 3)*.

The results for bias, variance, mean absolute deviation in table 7 show clear improvements as the sample size increases. The remaining biases for 50,000 observations are non-zero but appear small relative to the true values reported in table 6. The coverage check for the 95 % confidence intervals indicate undercoverage for smaller sample sizes in individual cases, part of which is attributable to remaining biases (column 1 of table 7). This is illustrated by the improved coverage of bias-corrected confidence intervals (i.e., confidence intervals centered at the debiased point estimates). The performance of these biased-corrected intervals suggests that uncertainty estimation remains approximately intact, despite remaining estimation biases (leading to unfavourable coverage if the confidence intervals are centered around the biased estimates). As the sample size increases, the confidence intervals tend to overcover, which is not a serious concern in practice, as it leads to conservative inference. As always, it remains unclear to what extent Monte Carlo results are generalizable so that more research would be useful.

9. Conclusion

Our analysis demonstrates that distributional random forests are a powerful and versatile tool for analysing income distributions with minimal parametric assumptions. Once trained, they allow the analyst to estimate a wide range of distributional indices – quantiles, means, Gini coefficients, poverty rates, etc. – without requiring separate model specifications. They also easily handle grouped income information as present in our application. Using German Microcensus data, we illustrated four applications that are relevant for both researchers and policymakers: (i) estimating income distributions for granular population subgroups, (ii) analysing changes in inequality and poverty over time, (iii) small-area estimation of income distributions, and (iv) purging spatial income distributions of differences in household and municipality characteristics.

From these analyses, we obtain several insights into the German income distribution. First, the shape and location of income distributions vary markedly across granular population subgroups, and income growth is highly heterogeneous. Second, while average incomes increased between 2005 and 2019, income inequality and the at-risk-of-poverty rate also rose. However, these increases were driven almost entirely by compositional shifts—population ageing, changes in educational attainment, and a rising share of immigrants—rather than by diverging income trajectories within fixed population subgroups.

Our geographical analysis provides new insights into the spatial structure of the German income distribution. We characterize regions with high or low income and

inequality and show that geographical units with higher mean incomes also tend to exhibit higher inequality. We also document that income growth was uneven across regions: poorer areas experienced faster relative growth than wealthier ones, suggesting partial convergence in mean incomes across space. Finally, we find that a large share of the observed regional variation in income and inequality is attributable to differences in municipal and household characteristics. After accounting for these compositional differences, systematic 'pure' spatial disparities remain, but they are substantially less pronounced.

We acknowledge that our analysis is ambitious and, to some extent, exploratory. It aims to extract fine-grained information from a large dataset while imposing little structure. The distributional random forest method is non-parametric in nature, computationally intensive, and data-hungry. Our results are highly plausible and likely exploit the available information to an extent not achieved in previous research. Nevertheless, more evidence on the method's finite-sample properties is needed. While our simulation results are encouraging, future research should scrutinize potential biases more closely and evaluate performance across alternative datasets, varying levels of observation, and outcome measures. Developing debiased variants and improved inference procedures is a promising direction for future work.

10. Acknowledgments

We thank Sarah Bohnensteffen, Michael Knaus, Niko Muffler, Jeffrey Näf, Henri Pfleiderer, Klaus Pforr, Julie Schnaitmann, and participants of the 11th Mikrocensus User Conference at GESIS Mannheim for valuable comments and discussions. We are gratefule to Kerstin Stockmayer, Kristin Nowak, and Lukas Köhler at the Research Data Center of the Statistical Office of Baden-Württemberg for their continued support. Martin Biewen is a member of the Machine Learning Cluster of Excellence, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2064/1 - Project 390727645.

References

Athey, S., J. Tibshirani, S. Wager (2019). Generalized Random Forests. *Annals of Statistics*, *47*, 1148-1178.

Boehle, M. Armutsmessung mit dem Mikrozensus: Methodische Aspekte und Umsetzung für Querschnitts- und Trendanalysen. *GESIS Papers*, *2015/16*.

Biau, G., E. Scornet (2016). A random forest guided tour. *TEST*, *25*, 197-227.

Biewen, M., S. Jenkins (2005). A framework for the decomposition of poverty differences with an application to poverty differences between countries. *Empirical Economics*, *30*, 331-358.

Blanchet, T., T. Piketty (2022). Generalized Pareto Curves: Theory and Applications. *Review of Income and Wealth*, *68*, 263-288.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5-32.

Cevid, D., L. Michel, J. Näf, P. Bühlmann, N. Meinshausen (2022). Distributional random forests: Heterogeneity adjustment and multivariate distributional regression. *Journal of Machine Learning Research*, *23*(333), 1-79.

Chernozhukov, V., I. Fernandez-Val, B. Melly (2013). Inference on counterfactual distributions. *Econometrica*, *81*, 2205-2268.

Chotikapanich, D., W.E. Griffiths, G. Hajargasht, W. Karunarathne, D.S. Prasada Rao (2018). Using the GB2 Income Distribution. *Econometrics*, *6*, 1-24.

Cowell, F., E. Flachaire (2015). Statistical Methods for Distributional Analysis. In *Handbook of income distribution* (eds. A.B. Atkinson, F. Bourguignon), vol. 2., ch. 6, pp. 359-460. Amsterdam: Elsevier.

De Nicolo, S., M.R. Ferrante, S. Pacei (2024a). Small area estimation of inequality measures using mixtures of Beta. *Journal of the Royal Statistical Society, Series A*, *187*, 85–109.

De Nicolo, S., M.R. Ferrante, S. Pacei (2024b). Small-Sample Bias Correction of Inequality Estimators in Complex Surveys. *Journal of Official Statistics*, *40*, 238–261.

Donald, S.G., D.A. Green, H. Paarsch (2000). Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates. *Review of Economic Studies*, *67*, 609-633.

Fabrizi, E., M.R. Ferrante, C. Trivisano (2020). A functional approach to small area estimation of the relative median poverty gap. *Journal of the Royal Statistical Society, Series A*, *183*, 1273-1291.

Fay, R.E., R.A. Herriot (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, *74*, 269-277.

Federal Statistical Office of Germany (2024). `https://www.forschungsdatenzentrum.de/de/haushalte/mikrozensus`. DOI: 10.21242/12211.2019.00.00.3.1.0 (Mikrozensus 2019, SUF), DOI: 10.21242/12211.2019.00.00.1.1.1 (Mikrozensus 2019, onsite), DOI: 10.21242/12211.2005.00.00.3.1.0 (Mikrozensus 2005, SUF), DOI: 10.21242/12211.2005.00.00.1.1.0 (Mikrozensus 2005, onsite).

Fortin, N., T. Lemieux, S. Firpo (2011). Decomposition Methods in Economics In *Handbook of Labor Economics* (eds. D. Card, O. Ashenfelter), vol. 4a, ch. 1, pp. 1-102. Amsterdam: Elsevier.

Frieden, I., A. Peichl, P. Schüle (2023). Regional Income Inequality in Germany. *EconPol Forum*, *24*, pp. 50-55.

Garbasevschi, A., H. Tabenböck, P. Schüle, J. Baarck, P. Hufe, M. Wurm, A. Peichl (2023). Learning Income Levels and Inequality from Spatial and Sociodemographic Data in Germany. *Applied Geography*, *159*, 103058.

Gardini, A., E. Fabrizi, C. Trivisano (2022). Poverty and inequality mapping based on a unit-level log-normal mixture model. *Journal of the Royal Statistical Society, Series A*, *185*(4), 2073-2096.

Gretton, A., K. Borgwardt, M. Rasch, B. Schölkopf, A. Smola (2007). A kernel method for the two-sample problem. In *Advances in Neural Information Processing Systems*, vol. 19., pp. 513-520.

Hochgürtel, T. (2019). Einkommensanalysen mit dem Mikrozensus. *Wirtschaft und Statistik*, *3*, pp. 53-64.

Halbmeier, C., A.K. Kreutzmann, T. Schmid, C. Schröder (2019). The fayherriot command for estimating small-area indicators. *Stata Journal*, *19*, pp. 624-644.

Hothorn, T., B. Lausen, A. Benner, M. Radespiel-Tröger (2004). Bagging Survival Trees. *Statistics in Medicine*, *23*, pp. 77-91.

Hothorn, T., T. Kneib, P. Bühlmann (2013). Conditional transformation models. *Journal of the Royal Statistical Society, Series B*, *75*, pp. 1-24.

Hothorn, T., A. Zeileis (2021). Predictive Distribution Modeling Using Transformation Forests. *Journal of Computational and Graphical Statistics*, *30*, pp. 1181-1196.

Immel, L., A. Peichl (2020). Regionale Ungleichheit in Deutschland: Wo leben die Reichen und wo die Armen? *Ifo Schnelldienst*, *73*, 43-47.

INKAR (2024). Indikatoren und Karten zur Raum- und Stadtentwicklung (INKAR), `http://www.inkar.de`.

Jenkins, S.P., P. Van Kerm (2009). The measurement of economic inequality. In *The Oxford Handbook on Economic Inequality* (eds. W. Salverda, B. Nolan, T. Smeeding), ch. 3, pp. 40-67. Oxford University Press.

Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.

Krenmair, P., T. Schmid (2022). Flexible domain prediction using mixed effects random forests. *Journal of the Royal Statistical Society, Series C*, *71*, 1865-1894.

Lin, Y., Y. Jeon (2006). Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association*, *101*, pp. 578-590.

McDonald, J.B., B.C. Jensen (1979). An Analysis of Some Properties of Alternative Measures of Income Inequality Based on the Gamma Distribution Function. *Journal of the American Statistical Association*, *74*, pp. 856-860.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*, pp. 983-999.

Muandet, K., K. Fukumizu, B. Sriperumbudur, B. Schölkopf (2017). Kernel Mean Embedding of Distributions: A Review and Beyond. *Foundations and Trends in Machine Learning*, *10*, pp. 1-144.

Molina, I., P. Corral, M. Nguyen (2022). Estimation of poverty and inequality in small areas: review and discussion. *TEST*, *31*(4), 1143-1166.

Näf, J., C. Emmenegger, P. Bühlmann, N. Meinshausen (2023). Confidence and Uncertainty Assessment for Distributional Random Forests. *Journal of Machine Learning Research*, *24*, 1-77.

Rigby, R.A., D.M. Stasinopoulos (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, *54*, pp. 507-554.

Schluter, C., M. Trede (2024). Spatial earnings inequality. *Journal of Economic Inequality*, *22*, pp. 531-550.

Schlosser, L., T. Hothorn, R. Stauffer, A. Zeileis (2019). Distributional Regression Forests for Probabilistic Precipitation Forecasting in Complex Terrain. *Annals of Applied Statistics*, *13*, pp. 1564-1589.

Sugasawa, S., G. Kobayashi, Y. Kawakubo (2020). Estimation and inference for area-wise spatial income distributions from grouped data. *Computational Statistics and Data Analysis*, *145*, 106904.

Tille, Y., M. Langel (2012). Histogram-based interpolation of the Lorenz curve and Gini index for grouped data. *The American Statistician*, *66*, pp. 225-231.

Tzavidis, N., L.C. Zhang, A. Luna (2018). From start to finish: a framework for the production of small area official statistics. *Journal of the Royal Statistical Society,*

Series A, *181*, pp. 927-979.

Walter, P., M. Groß, T. Schmid, K. Weimer (2022). Iterative Kernel Density Estimation Applied to Grouped Data: Estimating Poverty and Inequality Indicators from the German Microcensus. *Journal of Official Statistics*, *38*, pp. 599-635.