# A Appendix

## A.1 Appendix: Auxiliary Tables and Figures

Table 2
*Overview Data Sets with Grouped Data*

| Data set | Country | Measure | No. of Brackets |
|---|---|---|---|
| European Social Survey (ESS) | Multiple Countries | Family Income | 10 |
| Microcensus (MZ) | Germany | Individual Income | 24 |
| Socio-Economic Panel Study (SOEP, 1988) | Germany | Household Wealth | 7-9 |
| Household Economic Survey (HES) | New Zealand | Investment Income | 18 |
| Current Population Survey (CPS) | USA | Family Income | 16 |
| General Social Survey (GSS) | USA | Family Income | 26 |
| | | Personal Earnings | 26 |
| American National Election Studies (ANES) | USA | Family Income | 22 |
| Annual American Time Use Survey (ATUS) | USA | Weekly Earnings | Only top coded |

Table 3
*MCIB and other approaches*

| Type | Method | Author(s) | Procedure | Pro's/Con's |
|---|---|---|---|---|
| *Non-Parametric* | Midpoint Estimation | Hout (2004) | $\text{midpoint}_b = (L_b + U_b)/2$ | Pro's: Easy to estimate<br>Con's: Does not address within-bin variance; no imputation options |
| | Robust Pareto Midpoint Estimator (RPME) | Von Hippel, Scarpino, and Holas (2016) | Similar to Midpoint Estimation, but assigns arithmetic mean of a Pareto distribution to top-bracket as well as constraints and varies the Pareto parameter $\alpha$ and provides the median, geometric and harmonic mean as alternative midpoints for the top bin | Pro's: Easy to estimate and robust to small sample sizes<br>Con's: Does not address within-bin variance; not very accurate if the number of bins is low; no imputation option |
| | CDF Interpolation | Von Hippel, Hunter, and Drown (2017) | Fits interpolated CDFs to grouped data (counts in grouped and continuous distribution are equal for each bin); top bin: Pareto, rectangular, or exponential | Pro's: Values of the fitted distribution represent the number of observations observed within each bin<br>Con's: no imputation option |
| | Random Empirical Distribution Imputation (REDI) | King (2022) | Performs cold-deck imputation from a same bin of reference data set to impute continuous measures | Pro's: Imputes real metric values for each observation from the grouped distribution<br>Con's: Can not be used without a suitable reference data set |
| *Parametric* | Multimodel Generalized Beta Estimator (MGBE) | McDonald and Ransom (1979), McDonald and Xu (1995), and Jenkins (2012) | Fits various generalized beta functions to the grouped data and then uses the distribution with the best fit or averages estimates across models | Pros: Does not assume one specific distribution to adequately fit the data<br>Cons: Slow estimation and only provides specific moments of the distribution (i.e., no imputation option) |
| | Mean Constrained Integration over Brackets (MCIB) | Jargowsky and Wheeler (2018) | Approximates densities (with varying density functions) across bins and integrates over bins to estimate moments of the distribution | Pros: Utilizes more information from the grouped distribution and addresses within-bin variance<br>Cons: Only estimates specific moments of the distribution |

Table 4
*Further MCIB estimations: net income*

| | stat | original | 4 brackets | diff | 8 brackets | diff | 12 brackets | diff | 16 brackets | diff | 20 brackets | diff | 24 brackets | diff |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 90 | mean | 2,654 | 2,808 | 5.77 | 2,790 | 5.12 | 2,830 | 6.63 | 2,811 | 5.90 | 2,792 | 5.19 | 2,763 | 4.09 |
| | | (7) | (10) | (0.46) | (9) | (0.45) | (9) | (0.45) | (10) | (0.45) | (8) | (0.41) | (8) | (0.45 ) |
| | sd | 1,748 | 2,525 | 44.46 | 2,211 | 26.48 | 2,347 | 34.30 | 2,237 | 27.98 | 2,168 | 24.05 | 2,043 | 16.91 |
| | | (7) | (19) | (1.22) | (15) | (0.97) | (18) | (1.17) | (17) | (1.14) | (15) | (0.95) | (12) | (0.86) |
| | median | 2,250 | 2,330 | 3.53 | 2,250 | 0.02 | 2,238 | -0.52 | 2,269 | 0.85 | 2,248 | -0.08 | 2,256 | 0.26 |
| | p1 | 320 | 52 | -10.09 | 76 | -9.17 | 94 | -8.51 | 104 | -8.12 | 130 | -7.17 | 188 | -4.96 |
| | p5 | 518 | 260 | -9.74 | 382 | -5.13 | 473 | -1.68 | 528 | 0.39 | 559 | 1.54 | 578 | 2.26 |
| | p10 | 820 | 521 | -11.25 | 763 | -2.14 | 857 | 1.41 | 832 | 0.44 | 856 | 1.37 | 822 | 0.09 |
| | p25 | 1,400 | 1,301 | -3.73 | 1,413 | 0.48 | 1,428 | 1.06 | 1,423 | 0.88 | 1,417 | 0.63 | 1,413 | 0.49 |
| | p75 | 3,500 | 3,525 | 0.93 | 3,576 | 2.85 | 3,602 | 3.84 | 3,539 | 1.46 | 3,503 | 0.11 | 3,602 | 3.84 |
| | p90 | 5,000 | 5,109 | 4.10 | 5,091 | 3.42 | 5,099 | 3.75 | 5,092 | 3.48 | 5,091 | 3.41 | 5,081 | 3.05 |
| | p95 | 6,000 | 6,989 | 37.26 | 6,659 | 24.81 | 6,840 | 31.65 | 6,720 | 27.13 | 6,630 | 23.74 | 6,452 | 17.02 |
| | p99 | 9,000 | 14,308 | 199.97 | 12,294 | 124.10 | 13,381 | 165.06 | 12,564 | 134.26 | 12,076 | 115.90 | 11,106 | 79.35 |
| | Correlation | | 0.79 | | 0.85 | | 0.84 | | 0.86 | | 0.87 | | 0.89 | |
| | | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | |
| 95 | mean | 2,654 | 2,684 | 1.11 | 2,724 | 2.60 | 2,698 | 1.66 | 2,698 | 1.63 | 2,699 | 1.69 | 2,780 | 4.74 |
| | | | (9) | (0.00) | (8) | (0.40) | (7) | (0.38) | (8) | (0.39) | (8) | (0.40) | (9) | (0.43) |
| | sd | 1,748 | 2,123 | 21.49 | 1,951 | 11.61 | 1,920 | 9.87 | 1,916 | 9.62 | 1,885 | 7.88 | 2,210 | 26.47 |
| | | | (14) | (0.00) | (10) | (0.75) | (10) | (0.75) | (11) | (0.75) | (10) | (0.72) | (18) | (1.14) |
| | median | 2,250 | 2,352 | 4.53 | 2,287 | 1.65 | 2,233 | -0.78 | 2,272 | 0.96 | 2,259 | 0.40 | 2,245 | -0.21 |
| | p1 | 320 | 50 | -10.17 | 76 | -9.17 | 90 | -8.68 | 107 | -8.02 | 105 | -8.10 | 169 | -5.71 |
| | p5 | 518 | 250 | -10.09 | 381 | -5.17 | 455 | -2.36 | 538 | 0.74 | 516 | -0.07 | 601 | 3.14 |
| | p10 | 820 | 499 | -12.09 | 762 | -2.17 | 855 | 1.30 | 827 | 0.26 | 824 | 0.14 | 844 | 0.89 |
| | p25 | 1,400 | 1,249 | -5.70 | 1,442 | 1.57 | 1,417 | 0.63 | 1,431 | 1.17 | 1,382 | -0.68 | 1,434 | 1.29 |
| | p75 | 3,500 | 3,566 | 2.49 | 3,523 | 0.87 | 3,538 | 1.42 | 3,529 | 1.08 | 3,514 | 0.52 | 3,623 | 4.62 |
| | p90 | 5,000 | 4,856 | -5.41 | 5,163 | 6.14 | 5,036 | 1.35 | 5,006 | 0.22 | 5,082 | 3.10 | 4,915 | -3.22 |
| | p95 | 6,000 | 6,247 | 9.29 | 6,200 | 7.52 | 6,193 | 7.25 | 6,201 | 7.57 | 6,187 | 7.05 | 6,286 | 10.76 |
| | p99 | 9,000 | 11,502 | 94.25 | 10,229 | 46.32 | 10,123 | 42.31 | 10,201 | 45.23 | 9,747 | 28.15 | 12,679 | 138.61 |
| | Correlation | | 0.86 | | 0.93 | | 0.94 | | 0.95 | | 0.95 | | 0.90 | |
| | | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | |
| 99 | mean | 2,654 | 2,856 | 7.58 | 2,700 | 1.71 | 2,709 | 2.04 | 2,694 | 1.49 | 2,668 | 0.50 | 2,700 | 1.71 |
| | | | (8) | (0.00) | (8) | (0.38) | (7) | (0.36) | (7) | (0.36) | (7) | (0.38) | (7) | (0.37) |
| | sd | 1,748 | 2,127 | 21.71 | 1,887 | 7.99 | 1,858 | 6.32 | 1,812 | 3.70 | 1,794 | 2.67 | 1,815 | 3.84 |
| | | | (8) | (0.00) | (9) | (0.68) | (8) | (0.63) | (8) | 0.60 | (9) | (0.65) | (8) | (0.63) |
| | median | 2,250 | 2,393 | 6.36 | 2,289 | 1.72 | 2,188 | -2.73 | 2,286 | 1.60 | 2,260 | 0.43 | 2,278 | 1.25 |
| | p1 | 320 | 51 | -10.14 | 70 | -9.42 | 91 | -8.63 | 104 | -8.13 | 108 | -8.00 | 175 | -5.45 |
| | p5 | 518 | 254 | -9.96 | 352 | -6.27 | 453 | -2.47 | 520 | 0.07 | 537 | 0.72 | 592 | 2.78 |
| | p10 | 820 | 505 | -11.86 | 701 | -4.47 | 859 | 1.47 | 868 | 1.81 | 823 | 0.11 | 840 | 0.76 |
| | p25 | 1,400 | 1,263 | -5.17 | 1,435 | 1.34 | 1,441 | 1.55 | 1,407 | 0.27 | 1,411 | 0.41 | 1,419 | 0.72 |
| | p75 | 3,500 | 4,020 | 19.58 | 3,529 | 1.11 | 3,572 | 2.71 | 3,586 | 3.23 | 3,599 | 3.75 | 3,568 | 2.56 |
| | p90 | 5,000 | 5,964 | 36.31 | 5,278 | 10.46 | 5,105 | 3.94 | 4,955 | -1.69 | 4,988 | -0.46 | 5,081 | 3.05 |
| | p95 | 6,000 | 6,988 | 37.23 | 6,534 | 20.11 | 6,415 | 15.62 | 6,148 | 5.57 | 6,042 | 1.57 | 6,193 | 7.27 |
| | p99 | 9,000 | 9,243 | 9.14 | 9,215 | 8.12 | 9,200 | 7.53 | 9,209 | 7.87 | 9,222 | 8.35 | 9,192 | 7.22 |
| | Correlation | | 0.80 | | 0.93 | | 0.97 | | 0.98 | | 0.98 | | 0.99 | |
| | | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | |

*Compiled by authors based on SOEP v37. The table provides several descriptive statistics of the net income distribution at the household level in 2017, with non-negative and non-zero values. The original distribution is taken directly from the SOEP data. For the MCIB distribution, the data are artificially grouped into 4,8,12,16,20, and 24 brackets with n upper cut-off limits set to the 90th, 95th, and 99th percentiles. Variables are winsorized at the 0.5th and the 99.5th percentiles, respectively. The difference in percent is mean-adjusted for the percentile estimates. Bootstrapped standard errors are in parentheses based on 500 bootstrap weights.*

Table 5
*Further MCIB estimations: Gross wealth*

| | stat | original | 4 brackets | diff | 8 brackets | diff | 12 brackets | diff | 16 brackets | diff | 20 brackets | diff | 24 brackets | diff |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 90 | mean | 271,144 | 254,660 | -6.08 | 256,154 | -5.53 | 257,430 | -5.06 | 258,549 | -4.65 | 259,555 | -4.27 | 258,872 | -4.53 |
| | | (2,167) | (1,892) | (1.00) | (1,785) | (1.00) | ( 1,875) | (1.07) | (1,845) | (0.93) | (1,869) | (1.04) | (1,724) | (0.98) |
| | sd | 416,787 | 355,916 | -14.60 | 353,276 | -15.24 | 350,500 | -15.90 | 352,597 | -15.40 | 356,341 | -14.50 | 352,671 | -15.38 |
| | | (5,360) | (3,769) | (1.39) | (3,854) | (1.38) | (3,821) | (1.47) | (3,705) | (1.32) | (3,954) | (1.49) | (3,557) | (1.36) |
| | median | 156,000 | 117,154 | -24.90 | 153,682 | -1.49 | 155,957 | -0.03 | 156,217 | 0.14 | 155,539 | -0.30 | 156,286 | 0.18 |
| | p1 | 400 | 1,152 | 0.28 | 599 | 0.07 | 470 | 0.03 | 505 | 0.04 | 470 | 0.03 | 408 | 0.00 |
| | p5 | 2,300 | 6,060 | 1.39 | 3,002 | 0.26 | 2,506 | 0.08 | 2,627 | 0.12 | 2,391 | 0.03 | 2,615 | 0.12 |
| | p10 | 5,000 | 12,144 | 2.63 | 5,957 | 0.35 | 5,215 | 0.08 | 5,483 | 0.18 | 5,380 | 0.14 | 5,140 | 0.05 |
| | p25 | 23,000 | 30,444 | 2.75 | 22,285 | -0.26 | 22,710 | -0.11 | 22,517 | -0.18 | 23,670 | 0.25 | 22,914 | -0.03 |
| | p75 | 350,000 | 359,418 | 3.47 | 348,992 | -0.37 | 349,294 | -0.26 | 348,701 | -0.48 | 348,985 | -0.37 | 350,073 | 0.03 |
| | p90 | 610,987 | 611,035 | 0.02 | 611,236 | 0.09 | 611,035 | 0.02 | 611,035 | 0.02 | 611,035 | 0.02 | 611,035 | 0.02 |
| | p95 | 900,000 | 858,816 | -15.19 | 862,207 | -13.94 | 853,804 | -17.04 | 859,379 | -14.98 | 863,346 | -13.52 | 856,021 | -16.22 |
| | p99 | 2,300,000 | 1,903,808 | -146.12 | 1,916,268 | -141.52 | 1,873,696 | -157.22 | 1,891,589 | -150.63 | 1,879,604 | -155.05 | 1,891,589 | -150.63 |
| | Correlation | | 0.71 | | 0.72 | | 0.71 | | 0.73 | | 0.71 | | 0.72 | |
| | | | (0.00) | | (0.01) | | (0.01) | | (0.01) | | (0.01) | | (0.01) | |
| 95 | mean | 271,144 | 266,127 | -1.85 | 263,372 | -2.87 | 262,943 | -3.02 | 262,715 | -3.11 | 262,038 | -3.36 | 263,865 | -2.68 |
| | | (2,167) | (2,016) | (0.00) | (1,997) | (1.10) | (1,992) | (1.01) | (1,925) | (1.06) | (1,849) | (1.04) | (1,883) | (1.03) |
| | sd | 416,787 | 377,320 | -9.47 | 372,996 | -10.51 | 372,168 | -10.71 | 369,335 | -11.39 | 364,718 | -12.49 | 374,756 | -10.08 |
| | | (5,360) | (4,137) | (0.00) | (4,231) | (1.56) | (4,187) | (1.41) | (4,098) | (1.50) | (3,845) | (1.48) | (4,182) | (1.51) |
| | median | 156,000 | 116,434 | -25.36 | 153,018 | -1.91 | 155,241 | -0.49 | 156,608 | 0.39 | 156,297 | 0.19 | 158,143 | 1.37 |
| | p1 | 400 | 1,425 | 0.38 | 608 | 0.08 | 550 | 0.06 | 527 | 0.05 | 465 | 0.02 | 370 | -0.01 |
| | p5 | 2,300 | 7,068 | 1.76 | 3,048 | 0.28 | 2,811 | 0.19 | 2,507 | 0.08 | 2,324 | 0.01 | 2,182 | -0.04 |
| | p10 | 5,000 | 14,211 | 3.40 | 6,044 | 0.38 | 5,706 | 0.26 | 5,383 | 0.14 | 5,118 | 0.04 | 5,380 | 0.14 |
| | p25 | 23,000 | 35,329 | 4.55 | 22,824 | -0.07 | 22,662 | -0.12 | 23,160 | 0.06 | 23,211 | 0.08 | 22,492 | -0.19 |
| | p75 | 350,000 | 384,360 | 12.67 | 355,237 | 1.93 | 350,726 | 0.27 | 348,173 | -0.67 | 351,177 | 0.43 | 349,668 | -0.12 |
| | p90 | 610,987 | 641,795 | 11.36 | 615,552 | 1.68 | 609,677 | -0.48 | 604,628 | -2.35 | 611,033 | 0.02 | 610,379 | -0.22 |
| | p95 | 900,000 | 902,976 | 1.10 | 902,976 | 1.10 | 904,168 | 1.54 | 903,572 | 1.32 | 903,572 | 1.32 | 902,976 | 1.10 |
| | p99 | 2,300,000 | 2,012,113 | -106.17 | 1,992,490 | -113.41 | 2,018,785 | -103.71 | 1,998,967 | -111.02 | 1,948,854 | -129.51 | 2,039,206 | -96.18 |
| | Correlation | | 0.79 | | 0.82 | | 0.83 | | 0.83 | | 0.84 | | 0.83 | |
| | | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | |
| 99 | mean | 271,144 | 392,610 | 44.80 | 308,097 | 13.63 | 286,454 | 5.65 | 279,900 | 3.23 | 275,587 | 1.64 | 274,098 | 1.09 |
| | | (2,167) | (2,845) | (0.00) | (2,489) | (1.25) | (2,222) | (1.15) | (2,160) | (1.11) | (2,152) | (1.14) | (2,049) | (1.12) |
| | sd | 416,787 | 560,663 | 34.52 | 470,641 | 12.92 | 438,157 | 5.13 | 425,791 | 2.16 | 418,098 | 0.31 | 413,861 | -0.70 |
| | | (5,360) | (3,552) | (0.00) | (4,362) | (1.81) | (4,713) | (1.75) | (4,724) | (1.71) | (4,825) | (1.75) | (4,666) | (1.69) |
| | median | 156,000 | 117,949 | -24.39 | 153,383 | -1.68 | 154,540 | -0.94 | 156,770 | 0.49 | 155,753 | -0.16 | 157,207 | 0.77 |
| | p1 | 400 | 1,455 | 0.39 | 583 | 0.07 | 557 | 0.06 | 459 | 0.02 | 523 | 0.05 | 448 | 0.02 |
| | p5 | 2,300 | 7,700 | 1.99 | 2,946 | 0.24 | 2,781 | 0.18 | 2,287 | 0.00 | 2,504 | 0.08 | 2,305 | 0.00 |
| | p10 | 5,000 | 15,350 | 3.82 | 5,873 | 0.32 | 5,686 | 0.25 | 5,185 | 0.07 | 5,293 | 0.11 | 5,299 | 0.11 |
| | p25 | 23,000 | 38,187 | 5.60 | 21,850 | -0.42 | 22,633 | -0.14 | 23,264 | 0.10 | 22,635 | -0.13 | 22,784 | -0.08 |
| | p75 | 350,000 | 574,400 | 82.76 | 357,395 | 2.73 | 349,779 | -0.08 | 349,721 | -0.10 | 350,081 | 0.03 | 347,865 | -0.79 |
| | p90 | 610,987 | 1,243,345 | 233.22 | 862,724 | 92.84 | 611,760 | 0.29 | 622,505 | 4.25 | 610,125 | -0.32 | 611,167 | 0.07 |
| | p95 | 900,000 | 1,601,356 | 258.67 | 1,338,008 | 161.54 | 1,173,164 | 100.75 | 1,072,275 | 63.54 | 990,202 | 33.27 | 931,063 | 11.46 |
| | p99 | 2,300,000 | 2,300,179 | 0.07 | 2,300,179 | 0.07 | 2,307,808 | 2.88 | 2,300,179 | 0.07 | 2,300,179 | 0.07 | 2,300,179 | 0.07 |
| | Correlation | | 0.72 | | 0.86 | | 0.91 | | 0.93 | | 0.95 | | 0.96 | |
| | | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | | (0.00) | |

*Compiled by authors based on SOEP v37. The table provides several descriptive statistics of the gross wealth distribution at the household level in 2017, with non-negative and non-zero values. The original distribution is taken directly from the SOEP data. For the MCIB distribution, the data are artificially grouped into 4,8,12,16,20, and 24 brackets with upper cut-off limits set to the 90th, 95th, and 99th percentiles. Variables are winsorized at the 0.5th and the 99.5th percentiles, respectively. The difference in percent is mean-adjusted for the percentile estimates. Bootstrapped standard errors are in parentheses based on 500 bootstrap weights.*

Table 6
*Original data compared to MCIB estimations: Reduced datasets*

| | net household income | | | | | | gross household wealth | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (orig. 50%) | (MCIB 50%) | (Δ%) | (orig. 20%) | (MCIB 20%) | (Δ%) | (orig. 50%) | (MCIB 50%) | (Δ%) | (orig. 20%) | (MCIB 20%) | (Δ%) |
| mean | 2,656 | 2,656 | 0.03 | 2,679 | 2,681 | 0.08 | 276,277 | 273,590 | -0.97 | 278,675 | 275,305 | -1.21 |
| | (10) | (11) | (0.55) | (16) | (17) | (0.85) | (3,014) | (2,776) | (1.51) | (5,063) | (4,656) | (2.44) |
| sd | 1,751 | 1,817 | 3.76 | 1,812 | 1,882 | 3.83 | 422,880 | 398,664 | -5.73 | 456,134 | 402,283 | -11.81 |
| | (11) | (12) | (0.92) | (19) | (22) | (1.59) | (7,502) | (6,198) | (2.32) | (14,755) | (10,426) | (3.66) |
| median | 2,260 | 2,248 | -0.51 | 2,200 | 2,212 | 0.55 | 159,999 | 157,618 | -1.49 | 164,000 | 161,320 | -1.63 |
| p1 | 320 | 72 | -9.35 | 320 | 90 | -8.60 | 300 | 529 | 0.08 | 400 | 552 | 0.05 |
| p5 | 530 | 375 | -5.85 | 500 | 384 | -4.34 | 2,240 | 2,744 | 0.18 | 2,588 | 2,901 | 0.11 |
| p10 | 828 | 769 | -2.21 | 828 | 763 | -2.41 | 5,000 | 5,566 | 0.20 | 5,500 | 5,615 | 0.04 |
| p25 | 1,400 | 1,393 | -0.28 | 1,400 | 1,411 | 0.41 | 23,000 | 23,561 | 0.20 | 24,996 | 26,023 | 0.37 |
| p75 | 3,500 | 3,519 | 0.71 | 3,500 | 3,580 | 2.98 | 352,120 | 348,645 | -1.26 | 351,238 | 347,782 | -1.24 |
| p90 | 5,000 | 4,938 | -2.32 | 5,000 | 5,010 | 0.37 | 625,000 | 672,060 | 17.03 | 609,000 | 666,480 | 20.63 |
| p95 | 6,000 | 6,042 | 1.59 | 6,000 | 6,108 | 4.05 | 919,100 | 942,779 | 8.57 | 887,668 | 947,242 | 21.38 |
| p99 | 9,000 | 9,310 | 11.66 | 9,500 | 9,820 | 11.94 | 2,380,500 | 2,127,405 | -91.61 | 2,670,000 | 2,298,713 | -133.23 |
| N | 9,850 | 9,850 | | 3,939 | 3,939 | | 6,383 | 6,383 | | 2,560 | 2,560 | |
| | | | | | | Inequality Measures | | | | | | |
| Mean Log Dev | 0.22 | 0.28 | 26.30 | 0.23 | 0.28 | 22.58 | 1.15 | 1.12 | -2.47 | 1.14 | 1.09 | -3.79 |
| | (0.00) | (0.00) | (1.58 ) | (0.00) | (0.00) | (2.52) | (0.01) | (0.01) | (1.28) | (0.02) | (0.02) | (2.05) |
| Theil | 0.20 | 0.22 | 9.92 | 0.21 | 0.22 | 9.05 | 0.73 | 0.71 | -2.76 | 0.75 | 0.71 | -6.49 |
| | (0.00) | (0.00) | (1.22) | (0.00) | (0.00) | (2.07) | (0.01) | (0.01) | (1.64) | (0.02) | (0.01) | (2.58) |
| Gini | 0.35 | 0.36 | 3.48 | 0.35 | 0.37 | 3.18 | 0.62 | 0.62 | -0.44 | 0.62 | 0.62 | -1.16 |
| | (0.00 ) | (0.00) | (0.56) | (0.00) | (0.00) | (0.94) | (0.00) | (0.00) | (0.69) | (0.01) | (0.00) | (1.09) |
| COV | 0.66 | 0.68 | 3.72 | 0.68 | 0.70 | 3.75 | 1.53 | 1.46 | -4.80 | 1.64 | 1.46 | -10.73 |
| | (0.00) | (0.00) | (0.70) | (0.01) | (0.01) | (1.25) | (0.02) | (0.01) | (1.48) | (0.03) | (0.02) | (2.27) |
| Rel. Mean Dev | 0.25 | 0.26 | 3.10 | 0.26 | 0.26 | 3.01 | 0.45 | 0.45 | -0.14 | 0.45 | 0.45 | -0.64 |
| | (0.00) | (0.00) | (0.59) | (0.00) | (0.00) | (0.97) | (0.00) | (0.00) | (0.78) | (0.00) | (0.00) | (1.25) |
| Correlation | | 0.96 | | | 0.96 | | | 0.91 | | | 0.90 | |
| | | (0.00) | | | (0.00) | | | (0.00) | | | (0.01) | |

*Compiled by authors based on SOEP v37. The table provides several moments of the net income and gross wealth distribution at the household level in 2017, with non-negative and non-zero values. The original distribution is taken directly from the SOEP data. For the MCIB distribution, the data are artificially grouped into 10 brackets with an upper cut-off limited to the 98th percentile. Variables are winsorized at the 0.5th and the 99.5th percentiles, respectively. The difference in percent is mean-adjusted for the percentile estimates. Bootstrapped standard errors are in parentheses, based on 500 bootstrap weights. The table provides the same setting as Table 1 in the main text, only with randomly reduced datasets with only 50 and 20 percent of the observations.*

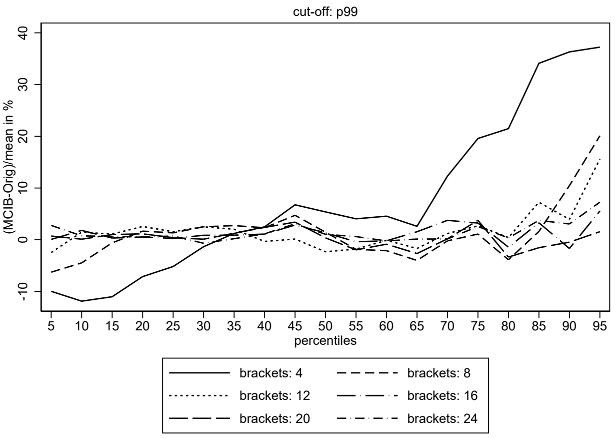


*Figure 7*. p99 cut-off: household net income. Compiled by authors based on SOEP v37. The panel shows the mean-adjusted difference in percent between the original distribution and the MCIB approximation (y-axis) along percentiles of households' net income (x-axis) in 2017. The panel depicts differences for various numbers of brackets. Variables are winsorized at the 0.5th and the 99.5th percentile, respectively.

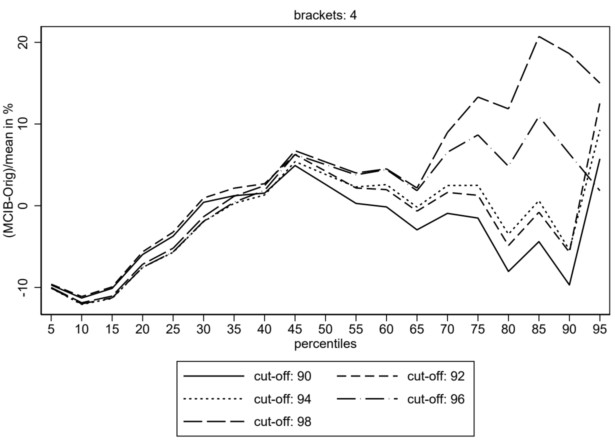


*Figure 8*. 4 brackets: household net income. Compiled by authors based on SOEP v37. The panel shows the mean-adjusted difference in percent between the original distribution and the MCIB approximation (y-axis) along percentiles of households' net income (x-axis) in 2017. The panel depicts differences for various cut-off limits. Variables are winsorized at the 0.5th and the 99.5th percentile, respectively.

Table 7

*Original data compared to MCIB estimations with 10 brackets with adjusted brackets settings*

| (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|
| gross household wealth | (orig.) | (MCIB) (orig. brackets) | (Δ%) | (MCIB) (adj. brackets) | (Δ%) |
| mean | 271,144 | 274,028 | 0.01 | 266,816 | - 0.02 |
| | (2,118) | (2,007) | (0.01) | (1,970) | (0.01) |
| sd | 416,787 | 401,723 | - 0.04 | 385,375 | - 0.08 |
| | (5,625) | (4,531) | (0.02) | (4,643) | (0.02) |
| median | 156,000 | 155,722 | 0.00 | 154,612 | - 0.01 |
| p1 | 400 | 533 | 0.00 | 538 | 0.00 |
| p5 | 2,300 | 2,728 | 0.00 | 2,915 | 0.00 |
| p10 | 5,000 | 5,420 | 0.00 | 6,100 | 0.00 |
| p25 | 23,000 | 23,042 | 0.00 | 227,66 | 0.00 |
| p75 | 350,000 | 346,950 | - 0.01 | 359,176 | 0.03 |
| p90 | 610,987 | 672,557 | 0.23 | 607,267 | - 0.01 |
| p95 | 900,000 | 997,407 | 0.36 | 884,713 | - 0.06 |
| p99 | 2,300,000 | 2,133,898 | - 0.61 | 2,113,580 | - 0.69 |
| N | 12,698 | 12,698 | | 12,698 | |
| Inequality Measures | | | | | |
| Mean Log Dev | 1.14 | 1.14 | - 0.01 | 1.10 | - 0.04 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Theil | 0.73 | 0.72 | - 0.01 | 0.70 | - 0.04 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Gini | 0.62 | 0.62 | 0.00 | 0.62 | - 0.01 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.01) |
| COV | 1.54 | 1.47 | - 0.05 | 1.44 | - 0.06 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Rel. Mean Dev | 0.45 | 0.46 | 0.01 | 0.45 | - 0.01 |
| | (0.00) | (0.00) | (0.01) | (0.00) | (0.01) |
| Correlation | | 0.91 | | 0.94 | |
| | | (0.00) | | (0.00) | |

*Compiled by authors based on SOEP v37. The table provides several moments of the gross wealth distribution at the household level in 2017, with non-negative and non-zero values. The original distribution is taken directly from the SOEP data. For the MCIB distribution, the data are artificially grouped in 10 brackets with an upper cut-off limited to the 98th percentile. Column (3) and (4) show the original approximation and column (5) and (6) show the estimates with adjusted bracket size. Variables are winsorized at the 0.5th and the 99.5th percentiles, respectively. The difference in percent is mean-adjusted for the percentile estimates. Bootstrapped standard errors are in parentheses, based on 500 bootstrap weights.*

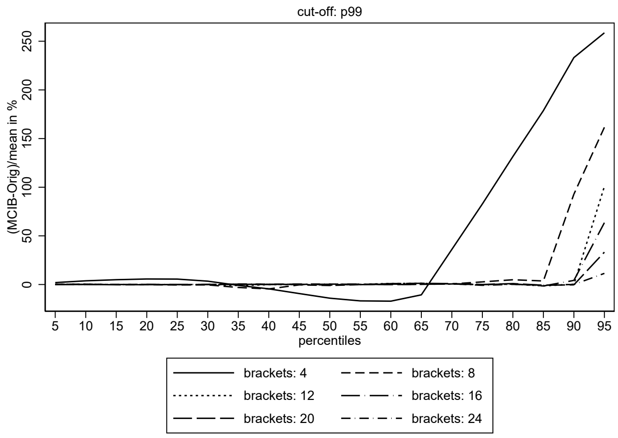


*Figure 9*. p99 cut-off: household gross wealth. Compiled by authors based on SOEP v37. The panel shows the mean-adjusted difference in percent between the original distribution and the MCIB approximation (y-axis) along percentiles of households' gross wealth (x-axis) in 2017. The panel depicts differences for various numbers of brackets. Variables are winsorized at the 0.5th and the 99.5th percentile, respectively.

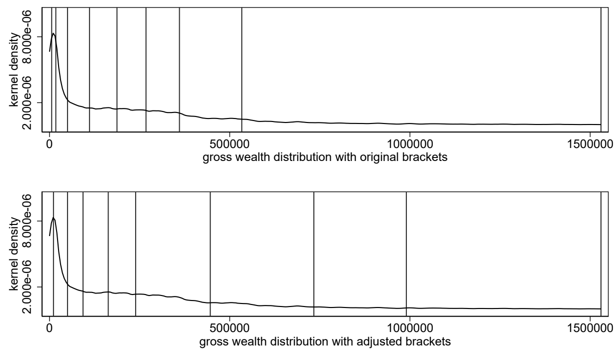


*Figure 11*. Kernel density plots with brackets settings. Compiled by authors based on SOEP v37. The panels show the brackets along the Epanechnikov kernel density function for the setting with 10 brackets (the last one is not depicted here) and a cut-off at the 98th percentile. The upper panel shows the original bracket setting. The lower panel shows a setting with approximately linear density functions within each bracket.

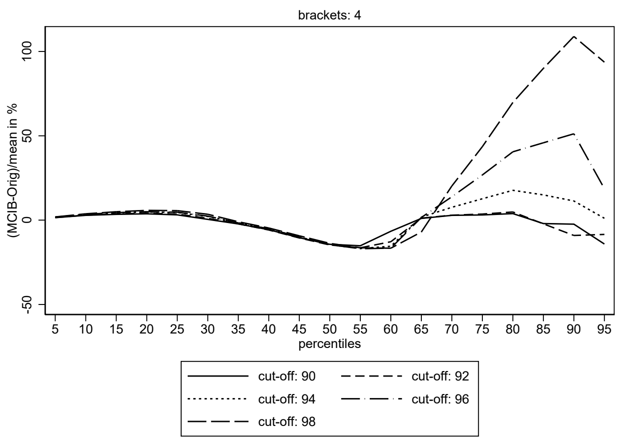


*Figure 10*. 4 brackets: household gross Wealth. Compiled by authors based on SOEP v37. The panel shows the mean-adjusted difference in percent between the original distribution and the MCIB approximation (y-axis) along percentiles of households' gross wealth (x-axis) in 2017. The panel depicts differences for various cut-off limits. Variables are winsorized at the 0.5th and the 99.5th percentile, respectively.

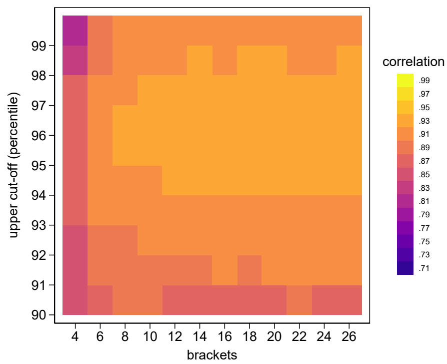


*Figure 12*. Correlation between original and simulated data by number of brackets and upper cut-off percentiles. Compiled by authors based on SOEP v37. The figure shows the correlation between the original distribution and the MCIB approximation based on equal width brackets. The y-axis depicts different upper cut-off limits; the x-axis different numbers of brackets. Variables are winsorized at the 0.5th and the 99.5th percentiles, respectively.

## A.2 Simulations

Our analysis relies on German survey data, but can our findings be generalized to other countries? We provide suggestive evidence that the results are relatively stable for different types of distributions and that the cut-off becomes more important if the distribution is more unequal. For the simulation, we take the following three steps: First, we estimate the Generalized Beta of the Second Kind (GB2) parameters $a$, $b$, $p$, $q$ based on the income distribution from our main analysis. This approach is based on Jenkins (2009), estimating the following probability density function using maximum likelihood:

$$f(y) = \frac{ay^{ap-1}}{b^{ap}B(p,q)[1+(y/b)^a]^{p+q}}, \quad (3)$$

where parameters $a, b, p, q$ are each postive, $B(p, q)$ represents the Beta function. Estimating the parameters on household income using the SOEP data, the parameters take the values $a = 2.46$, $b = 4048.35$, $p = .76$, $q = 1.64$.

Second, we manipulate the $a$ parameter to simulate different distributions, leaving $b$, $p$, $q$ constant. Increasing (decreasing) the $a$ parameter leads to lower (higher) levels of inequality, here measured by the Gini coefficient. We manipulate the $a$ parameter so that the simulated distributions exhibit Gini coefficients of 0.2, 0.4, and 0.6 (indicating low, medium and higher inequality). The corresponding $a$ parameters are provided in Table 8. The univariate kernel estimates of the distributions are provided in Figure 13.

Third, we ran the same methodology as in the main analysis on the simulated distributions. Using different brackets and cut-off specifications, we artificially group the data, and impute a metric distribution based on the grouped information. The correlations are provided in Figure 14.

The simulations show that our baseline rule from the main analysis, more than six brackets and an upper cut-off at the 95th holds for typical income distributions, i.e., simulations 1 and 2. A more equal distribution than in the basic case, as shown in the upper right panel of Figure 14, does not necessarily require fewer brackets to receive a similar level of correlation as in our baseline approach. However, the cut-off could be slightly lower for a more equal distribution providing slightly higher correlations for lower cut-offs. Interestingly, the smaller variance does not improve the approximation, and we observe a slightly smaller correlation for 10 to 18 brackets. Increasing the Gini coefficient to 0.4 (Simulation 2, lower left panel in Figure 14) also does not change the baseline rule. However, setting lower cut-offs leads to more imprecise estimates. Finally, simulation 3 (lower right panel in Figure 14) would require more brackets and a later cut-off to receive precise estimates. Nevertheless, a Gini coefficient of 0.6 is relatively high for an income distribution and points toward the baseline rule measured for the wealth distribution. We summarize from this simulation exercise that our baseline estimates seem to hold for typical levels of dispersion of income and, potentially, wealth and, as discussed in the main analysis, the more unequal the distribution is, the more important the right cut-off point.

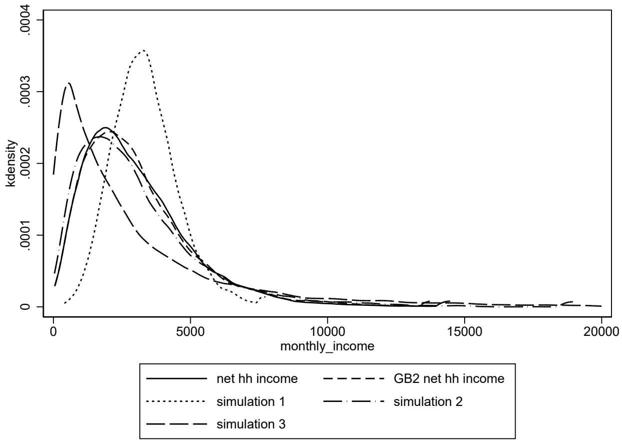


*Figure 13.* Net household income: kdensity functions. Compiled by authors based on SOEP v37. The figure shows the kdensity estimates of the net household income variable, its GB2 approximation, and three simulations. Estimates are based on the Epanechnikov function. For readability, the figure is cut-off at the x-axis at €20,000.

Table 8
*Simulations based on different a parameters from the GB2 estimates*

| Variable | GB2 $a$ estimate | Gini coefficient |
|---|---|---|
| net household income | n/a | 0.35 |
| GB2 net household income | 2.46 | 0.36 |
| GB2 Simulation 1 | 4.68 | 0.20 |
| GB2 Simulation 2 | 2.17 | 0.40 |
| GB2 Simulation 3 | 1.30 | 0.60 |

Compiled by authors based on SOEP v37. The table shows the estimated $a$ parameter based on the GB2 distribution if applicable, and the Gini coefficient.

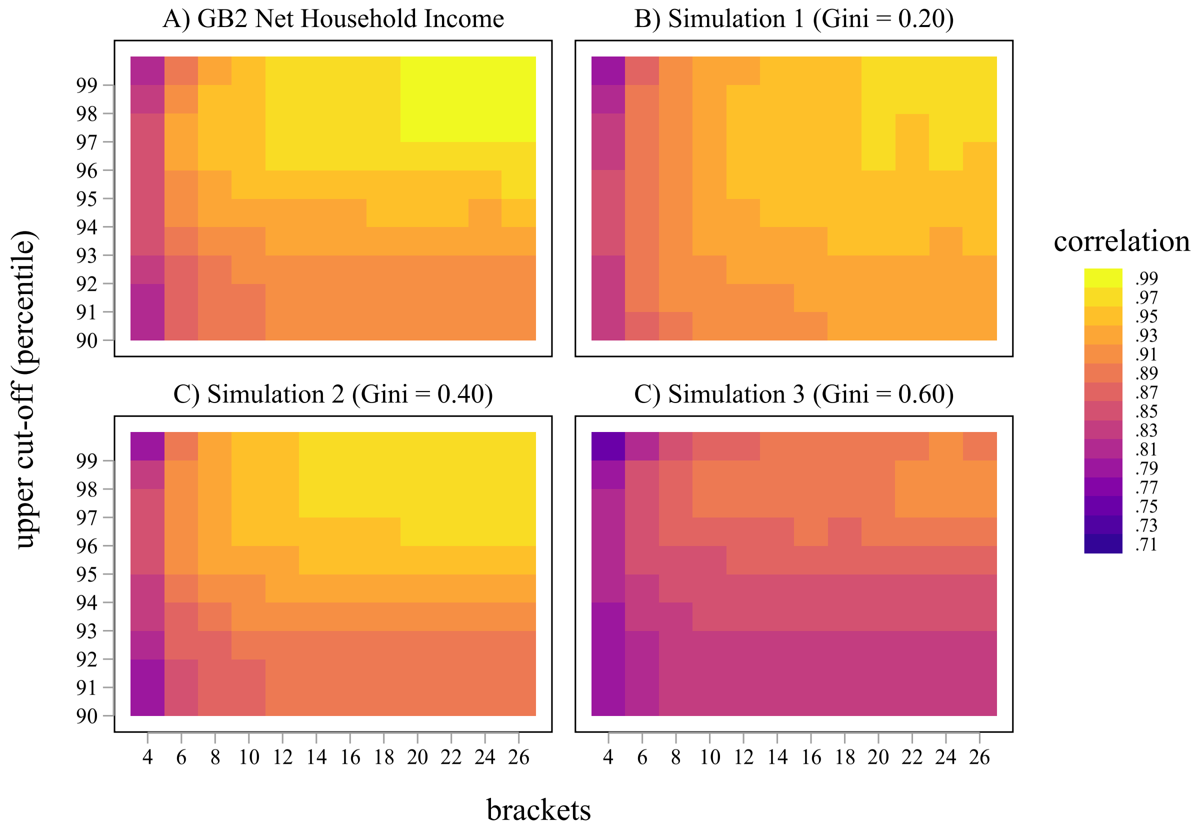


*Figure 14*. Correlation between original and simulated data by number of brackets and upper cut-off percentiles. Compiled by authors based on SOEP v37. The figure shows the correlation between the original distribution and the MCIB approximation for household net income and the three simulations. The y-axis depicts different upper cut-off limits; the x-axis different numbers of brackets. Variables are winsorized at the 0.5th and the 99.5th percentiles, respectively.