EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 12, No. 3, 2019, 689-708
ISSN 1307-5543 – www.ejpam.com
Published by New York Business Global

Decomposing the Quantile Ratio Index with Applications to Australian Income and Wealth Data

Luke A. Prendergast[1], Robert G. Staudte[1,*]

[1] *Department of Mathematics and Statistics, College of Science, Health and Engineering, La Trobe University, Melbourne, Australia*

Abstract. The quantile ratio index is a simple and effective measure of relative inequality for income data that is resistant to outliers. A useful property of this index is investigated here: given a partition of the income distribution into a union of sets of symmetric quantiles, one can find the inequality for each set and readily combine them in a weighted average to obtain the index for the entire population. When applied to data for various years, one can track how these contributions to inequality vary over time, as illustrated here for Australian Bureau of Statistics income and wealth data.

2010 Mathematics Subject Classifications: 62P20, 91B82

Key Words and Phrases: Confidence intervals, Gini index, inequality measures, quantile density, robust statistics

1. Introduction

It is desirable to break down a measure of inequality for a population of incomes into contributions to inequality from sub-populations. A natural partition of a population for the quantile ratio index (QRI) of [14] is provided by unions of symmetric quantiles. It allows one to determine how inequality in the middle half, for example, affects the QRI compared to how inequality between the smallest and largest quartiles does.

1.1. Background

Given the vast literature on inequality measures and many possible decompositions of them, it is reasonable to ask why should we start over in these endeavours with ratios of quantiles or averages of ratios of symmetric quantiles? The answer is two-fold: first, estimation of the traditional inequality measures is quite difficult. It has been shown by [5] that many inequality measures have unbounded influence functions, which leads their

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v12i3.3436

Email addresses: luke.prendergast@latrobe.edu.au (L. Prendergast), r.staudte@latrobe.edu.au (R. Staudte)

sample versions (estimators) to be overly sensitive to outliers, which are not uncommon in income data. Moreover, these estimators can be extremely slow to converge to normality (even when the population variance is finite), which undermines confidence intervals based on the fact of asymptotic normality, see [12, Fig.7]. Second, only recently have good inferential methods become available for ratios of quantiles. That is because confidence intervals for quantiles or ratios of quantiles require good estimates of the quantile density, recently found in [11]. These estimates are the basis for confidence intervals for ratios of quantiles, [13], quantile versions of the Lorenz curve and associated inequality measures in [12] and the quantile ratio index itself [14].

When we talk about 'decomposability' of inequality indices we usually mean an ANOVA type breakdown of the inequality index between and within subpopulations, [2, 4, 6, 8, 10, 15–17, and references therein]. The QRI has not yet been shown to satisfy such a decomposition, but readily lends itself to partitions of unions of symmetric quantiles. So, for example, while one might use the popular P90/P10 ratio of percentiles to compare large with small incomes, here we can apply the QRI to the union of the largest 20% of incomes with the smallest 20%, say, and also see how this contributes to the overall QRI for the entire population, as summarized in the weighted average (3). By estimating these components of the QRI for data in different years, we can see which portions of the population are changing over time, as we illustrate with Australian income and wealth data in Section 3.

To make these statements precise, we next formally introduce the QRI. Let F be the cumulative distribution function (cdf) describing a population of non-negative incomes with possible positive mass on zero $F(0) < 1/2$, and define the quantile function of F by $Q(p) = \inf\{x : F(x) \geq p\}$, $0 \leq p < 1$. Further define $Q(1) = \lim_{p\to 1} Q(p)$, which equals $+\infty$ if F has infinite support. Often we write x_p for $Q(p)$. Following [14], define the *ratio of symmetric quantiles* by $R(p) = Q(p/2)/Q(1-p/2) = x_{p/2}/x_{1-p/2}$ for $0 \leq p \leq 1$. A plot of $R(p)$ against p shows how the typical (median) income of those with the lowest $100p\%$ incomes, divided by typical (median) income of those with the highest $100p\%$ incomes varies with p.

Relative inequality in the population of incomes is measured by the *quantile ratio index*, the area above $R(p)$ and less than the horizontal line at one: $I = \int_0^1 \{1 - R(p)\}\, dp$. Each of the $(1 - R(p))$s is itself a measure of relative inequality; for example, $1 - R(0.2) = 1 - x_{0.1}/x_{0.9}$ is the ratio of percentiles P90/P10, after transformation to the unit interval so that larger values indicate more relative inequality. Thus I is a simple average of these relative inequality measures, one for each p, while most such measures, including the Gini index, are a ratio of two measures, concentration and scale. It is shown in [14] that $I = I(F)$ has a bounded influence function which explains the good robustness properties of its estimator $\widehat{I}$, defined below in Section 2. For extensive comparison of the Gini index with $\hat{I}$ and other outlier resistant measures of relative inequalty, see [12–14].

1.2. Derivation of the decomposition and examples

Partition the unit interval into symmetric unions of intervals as follows: given $K \geq 1$ and $0 = p_0 < p_1 < \cdots < p_{K-1} < p_K = 1/2$ define $A_k = [p_{k-1}, p_k) \cup (1 - p_k, 1 - p_{k-1}]$ for $k = 1, \ldots, K-1$ and let the last $A_K = [p_{K-1}, 1 - p_{K-1}]$. For our purposes it is useful to think of A_K as essentially the union of two intervals $[p_{K-1}, 1/2)$ and $(1/2, 1 - p_{K-1}]$, the central point 1/2 playing a trivial role in what follows. We call $\{A_1, \ldots, A_K\}$ a *symmetric K-partition* of [0,1]. When $K = 2$ for any $0 < p_1 < 1/2$ the symmetric 2-partition consists of $A_1 = [0, p_1) \cup (1 - p_1, 1]$ and $A_2 = [p_1, 1 - p_1]$; and in particular for $p_1 = 1/4$ we obtain the *quartile partition*: the set A_1 describes the outer quartiles while A_2 the inner quartiles or central half. Another example is the *quintile partition* obtained by taking $K = 3$ and $p_1 = 0.2$, $p_2 = 0.4$. This symmetric 3-partition consists of $A_1 = [0, 0.2) \cup (0.8, 1.0]$, $A_2 = [0.2, 0.4) \cup (0.6, 0.8]$ and $A_3 = [0.4, 0.6]$.

Next we derive the decomposition of I into a weighted sum of inequality contributions in the partitioning of F inherited from a symmetric K-partition. To this end, when X has cdf F, write $X \sim F$; and denote $U \sim \mathcal{U}[0,1]$ for U uniformly distributed on [0,1]. Fix k and let $Q\{A_k\}$ denote the image of A_k under $Q = F^{-1}$. Also introduce $w_k = \Pr(X \in Q\{A_k\}) = \Pr(U \in A_k) = 2(p_k - p_{k-1})$. For simplicity of notation in the coming paragraphs, temporarily let $a = p_{k-1}$ and $b = p_k$, so $A_k = [a, b) \cup (1-b, 1-a]$ and $c = 2(b - a) = w_k$. The conditional distribution of F given $X \in Q\{A_k\}$ has cdf F_k given by:

$$F_k(x) = \begin{cases} \frac{F(x)-a}{c}, & \text{for} \quad Q(a) \leq x < Q(b) \; ; \\ \frac{1}{2}, & \text{for} \quad Q(b) \leq x < Q(1-b) \; ; \\ \frac{1}{2} + \frac{F(x)-(1-b)}{c}, & \text{for} \quad Q(1-b) \leq x < Q(1-a) \; . \end{cases}$$

The quantile function Q_k of F_k can be obtained by solving for $x = Q_k(u)$ in the above expression to obtain:

$$Q_k(u) = \begin{cases} Q(a + cu), & \text{for} \quad 0 < u \leq 1/2 \; ; \\ Q(cu + 1 + a - 2b), & \text{for} \quad 1/2 < u < 1 \; . \end{cases}$$

Given $X \in Q\{A_k\}$, the *conditional quantile inequality curve* is defined for $0 < p < 1$ by

$$R_k(p) = \frac{Q_k(p/2)}{Q_k(1-p/2)} = \frac{Q(a + c\,p/2)}{Q(1 - a - c\,p/2)} \; . \tag{1}$$

Given $X \in Q\{A_k\}$, the *conditional QRI* is denoted I_k and determined by

$$1 - I_k = \int_0^1 R_k(p)\, dp = \int_0^1 \frac{Q(a + c\,p/2)}{Q(1 - a - c\,p/2)}\, dp = \frac{1}{c} \int_{2a}^{2b} R(u)\, du \; , \tag{2}$$

where we have made the change of variable $u = 2a + cp$. Rewriting $1 - I_k$ in terms of our earlier notation $a = p_{k-1}$, $b = p_k$ and $c = 2(b - a) = w_k$, multiplying both sides of equation (2) by w_k and summing over k leads to the weighted average:

$$I = \sum_k w_k I_k \; . \tag{3}$$

It is evident that $I_1 \geq \cdots \geq I_K$ because Q is non-decreasing. Moreover, we can interpret each I_k as the conditional inequality, given $X \in Q\{A_k\}$. The product $w_k I_k$ gives the amount that the kth partition member contributes to I.

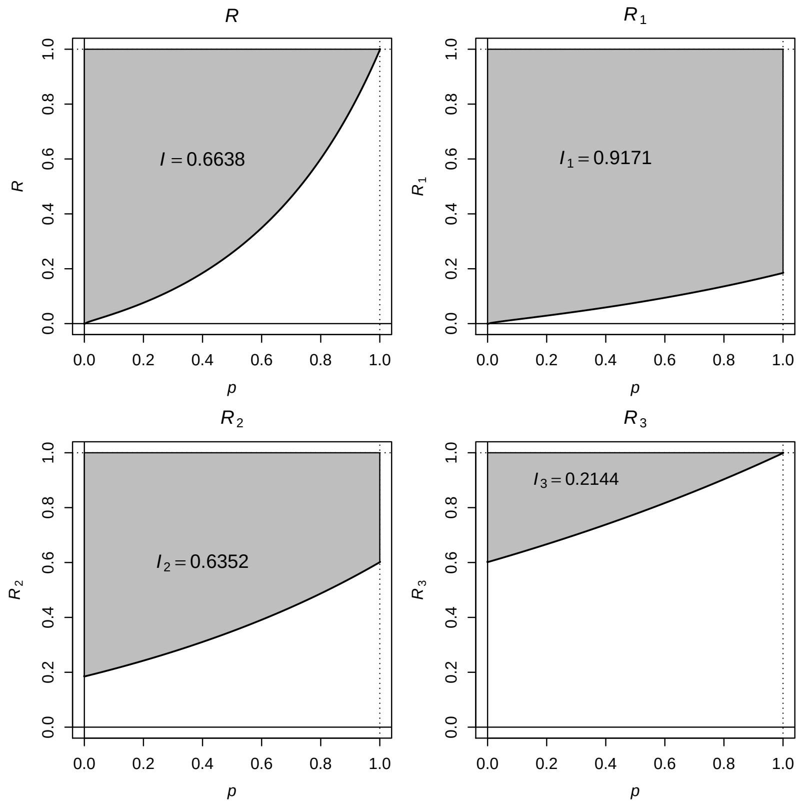

Figure 1: Plots of the quantile ratio $R(p) = Q(p/2)/Q(1-p/2)$ for the standard lognormal distribution F and, for the quintile partition, $R_k(p)$ defined by (1) for $k = 1, 2, 3$. The inequality measure I_k for each R_k is the shaded area above its graph and below the horizontal line at one.

1.3. Examples of symmetric K-partitions and inequality decompositions for the standard lognormal distribution

For F the standard lognormal distribution and the *quintile partition* defined by $K = 3$ with $p_1 = 0.2$, $p_2 = 0.4$, the graphs of R and R_1, R_2, R_3 are shown in Figure 1. The shaded areas above the inequality curves and below the horizontal lines at one are respectively $I = 0.6638$, $I_1 = 0.9171$, $I_2 = 0.6352$ and $I_3 = 0.2144$. By (3) $I = 0.6638 = 0.4 \times 0.9171 + 0.4 \times 0.6352 + 0.2 \times 0.2144$. For this last partition I_1 and I_2 each contribute 40% to the overall index I while the middle group contributes only 20%. (If one desires all partition members to contribute equally to I, one needs $p_1 = 1/6$ and $p_2 = 1/3$. Then I is the average of $I_1 = 0.9334$, $I_2 = 0.7325$ and $I_3 = 0.3254$.)

Next consider the *decile partition* $p_k = k/10$, $k = 1, 2, 3, 4$. It has $I = (I_1 + I_2 + \cdots +$

$I_5)/5 = (0.9619 + 0.8723 + 0.7376 + 0.5327 + 0.2144)/5$. From this decomposition, one can recover results for the coarser quintile partition obtained at the start of this example:

$$I = 0.4\left\{\frac{0.9619 + 0.8723}{2}\right\} + 0.4\left\{\frac{0.7376 + 0.5327}{2}\right\} + 0.2\left\{0.2144\right\} .$$

1.4. More examples of decompositions of I

In general the calculation of $\int_0^r R(p)\,dp$ requires numerical integration, but for the lognormal distribution closed form expressions for the components I_k of I are obtainable. The quantile function of the lognormal distribution with parameters μ, σ on the log-scale is $Q_{\mu,\sigma}(p) = \exp\{\mu + z_p\sigma\}$, where $z_p = \Phi^{-1}(p)$ is the p-quantile of the standard normal distribution having cdf Φ. Therefore $R_\sigma(p) = \exp\{2\sigma z_{p/2}\}$. Using elementary calculations relegated to Appendix A, one can show

$$\int_0^r R_\sigma(p)\,dp = 2\exp\{2\sigma^2\}\,\Phi\left\{\Phi^{-1}\left(\frac{r}{2}\right) - 2\sigma\right\} . \tag{4}$$

It follows immediately that $I = I(\sigma) = 1 - 2\exp\{2\sigma^2\}\,\Phi(-2\sigma)$. This QRI is monotone increasing from 0 to 1 as the shape parameter σ increases from 0 to ∞.

We can also find closed form expressions for the ingredients in the decomposition (3) using (4). For a K-partition A_k defined by $0 = p_0 < p_1 < \cdots < p_{K-1} < p_K = 1/2$ we have weights $w_k = 2(p_k - p_{k-1})$ and I_k determined by

$$1 - I_k(\sigma) = \frac{1}{w_k}\int_{2p_{k-1}}^{2p_k} R_\sigma(p)\,dp = \frac{2\exp\{2\sigma^2\}}{w_k}\left\{\Phi(z_{p_k} - 2\sigma) - \Phi(z_{p_{k-1}} - 2\sigma)\right\} . \tag{5}$$

In the top left of Figure 2 are shown the graphs of $I(\sigma)$, $I_1(\sigma)/2$ and $I_2(\sigma)/2$ for the quartile partition. The inequality $I_1(\sigma)$ rises rapidly to 1 for $\sigma < 2$. Therefore for $\sigma > 2$ nearly all the change in the index $I(\sigma)$ for the lognormal is due to that in the central half.

Quartile partition inequality graphs for the Type II Pareto distribution with shape parameter a are quite different, see the top right of Figure 2. In this case $I = I(a)$ is monotone decreasing in a from a high of 1 to its asymptote value of 0.7016. The contribution $I_1(a)/2$, inequality due to the outer quartiles is almost constant, decreasing from a high of 0.5 to a low of 0.4615 as a increases without bound. The contribution of inequality in the middle half of the population $I_2(a)/2$ descends from 0.5 to a low of 0.2401 with increasing a. Thus nearly all the change in inequality comes from the central half of the population. The asymptotic values are those belonging to the exponential distribution, which is the limit of the Pareto(a) distributions as $a \to \infty$.

The lower plots in Figures 2 exhibit different behavior from the previous plots, with I descending from 1 to 0 as the shape parameter increases. The contributions to inequality of the inner and outer quartiles in each case start at 0.5 and descend slowly to 0. For the symmetric Beta(α, α) family, the graphs (not shown) are similar in shape to those for the Gamma(α) family, but they descend to 0 faster as α increases.

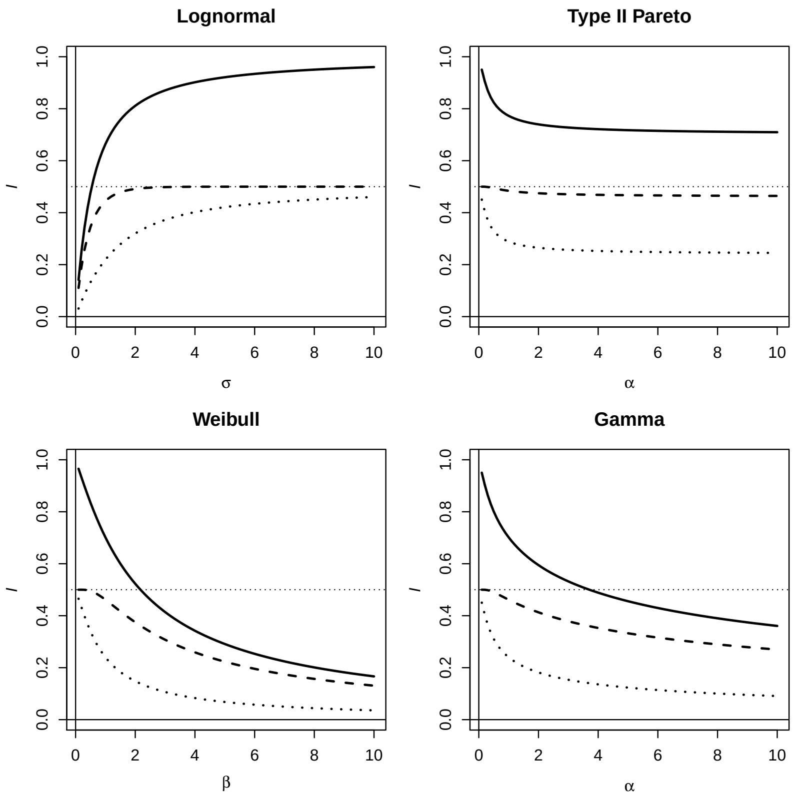

Figure 2: Plots in solid lines of the graphs of I versus the shape parameters for four standard families. For the quartile partition, $I = (I_1 + I_2)/2$, and the dashed lines show the respective graphs of $I_1/2$ defined by (2) and the dotted lines those of $I_2/2$.

Equi-K-partitions for large K.

In practice we are usually concerned only with small K-partitions. But what happens to the decomposition (3) if we fix F and take an *equi-K-partition* defined by $p_k = k/(2K)$, for $k = 0, 1, 2, \ldots, K$, and let $K \to \infty$? There are then K partition members with equal weights $w_k = 2(p_k - p_{k-1}) = 1/K$. The inequality in the kth subpopulation F_k contributes $w_k I_k = I_k/K$ to the sum in (3); and, for $k/K \to p$

$$1 - I_k = \frac{1}{w_k} \int_{2p_{k-1}}^{2p_k} R(u)\, du = K \int_{(k-1)/K}^{k/K} R(u)\, du \to R(p)\ .$$

Thus for a large equi-K-partition with $k/K \to p$, the conditional inequality I_k is approximately $1 - R(p)$.

1.5. Summary of results to follow

In the next Section 2 we explain how to find asymptotic standard errors and confidence intervals for the $\widehat{I}_k$s, and confirm by simulation studies the reliability of coverage probabilities for them for a wide range of possible income distributions. Applications of the symmetric decomposition theory and methodology to Australian income and wealth data are in Section 3. Details for two web-based applications that we have developed for readers to further study the IQR and to estimate the IQR for their own data are found in Section 4. Further applications and extensions are suggested in Section 5.

2. Inference for QRI component estimates

In this section we obtain large sample confidence intervals for the I_ks using methodology of [14, Section 3.2] to find such intervals for I. The basic idea there was to choose a positive integer J, define a grid $p_j = (j-1/2)/J$, $j = 1, \dots, J$ on the unit interval, and then estimate $I = \int_0^1 (1-R(p))\,dp$ by the average $\widehat{I}^{(J)} = \{\sum_{j=1}^{J}[1-\widehat{R}(p_j)\}/J$ where $\widehat{R}(p_j)$ is an estimate of the quantile ratio $R(p_j) = Q(p_j/2)/Q(1-p_j/2)$. Using standard results for the asymptotic normality and covariance structure of sample quantiles, nominal $100(1-\alpha)\%$ confidence intervals for I of the form $\widehat{I}^{(J)} \pm z_{1-\alpha/2}\{\widehat{\mathrm{Var}}(\widehat{I}^{(J)})\}^{1/2}$ are obtained. The required quantile function estimates are the continuous Type 8 estimates recommended by [9] and the quantile density estimates are those developed by [11]. These large sample intervals are shown to have good coverage probabilities for nearly all of the distributions listed below in Table 1 and n =100, 200, 500, 1000 and 5000. Further the choice of grid size $J = 100$ is large enough to obtain this good coverage. We have developed applications with the R software [R 7], see [14] to compute the standard errors and confidence intervals. For details see Section 4.

2.1. Estimating components of I

Given a symmetric K-partition with kth element $A_k = [p_{k-1}, p_k) \cup (1-p_k, 1-p_{k-1}]$, we found that for $w_k = 2(p_k - p_{k-1})$ the kth conditional inequality is determined by (2), which is simply $1 - I_k = \left\{\int_{2p_{k-1}}^{2p_k} R(u)\,du\right\}/w_k$.

We can estimate I_k by first estimating the integral of R over the interval $[2p_{k-1}, 2p_k]$. To this end define a grid on it by $p_{k_j} = 2p_{k-1} + w_k(j-1/2)/J$ for $j = 1, \dots, J$. (There does not seem to be any benefit in allowing J to depend on k.) Our estimate of I_k is then

$$\widehat{I}_k^{(J)} = \frac{1}{w_k\,J}\sum_{j=1}^{J}[1-\widehat{R}(p_{k_j})]\ , \tag{6}$$

where $\widehat{R}(p_{k_j})$ is an estimate of the quantile ratio $R(p_{k_j}) = Q(p_{k_j}/2)/Q(1-p_{k_j}/2)$. The nominal $100(1-\alpha)\%$ confidence interval for I_k is then

$$\widehat{I}_k^{(J)} \pm z_{1-\alpha/2}\{\widehat{\mathrm{Var}}(\widehat{I}_k^{(J)})\}^{1/2}\ . \tag{7}$$

Details of the formula for the asymptotic variance $\mathrm{Var}(\widehat{I}_k^{(J)})$ of $\widehat{I}_k^{(J)}$ and how to estimate it are essentially the same as those for $\widehat{I}^{(J)}$ found in [14, Appendix A] and so are omitted here.

Table 1: Simulated coverage probabilities for the quartile partition with sample sizes $n = 100, 500$ and 1000 for various choices of standard income distributions F, also studied in [14, Tables 1 and 3]. A total of 1000 trials were conducted.

		$n = 100$		$n = 500$		$n = 1000$	
#	F	I_1	I_2	I_1	I_2	I_1	I_2
1	Lognormal	0.966	0.960	0.970	0.968	0.955	0.957
2	Beta(0.1,0.1)	1.000	0.918	1.000	0.976	1.000	0.965
3	Beta(0.5,0.5)	0.930	0.926	0.940	0.941	0.952	0.927
4	Beta(1,1)	0.943	0.935	0.953	0.946	0.957	0.952
5	Beta(10,10)	0.969	0.972	0.976	0.960	0.972	0.961
6	χ^2_1	0.957	0.973	0.961	0.964	0.946	0.940
7	χ^2_4	0.966	0.959	0.963	0.955	0.960	0.958
8	χ^2_{25}	0.974	0.965	0.973	0.955	0.959	0.945
9	Pareto(1)	0.966	0.989	0.966	0.963	0.959	0.962
10	Pareto(2)	0.965	0.976	0.962	0.960	0.954	0.958
11	Pareto(100)	0.952	0.965	0.955	0.969	0.956	0.948
12	Exp(1)	0.956	0.963	0.945	0.960	0.949	0.962
13	Weibull(0.5)	0.968	0.993	0.967	0.969	0.962	0.969
14	Weibull(2)	0.959	0.961	0.966	0.953	0.960	0.957
15	Weibull(10)	0.980	0.983	0.985	0.955	0.987	0.960
16	LN-Frechet	0.983	0.970	0.973	0.968	0.968	0.959

Simulated coverage probabilities for the interval in (7) for varying sample sizes n from 16 different possible income distributions F are listed in Table 1 for the quartile partition. A total of 1000 trials were conducted for each choice of n and distribution and the nominal coverage was set to 0.95. With the exception of the extreme U-shaped Beta(0.1, 0.1) distribution, excellent coverages are achieved even for the smaller sample size $n = 100$ for estimation of both I_1 and I_2. Coverage tends to be slightly conservative and is typically closer to nominal for larger sample sizes. The simulations were repeated with $J = 50$ and $J = 200$ and similar results were obtained.

For the *quintile* partition, simulated coverage probabilities are presented in Table 2. These simulations and others not shown here, convince us that the interval estimators of the I_ks are reliable for a wide number of income distributions F.

An estimate of I itself can be obtained by applying the decomposition (9) to the $\hat{I}_k$s, but its standard error is not readily obtainable from the standard errors of the components, because although vector $\widehat{\boldsymbol{I}} = (\hat{I}_1, \ldots, \hat{I}_K)$ is asymptotically multivariate normal, its limiting covariance matrix $\lim_{n\to\infty} n^{1/2}\mathrm{Cov}(\widehat{\boldsymbol{I}})$ is not diagonal. Exact expressions for $\widehat{I}$ and subcomponents $\widehat{I}_k$ based on ordered data are presented in Appendix B.

Table 2: Simulated coverage probabilities for the quintile partition with sample sizes $n = 100, 500$ and 1000. A total of 1000 trials were conducted.

		$n=100$			$n=500$			$n=1000$		
#	F	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
1	Lognormal	0.976	0.965	0.964	0.966	0.958	0.959	0.968	0.969	0.943
2	Beta(0.1,0.1)	0.999	0.991	0.799	1.000	1.000	0.882	1.000	0.999	0.899
3	Beta(0.5,0.5)	0.926	0.934	0.935	0.937	0.937	0.941	0.944	0.933	0.925
4	Beta(1,1)	0.946	0.947	0.951	0.957	0.954	0.953	0.948	0.938	0.947
5	Beta(10,10)	0.955	0.968	0.961	0.984	0.960	0.948	0.975	0.970	0.961
6	χ^2_1	0.955	0.967	0.979	0.947	0.944	0.958	0.962	0.963	0.955
7	χ^2_4	0.973	0.962	0.959	0.953	0.945	0.943	0.971	0.953	0.952
8	χ^2_{25}	0.957	0.974	0.969	0.978	0.954	0.955	0.971	0.960	0.959
9	Pareto(1)	0.968	0.982	0.987	0.967	0.962	0.959	0.962	0.958	0.966
10	Pareto(2)	0.949	0.966	0.972	0.950	0.948	0.955	0.962	0.968	0.955
11	Pareto(100)	0.955	0.968	0.958	0.958	0.962	0.949	0.948	0.954	0.950
12	Exp(1)	0.944	0.965	0.971	0.960	0.952	0.947	0.942	0.948	0.952
13	Weibull(0.5)	0.969	0.989	0.996	0.964	0.966	0.973	0.956	0.969	0.958
14	Weibull(2)	0.966	0.960	0.962	0.971	0.940	0.940	0.956	0.958	0.954
15	Weibull(10)	0.973	0.990	0.961	0.985	0.971	0.954	0.979	0.957	0.955
16	LN-Frechet	0.985	0.979	0.980	0.977	0.974	0.967	0.971	0.964	0.953

3. Applications of QRI decompositions

3.1. Example 1: Australian disposable weekly income

Measuring household and personal weekly income is a complicated task carried out by governmental departments, including the Australian Bureau of Statistics (ABS), whose reports are available at [1]. The gross household income per week is published, but households differ so much in size that the *equivalized* disposal weekly income (DWI) is also found. The ABS defines the DWI as '... the amount of disposable cash income that a single person household would require to maintain the same standard of living as the household in question, regardless of the size or composition of the latter.'

Table 3 provides ABS grouped data on DWI for selected years, based on representative samples of households converted to 2014 dollar values. Figure 3 is our depiction of these data with kernel density plots. They are multi-modal distributions and reveal a clear shift to the right of DWI values over the period 2004–2014 (the solid line with the highest mode is for 2004). We are interested in tracking inequality over this period.

A density plot of the DWI data constructed from Table 3 and a plot of some of the quantiles over time, where the quantiles are relative to what the corresponding quantile was in 2004 (e.g. the 0.9 quantiles for 2006, 2008 etc. are each divided by the 2004 0.9 quantile) for clearer interpretation, are provided in Figure 3. The latter plot suggests increasing disparity in DWI up until 2010 between the wealthier and poorer sections of the population.

To understand how we constructed the plots in Figure 3, we need to examine Table 3 in some detail. For a partition of 29 classes of dollar incomes listed in the left-most column, and selected financial years, each table entry gives the estimated number of DWIs (in

Table 3: Equivalized disposable weekly income (DWI) in Australian dollars, adjusted for inflation to 2013-2014 dollars, for selected financial years. The tabled entries represent thousands of persons. Source: [1, Table 1.3], downloaded 27 July, 2017.

	2003–2004	2005–2006	2008–2010	2011–2012	2013–2014
$[0, 0]$ [1]	87.3	73.7	89.0	87.4	86.4
$[1, 49]$	94.1	90.1	95.8	81.6	95.3
$[50, 99]$	49.7	63.1	61.3	85.3	78.9
$[100, 149]$	94.0	66.2	84.0	92.3	47.6
$[150, 199]$	129.9	108.6	125.1	107.3	134.9
$[200, 249]$	273.7	219.6	164.7	185.6	151.1
$[250, 299]$	657.6	443.7	351.5	335.0	373.4
$[300, 349]$	1385.5	1152.0	596.3	373.9	397.9
$[350, 399]$	1301.8	1187.5	1195.8	913.3	636.7
$[400, 449]$	1231.7	1111.8	1172.4	1184.1	1135.2
$[450, 499]$	1093.7	1052.3	933.4	1044.7	1175.2
$[500, 549]$	1043.0	1097.4	991.3	1019.7	1171.7
$[550, 599]$	1092.2	1057.0	1009.7	980.8	1093.0
$[600, 649]$	1087.5	1016.2	1046.4	926.3	956.6
$[650, 699]$	1083.5	1066.9	987.0	1021.9	972.7
$[700, 749]$	1092.8	1023.3	996.9	999.2	938.9
$[750, 799]$	959.9	834.1	1037.1	1038.1	1009.6
$[800, 849]$	878.1	940.4	829.3	989.4	1013.4
$[850, 899]$	718.3	828.5	806.5	959.7	1099.5
$[900, 949]$	612.2	746.6	793.0	896.4	826.2
$[950, 999]$	631.8	731.9	757.8	714.9	885.6
$[1000, 1049]$	506.8	547.5	630.3	690.1	692.6
$[1050, 1099]$	492.3	515.3	730.8	803.1	695.8
$[1100, 1199]$	750.3	933.9	1118.5	1245.7	1379.5
$[1200, 1299]$	529.4	674.2	906.1	985.3	1027.2
$[1300, 1499]$	706.4	863.9	1400.8	1499.2	1447.8
$[1500, 1699]$	387.9	469.6	889.7	995.4	938.5
$[1700, 1999]$	263.2	427.0	682.8	850.2	862.3
$[2000, +\infty)$ [2]	371.9	588.4	1106.3	1082.9	1355.6
Total	19,606.5	19,930.7	21,589.6	22,188.8	22,679.1

[1] Some DWIs are negative, but these values have been rounded up to zero.
[2] An upper bound on DWIs greater than $2000 is not reported.

thousands). All amounts have been converted to 2013-2014 dollars using the Consumer Price Index. The financial year in Australia ranges from 1 July of one year to 30 June of the next; for simplicity we hereafter write '2004' for '2003-2004', and similarly for other years.. For the 2004 data the first entry tells us that there were (approximately) 87,300 zero DWIs, the second entry that there were 94,100 DWIs between 1 and 49 dollars, and for the last class 371,900 DWIs of 2000 dollars or more. The last class has lower bound $x_q = 2000$, where $q = 1 - 379.1/19606.5 = 0.981$.

Lacking the individual data, we created a population to take samples from, which allowed for the ambiguity of missing data in the right hand tail, which is often modelled by the Pareto distribution. We did this by generating 873 zeros, 941 random uniform

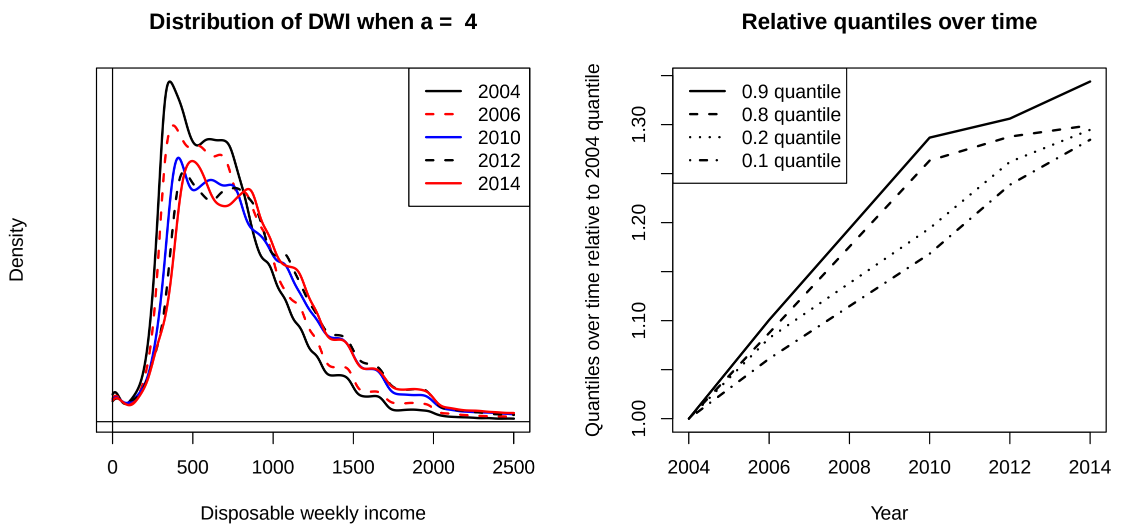

Figure 3: Kernel density estimates and relative quantile plots over time for the five populations generated in Section 3.1, truncated to [0,2500]. The small positive mass on 0 is smoothed out by these density estimates.

values from 1 to 49, 497 uniform values from 55 to 99, and so on. For the last category we generated 3,719 random Pareto(a, λ) values as follows: first, for a given shape parameter $a > 0$, we computed the scale parameter $\lambda = x_q/\{(1-q)^{-1/a}-1\} = 2000/\{(1-0.981)^{-1/a}-1\}$; secondly, we generated 3,719 uniform values u_i from [0.981,1]; and thirdly, we applied the quantile function to these values $Q_{a,\lambda}(u_i) = \lambda\{(1-u_i)^{-1/a} - 1\}$. As we will see, the choice of a does not affect the QRI much, although it would significantly affect most inequality measures.

Standard kernel density plots for the five populations were generated in this way using the default `density` command on R [7], one for each of the selected years, and these are shown in Figure 3 when the choice of Pareto shape parameter $a = 4$. They are truncated at income 2500, although their maxima can be much larger, as shown in Table 4. It is evident that the distributions are moving to the right. In fact the distribution of 1996 DWI (not shown) is unimodal with mode near 500, while these populations have two or more modes. The 2014 density plot is very similar to Graph 2 of the section 'Household Income and Wealth Distribution' [1].

The ABS also provides the relative standard error (RSE), defined as standard error of estimate divided by the estimate, for most of their results. For example, in the same source [1, Table 1.3] from which our Table 3 was excerpted, they give the RSE = 11.8% for the first table entry 87.3 (thousands of persons with zero DWI in 2004); that is, the standard error is nearly 10.3 thousand. For later years where the sample number of households was much higher, such as in 2014, the RSE's are in the range 5–8 %. The main point is that even with the best of survey methods, the resulting summary data listed in the ABS Table 1.3 only approximately describes the exact populations of DWI. Our five populations, truncated at \$2500 to fit in Figure 3, are also approximations.

Table 4: Percentiles for five distributions of DWI depicted in Figure 3, and values of $\widehat{I}$ rounded to two places. The minimum value for each distribution was 0.

	P05	P10	P20	P25	P50	P75	P80	P90	P95	max	$\widehat{I}$
2004	269	320	394	433	658	928	1008	1255	1521	33520	0.51
2006	292	340	426	472	707	1003	1096	1383	1714	39441	0.51
2010	309	374	470	526	793	1163	1273	1615	2024	39666	0.52
2012	317	396	497	552	831	1188	1298	1642	1989	34226	0.52
2014	321	411	509	558	843	1196	1309	1688	2179	32970	0.52

The percentiles of our five populations are listed in Table 4. They are in good agreement with the percentile estimates of [1, Table 1.1]. For example, for the year 2004 they obtain P10=324, P20=395, P50=657, P80=1,008 and P90=1,255. And for 2014 they obtain P10=415, P20=511, P50=844, P80=1308 and P90=1688.

The maximum values in the second-last column of Table 4 could have been quite different, (and much larger for a near 1). However, extremely large incomes (outliers) do not affect the quantile ratio index estimator $\widehat{I}$ very much. Note that $\widehat{I}$, also listed in Table 4, does not appear to be changing over the years 2004–2014, indicating a stable level of inequality for this period. Changing the Pareto tail shape a from 4 to 1 greatly increases the maximum values in this table, but has no effect on the other quantiles and little effect on $\hat{I}$. This is because a small percentage of extreme outliers do not affect the index, because their magnitudes contribute little after transformation to the [0,1] scale (e.g, note that the ratio of a small quantile and an a very large quantile due to extreme outliers is approximately zero, and increasing the large quantile only makes this closer to zero). Robustness properties of the QRI are discussed in more detail in [14].

Quartile partition estimates for DWI.

Next consider the 'quartile partition' with members $A_1 = [0, 0.25] \cup [0.75, 1]$ and $A_2 = [0.25, 0.75]$. Incomes in A_2 can be considered as belonging to the 'middle class'. Here $\widehat{I}_k$ is defined by (6) for $k = 1, 2$. These estimates, as well as those for I were based on samples of size $n = 10,000$ for each of the five populations generated with $a = 4$ and are listed in Table 5.

Table 5: Estimates of I and I_k for the quartile partition of the five distributions of Figure 4, based on samples of size $n = 10,000$. In parentheses are the values of $\sqrt{n}\ \widehat{\mathsf{SE}}[\widehat{I}_k]$.

	2004	2006	2010	2012	2014
$\widehat{I}_1$	0.721(0.27)	0.722(0.27)	0.742(0.27)	0.733(0.28)	0.736(0.28)
$\widehat{I}_2$	0.285(0.34)	0.289(0.33)	0.297(0.34)	0.285(0.34)	0.287(0.34)
$\widehat{I}$	0.503(0.27)	0.506(0.26)	0.520(0.27)	0.509(0.27)	0.512(0.27)

The bottom row shows the estimates of I are only increasing slightly over time. We have already commented upon this stability earlier for the right-most column of Table 4. However, a two-sided level 0.05 test for a difference between the 2004 and 2014 values of $\widehat{I}$ is just significant. We can now find the source of this increasing inequality.

The middle row of Table 5 shows that inequality in the middle class, measured by $\widehat{I}_2$, has not significantly changed over this period. But the inequality for partition A_1, containing the lower quartile and upper quartiles, has increased from 0.721 to 0.736 which is significant at the 0.05 level . Further, inequality for A_1 peaks at 0.742 in 2010, which was suggested in the righthand plot of Figure 3 where disparity grew at an increasing rate up until this time. It is interesting that the standard errors of $\widehat{I}_1$ and $\widehat{I}$ are roughly the same, while those of $\widehat{I}_2$ for the middle class are about 1/4 larger. Similar results were obtained for samples based on $n = 4000$ observations, but for $n = 1000$ changes in inequality over this time period were not quite statistically significant at the 0.05 level.

3.2. Example 2: Australian wealth data

In its explanatory notes of Manual 6523.0, the ABS defines household wealth by 'Net worth, often is the value of a household's assets less the value of its liabilities.' and then goes on to explain what it means by assets and liabilities.

Table 6 is obtained from [1, Table 2.3]. Because of the large numbers involved, we have written the dollar classes in thousands of dollars. For example, the second entry in the column labelled 2004 is 1098.9, which means that there were an estimated 1,098,900 households in that year whose net wealth was between 0 and $50,000 in 2004.

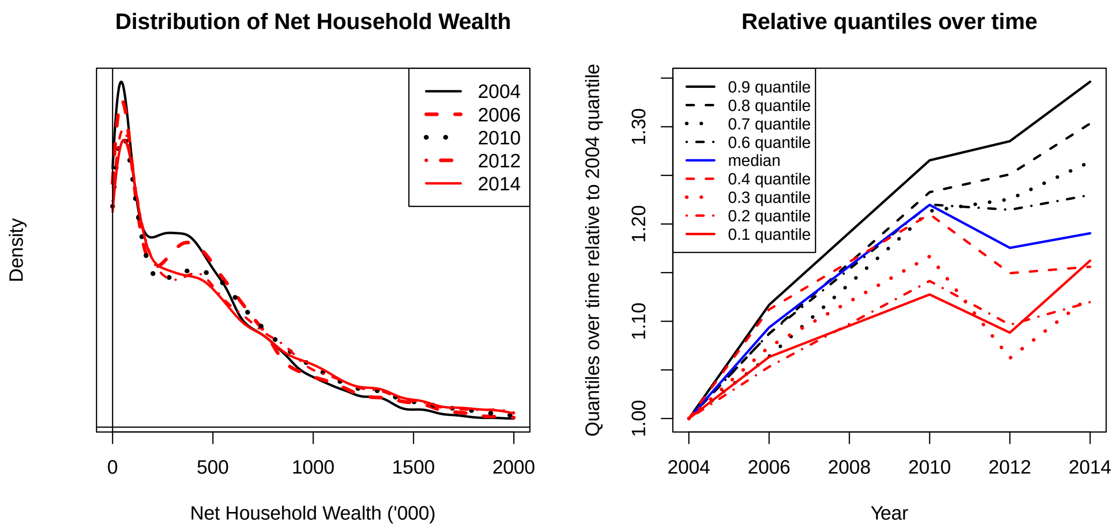

Figure 4: Kernel density estimates and relative quantile plots over time for the five populations of wealth data. For the densities, the solid line with the single, highest mode is the graph for 2004; it crosses the graph for 2014, also in solid line, which depicts a much more dispersed population of incomes that is bimodal.

We used the methods of Example 1 to generate a population that reflects the information in Table 6, modulo the shape parameter for a Pareto tail for the last unbounded dollar class. Density plots for the five populations and plots of the relative quantiles over time are shown in Figure 4. The density graph for 2004 is unimodal, while for subsequent

Table 6: Net Household Wealth (NHW) in thousands of Australian dollars, adjusted for inflation to 2013-2014 dollars, for all available financial years. The tabled entries represent thousands of households. Source: [1, Table 2.3], downloaded 27 July, 2017.

	2003–2004	2005–2006	2009–2010	2011–2012	2013–2014
$(-\infty, 0)$ [1]	56.6	75.6	77.3	113.7	93.8
[0, 49]	1098.9	1044.6	1058.2	1075.2	1052.7
[50, 99]	547.0	583.4	577.4	617.1	667.3
[100, 149]	364.3	374.1	408.5	441.9	433.3
[150, 199]	365.4	308.0	311.5	368.5	334.1
[200, 249]	372.7	305.4	294.6	337.7	360.1
[250, 299]	393.3	354.3	309.1	319.8	363.1
[300, 349]	372.8	356.5	348.6	306.4	329.5
[350, 399]	397.5	397.0	331.0	335.9	343.3
[400, 449]	353.5	351.6	332.1	360.7	329.3
[450, 499]	335.5	361.8	348.2	333.0	342.4
[500, 599]	574.9	601.4	621.8	554.7	570.4
[600, 699]	402.6	492.9	508.6	499.4	451.0
[700, 799]	365.9	400.8	425.8	410.4	420.2
[800, 899]	295.6	252.9	344.0	377.3	325.8
[900, 999]	211.1	220.5	283.1	267.5	298.0
[1000, 1099]	179.5	189.5	235.5	234.0	261.9
[1100, 1199]	147.3	161.9	185.0	203.9	202.4
[1200, 1399]	233.7	241.9	314.7	323.6	346.4
[1400, 1599]	138.3	196.4	213.1	228.6	242.9
[1600, 1799]	92.8	122.6	161.8	158.6	171.5
[1800, 1999]	69.9	83.7	118.9	153.3	149.0
[2000, 2199]	64.9	77.4	75.6	94.7	100.8
[2200, 2399]	50.9	53.8	72.3	68.0	78.8
[2400, 2599]	30.1	44.4	62.3	55.6	62.0
[2600, 2999]	55.8	73.6	98.4	87.6	91.4
[3000, 3999]	83.6	90.1	111.1	126.3	149.2
[4000, 4999]	19.4	41.2	61.8	68.7	70.5
[5000, 6999]	34.8	36.3	51.7	55.3	63.9
[7000, 9999]	11.7	14.5	25.4	33.1	31.8
$[10000, +\infty)$ [2]	15.5	18.3	31.4	19.9	29.5
Total	7,735.8	7,926.4	8,398.8	8,630.4	8,766.3

[1] The unknown NHWs less than 0 will be assigned 0 in our analysis.
[2] An upper bound on NHWs greater than $10,000,000.00 is not reported.

years a clear shift to bimodality is apparent. The decrease in relative quantiles for the the less wealthy half of the population between 2010 and 2012 is not seen for the wealthier half.

Empirical percentiles for these data in Table 7 reveal that while the lower percentiles are not changing over the decade, the median P50 and larger percentiles appear to be steadily increasing. The estimated inequality of wealth $\widehat{I}$ appears to be increasing only slightly over the decade. Nevertheless $\widehat{I}$ values near 0.7 are certainly much higher than those for disposable income, which was near 0.5, see the last column of Table 5. We now

Table 7: Percentiles for five distributions of NHW depicted in Figure 4, and values of $\widehat{I}$. The minimum value for each distribution was 0.

	P05	P10	P20	P25	P50	P75	P80	P90	P95	max	$\widehat{I}$
2004	15	32	86	132	388	747	867	1319	1984	43203	0.71
2006	15	34	89	137	424	787	942	1477	2182	42635	0.71
2010	16	35	97	147	474	909	1070	1670	2567	62912	0.72
2012	15	34	93	140	458	913	1085	1697	2526	40650	0.73
2014	16	36	95	143	463	961	1129	1770	2717	41328	0.73

examine the possible change in inequality over the Australian population of households and certain sub-populations of these wealth data.

3.2.1. Quartile partition estimates for NHW.

Next consider the quartile partition with members $A_1 = [0, 0.25] \cup [0.75, 1]$ and $A_2 = [0.25, 0.75]$. Incomes in the population image of A_2 can be considered as belonging to the 'middle class'. Here $\widehat{I}_k$ is defined by (6) for $k = 1, 2$. These estimates are found for each of the five populations generated (as was done for the DWI data) starting with the grouped data listed in Table 6. They are based on 10,000 observations from each of the respective populations.

Table 8: Estimates of I_k for the quartile partition of the five distributions of Figure 4 based on 10,000 observations selected at random from each of them. In parentheses are the values of $\sqrt{n}\ \widehat{\mathsf{SE}}[\widehat{I}_k]$.

	2004	2006	2010	2012	2014
$\widehat{I}_1$	0.949(0.16)	0.948(0.16)	0.952(0.16)	0.956(0.15)	0.955(0.14)
$\widehat{I}_2$	0.480(0.54)	0.466(0.55)	0.473(0.56)	0.500(0.53)	0.498(0.53)
$\widehat{I}$	0.714(0.33)	0.707(0.33)	0.712(0.34)	0.726(0.32)	0.726(0.31)

Estimates for the quartile partition are provided in Table 8. A level 0.05 test for a difference between the 2004 and 2014 values of $\widehat{I}$, namely $|0.726 - 0.714| = 0.012$ would reject for $n = 10,000$ because then the standard error of the difference between them is $\mathrm{SE} = \sqrt{(0.33^2 + 0.31^2)/n} = 0.004$. A similar test for significant change in inequality for the middle class over the same range of time is just significant at the 0.05 level, because the difference is 0.018 and the standard error of the difference is 0.008. The standard errors for the estimates of inequality in the outer quartiles $\widehat{I}_1$ are much smaller, so while the difference between the 2014 and 2004 results is only 0.955-0.949 =0.006 but its standard error is also much smaller at 0.002, leading to a statistically significant result. Thus most of the change in wealth inequality over this period is due to the change in the lower and upper quartiles. Note that such tests are correlated.

Decile partition estimates for NHW.

Next we look at a finer partition, the decile partition, to further pinpoint where the wealth inequality is changing most.

Table 9: Estimates of I_k for the decile partitioning of the five distributions deicted in Figure 4. As in Table 8, estimates are based on 10,000 observations selected at random from each of them, and values in parentheses are $100 \times \widehat{\mathsf{SE}}[\widehat{I}_k]$.

	2004	2006	2010	2012	2014
$\widehat{I}_1$	0.991(0.04)	0.991(0.04)	0.992(0.03)	0.993(0.03)	0.993(0.03)
$\widehat{I}_2$	0.944(0.19)	0.945(0.19)	0.949(0.18)	0.949(0.18)	0.951(0.16)
$\widehat{I}_3$	0.812(0.51)	0.811(0.52)	0.827(0.48)	0.832(0.44)	0.842(0.42)
$\widehat{I}_4$	0.565(0.69)	0.557(0.69)	0.584(0.68)	0.606(0.69)	0.618(0.66)
$\widehat{I}_5$	0.223(0.51)	0.218(0.49)	0.232(0.52)	0.235(0.54)	0.229(0.55)
$\widehat{I}$	0.707(0.33)	0.705(0.33)	0.717(0.32)	0.723(0.32)	0.729(0.31)

Estimates and their standard errors are given in Table 9. Note that the standard errors of the estimates can vary from 0.0003 to 0.0069.

Using (3) the results in the second column show that for 2004 the overall inequality can be broken down into $\widehat{I} = 0.707 = 0.198 + 0.189 + 0.162 + 0.113 + 0.045$, so the first three partition members (outer six deciles) contribute 0.549/0.707 or almost 80% to the overall QRI and the fifth partition (central two deciles) only 0.045/0.707, or 6%.

Comparing the estimates of I_5 in 2004 and 2014 for the central partition shows a less than one standard error of increase. However, for every other partition member the QRI has increased by more than two standard errors over this time period. We can conclude that wealth inequality is becoming more unequal except within the central 1/5 of the population.

4. Shiny applications

For convenience, we have created two Shiny [3] applications that some readers may find useful. These applications can be found at

`https://lukeprendergast.shinyapps.io/Decomp/`

and

`https://lukeprendergast.shinyapps.io/QRIestimation/`.

The first calculates the QRI for several distributions considered within this manuscript. This application also calculates the quartile and quintile decompositions for the QRI if requested. The second allows the user to upload a csv data file for which to estimate the QRI and its quartile and quintile decompositions. The standard errors and large sample confidence interval estimators for the QRI and its decompositions are included in the tabulated output. We will continue to improve these applications and are grateful for any feedback.

5. Conclusion

We have learned that for Australian data, inequality of DWI as measured by the QRI is almost steady at about $\widehat{I} = 0.5$ over the years 2004 to 2014, although for a large enough sample it has increased by a statistically significant amount. Samples of size 1000 would not detect such an increase. Further, by examining the QRI estimates for the quartile partition, we found that all the change in inequality of incomes over the time period 2004 to 2014 can be attributed to the outer classes, with incomes inequality in the middle class remaining stagnant. The inequality in wealth NHW for Australian households over this same period was much higher $\widehat{I} \approx 0.7$ and very significantly increases from 2004 to 2014. Moreover, hardly any of this increase is due to the middle two deciles.

These examples illustrate the simple utility of measuring inequality with the QRI. Not only can we find reliable confidence intervals for the QRI of a population using relatively small samples in the hundreds, we can also find them for symmetric partitions of quantiles, and use them to identify those which contribute most to the population QRI and how much them contribute. Finally we can use these results to detect changes over time. Given income or wealth data from various countries, it would be straightforward, using the accompanying online scripts, to compute the QRI for each of them and/or desired symmetric sub-populations. One could also study income data for those in the top 10% by means of the QRI, or any other region of interest.

Testing for changes in the QRI over time and for several components of the decomposition means that it would be wise to take the usual care when dealing with multiple testing. While the Bonferroni correction is often seen as too conservative, methods for controlling the false discovery rate for dependent tests may be appealing.

Given that the QRI is a simple average of quantile inequality measures of the form $R(p) = 1 - Q(p/2)/Q(1-p/2)$, one for each $0 < p < 1$, it should be possible to introduce weights or otherwise extend it to other partitions where meaningful decompositions arise.

Acknowledgements

The authors would like to thank the three referees and the Editor for taking their time to review the paper and for the suggestions made.

References

[1] ABS. Household data and income distribution, Australian Bureau of Statistics Report 6523.0. Canberra, ACT, Australia, 2016. Available on www.ausstats.abs.gov.au.

[2] F. Bourguignon. Decomposable income inequality measures. *Econometrica*, 47(4):901–920, 1979.

[3] W. Chang, J. Cheng, J. J. Allaire, Y. Xie, and J. McPherson. *shiny: Web Application Framework for R*, 2017. R package version 1.0.5.

[4] F.A. Cowell. *Measuring Inequality.* Oxford University Press, Oxford, Third edition, 2011.

[5] F.A. Cowell and M.P. Victoria-Feser. Robustness properties of inequality measures. *Econometrica*, 64(1):77–101, 1996.

[6] T. Das and A. Parikh. Decomposition of inequality measures and a comparative analysis. *Empirical Economics*, 7:23–48, 1982.

[7] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008.

[8] B.V. Frosini. Approximation and decomposition of Gini, PietraRicci and Theil inequality measures. *Empirical Economics*, 43:175–197, 2012.

[9] R.J. Hyndman and Y. Fan. Sample quantiles in statistical packages. *The American Statistician*, 50:361–365, 1996.

[10] L Pasquazzi and M Zenga. Components of Gini, Bonferroni, and Zenga inequality indexes for EU income data. *Journal of Official Statistics*, 34:149–180, 2018.

[11] L.A. Prendergast and R.G. Staudte. Exploiting the quantile optimality ratio in finding confidence intervals for quantiles. *Stat*, 5:70–81, 2016a.

[12] L.A. Prendergast and R.G. Staudte. Quantile versions of the Lorenz curve. *Electronic Journal of Statistics*, 10(2):1896–1926, 2016b.

[13] L.A. Prendergast and R.G. Staudte. When large n is not enough—distribution-free interval estimators for ratios of quantiles. *Journal of Economic Inequality*, 15:277–293, 2017.

[14] L.A. Prendergast and R.G. Staudte. A simple and effective inequality measure. *The American Statistician*, 72(4):328–343, 2018.

[15] A.F. Shorrocks. The class of additively decomposable inequality measures. *Econometrica*, 48(3):613–625, 1980.

[16] S. Subramanian. Inequality measurement with subgroup decomposability and level-sensitivity. *Economics: The Open-Access, Open-Assessment E-Journal*, (9):1–18, 2011. dx.doi.org/10.5018.

[17] B. Zheng. Unit-consistent decomposable inequality measures. *Economica*, 74(293):97–111, 2007.

Appendix

A. Derivation of the indefinite integral (4)

Starting with $R_\sigma(p) = \exp\{2\sigma z_{p/2}\}$ and making the change of variable $y = 2\sigma z_{p/2}$

$$\int_0^r R_\sigma(p)\,dp = \frac{1}{\sigma}\int_{-\infty}^{2\sigma z_{r/2}} e^y \varphi\left(\frac{y}{2\sigma}\right)\,dy\ ,$$

where $\varphi(z) = e^{-z^2/2}/\sqrt{2\pi}$ is the standard normal density. Completing the square within the exponential of the integrand $\exp\{y - (y/2\sigma)^2/2\}$ leads to

$$\frac{8\sigma^2 y - y^2}{8\sigma^2} = \frac{16\sigma^4 - (y-4\sigma^2)^2}{8\sigma^2}\ ,\ \text{so}$$

$$\int_0^r R_\sigma(p)\,dp = 2\exp\{2\sigma^2\}\int_{-\infty}^{2\sigma z_{r/2}} \frac{1}{2\sigma}\varphi\left(\frac{y-4\sigma^2}{2\sigma}\right)\,dy = 2\exp\{2\sigma^2\}\,\Phi(\Phi^{-1}(r/2) - 2\sigma)\ .$$

B. Exact decomposition formula for $\hat{I}$

Taking the weighted average of the $\widehat{I}_k^{(J)}$s defined in (6) does not guarantee that it will be exactly equal to the estimator of I found by $\widehat{I}^{(J)} = \{\sum_{j=1}^{J}[1-\widehat{R}(p_j)\}/J$. However, [14] showed the estimates are stable for moderate to large choices of J so that the resulting weighted average of the $\widehat{I}_k^{(J)}$s is expected to be very close to $\widehat{I}^{(J)}$. However, if $n_0 = np_0 = 0$, $n_1 = np_1$, $n_2 = np_2 \ldots, n_{K-1} = np_{K-1}$ and $n_K = n/2$ are all distinct integers, then it is possible to define estimates of the $\widehat{I}_k$s such that their weighted average is equal to a simple estimator of I.

Given ordered incomes $0 \le x_1 \le x_2 \le \cdots \le x_n$, where $n \ge 2$ and the frequency of 0's is less than 0.5, let $k = \lfloor n/2 \rfloor$. It is shown in [14, Equation 3] that an exact estimate of I is given by

$$I_n = I(F_n) = \frac{2}{n}\sum_{j=1}^{k}\left(1 - \frac{x_j}{x_{n-j+1}}\right)\ . \qquad (8)$$

Given a symmetric K-partition $\{A_1, \ldots, A_K\}$ of the unit interval determined by $0 = p_0 < p_1 < \cdots < p_K = 1/2$, we want to decompose $\widehat{I}$ into a weighted average of individual estimates $\widehat{I}_k$ of I_k. To this end assume $n_0, n_2, \ldots$ as above and assume that they are distinct integers. Define the disjoint sets $B_k = \{n_{k-1}+1, n_{k-1}+2, \ldots, n_k\}$, for $k = 1, \ldots, K$. The length of B_k is $m_k = n_k - n_{k-1}$ and $\sum_{k=1}^{K} m_k = n_K = n/2$. We estimate I_k by

$$\widehat{I}_k = \frac{1}{m_k}\sum_{j\in B_k}\left(1 - \frac{x_j}{x_{n-j+1}}\right)\ .$$

Then a simple estimator of $\widehat{I}$ based on the order statistics can be written

$$\widehat{I} \;=\; \frac{2}{n}\sum_{j=1}^{n/2}\left(1 - \frac{x_j}{x_{n-j+1}}\right)$$

$$
\begin{aligned}
&= \frac{2}{n} \sum_{k=1}^{K} \sum_{j \in B_k} \left(1 - \frac{x_j}{x_{n-j+1}}\right) \\
&= \frac{2}{n} \sum_{k=1}^{K} m_k \, \widehat{I}_k \ , \qquad (9)
\end{aligned}
$$

where $m_k = n_k - n_{k-1}$. Note that the sum of the weights $\sum_k 2m_k/n = 1$.