Robust Inequality Measures

R.G. Staudte[1*] and L.A. Prendergast[1]

[1] *La Trobe University, Melbourne 3086 Australia;*
r.staudte@latrobe.edu.au, luke.prendergast@latrobe.edu.au
[*] *Presenting author*

Keywords. *Gini index; Influence function; Density quantile*

The Lorenz curve and the associated Gini coefficient are routinely employed for comparisons of income inequality in various countries. These concepts have nice mathematical properties, and thus are the subject of numerous theoretical studies. But when it comes to statistical inference for them, thorny issues arise: in particular, **?** show that these and many other inequality measures in the econometrics literature have unbounded influence functions. While robust methods are available for income distributions with heavy tails, usually only grouped data are available for privacy reasons. To deal with such situations, we redefine the basic concept of the Lorenz curve in terms of quantiles instead of moments, and see what has been gained and lost in terms of conceptual clarity, inference and estimator resistance to contamination.

Let x_p be the pth quantile of a population of incomes, and let μ_p be the mean of smallest incomes $x \leq x_p$. Then if $\mu = \lim \mu_p$ exists as p approaches one, the Lorenz curve is defined by $L_0(p) = p\,\mu_p/\mu$ for all $0 < p < 1$. What we propose is to replace in this definition the mean μ_p by the median $x_{p/2}$ of incomes $x \leq x_p$; and, in the denominator to replace the mean μ by one of three quantities: $d_1(p) = x_{0.5}$, $d_2(p) = x_{1-p/2}$ or $d_3(p) = (x_{p/2} + x_{1-p/2})/2$. This leads to three quantile based versions of the Lorenz curve $L_i(p) = p\,x_{p/2}/d_i(p)$ for $i = 1, 2$ and 3. The associated inequality coefficients G_i are defined as one minus twice the area between the graph of L_i and the diagonal line for $i = 0, 1, 2$ and 3.

The quantile inequality curves measure inequality in that for any non-decreasing transfer of income function that does not increase the distance of any quantile from the median, the curve ordinates $L_i(p)$ can only increase for each p, which means that the corresponding coefficient of inequality can only decrease. The quantile inequality curves have most of the nice properties of the Lorenz curve, including convexity for all income models commonly used in the literature. In addition, they and the associated inequality coefficients have bounded influence functions.

By substituting sample quantile estimates into the formulae defining the L_i and G_i one obtains estimators $\hat{L}_i$ and $\hat{G}_i$ with standard errors depending on unknown quantile densities as well as the quantile function. Using quantile density estimators from **?**, this leads to large-sample distribution-free confidence intervals for the quantile inequality coefficients that have good coverage probabilities, even for moderate sample sizes, **?**.

The Gini coefficient has been criticized for placing too much emphasis on the middle incomes, and the quantile versions can be criticized for the same reason. To see why, let Y_1, Y_2 be independent, random incomes less than the median, and let $V = \max\{Y_1, Y_2\}$. Then by a change of variable, one finds $G_1 = \mathrm{E}[(m - V)/m]$. That is, G_1 is the average relative distance of V from the median. Next define $W = x_{1-r}$ to be the $(1-r)$th quantile of incomes whenever $V = x_r$ is the rth quantile of incomes. Then it follows that $G_2 = \mathrm{E}[(W - V)/W]$ and $G_3 = \mathrm{E}[(W - V)/(V + W)]$.

The maximum $V = \max\{Y_1, Y_2\}$ arises because of the multiplier p in the definition of $L_i(p)$, as one can see by omitting it. For example, if L_1 were redefined to be $L_1^*(p) = x_{p/2}/x_{0.5}$ taking values in [0,1], and G_1 redefined to G_1^*, the area between $L_1^*(p)$ and the horizontal line over the unit interval, then $G_1^* = (m - \mathrm{E}[Y])/m$, where Y has the conditional distribution of X, given that X is less than its median. Thus G_1^* is the average relative distance of a *single* randomly chosen income less than the median from the median.

If one wants an inequality curve and associated coefficient of inequality that do not suffer from down-weighting small incomes, and is willing to give up the property of convexity, then L_1^*, G_1^* is a possibility. The analogous definitions L_2^*, G_2^* and their robust estimators are investigated in **?**.

References

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

L.A. Prendergast and R.G. Staudte. Exploiting the quantile optimality ratio in finding confidence intervals for quantiles. *Stat*, 5:70–81, 2016a. DOI: 10.1002/sta4.105.

L.A. Prendergast and R.G. Staudte. Quantile versions of the Lorenz curve. *http://arxiv.org/abs/1510.06085*, 2016b.

L.A. Prendergast and R.G. Staudte. A simple and Effective Inequality Measure. *http://arxiv.org/abs/1603.03481*, 2016c.