Robust Lorenz Curves: A Semi-Parametric Approach

M.-P. Victoria-Feser[1], and F.A. Cowell[2]

[1] University of Geneva, 40, bd du Pont d'Arve, 1211 Geneva, Switzerland. (presenting author)
[2] London School of Economics, Houghton street, London WC2A 2AE, UK.

Keywords: Income distribution, stochastic dominance, Pareto model, M-estimators.

1 Abstract

The Lorenz curve is central to the analysis of income distributions: it is a powerful tool used for inequality and welfare comparisons from one or several income samples. In practice however income-distribution data may be contaminated by recording errors, measurement errors and the like and, if the data cannot be purged of these, welfare conclusions drawn from the data can be seriously misleading. The purpose of this paper is to provide a method for handling these potential problems, one that accords well with pragmatic procedures that are sometimes adopted by applied researchers in this field.

The point of departure is recent research which has shown that Lorenz and stochastic dominance results are non-robust (Cowell and Victoria-Feser, 2002). This means that small amounts of data contamination in the wrong place can reverse unambiguous welfare conclusions: the "wrong place" usually means in the upper tail of the distribution. So it is important to have an approach that enables one to control for the distortionary effect of upper-tail contamination in a systematic fashion. We need a robust method of estimating Lorenz curves and implementing stochastic dominance criteria.

We explore here approaches based on the specification of parametric models for the distribution of the data and use robust estimators of the parameters. These approaches are of particular interest because of their use in practical treatment of problems associated with the distribution in the upper tails like in income and wealth distributions as well as in extreme value theory. An example is the use of a Pareto distribution for fitting the tail of income data in cases where data are sparse in order to provide better estimates of inequality measures. Another example is the use of a generalized Pareto distribution to model extremal events in order to provide better extreme quantile estimates.

More precisely, let

$$C(F;q) := \int_{\underline{x}}^{Q(F;q)} x dF(x) = c_q, q \in (0,1)$$

be the *cumulative income* the functional based on the income distribution F with support $(\underline{x}, \overline{x})$, where $Q(F;q) = \inf\{x|F(x) \geq q\} = x_q$ is the quantile functional. $C(F;q)$ actually defines the generalized Lorenz curve (GLC), whereas the (more popular) relative Lorenz curve (RLC) is defined as

$$L(F;q) := \frac{C(F;q)}{C(F,1)}$$

Note that $C(F,1) = \mu(F)$, the mean functional. Two distributions can then be compared in terms of their respective RLC. For example, if $L(F;q) > L(F;q), \forall q$ then distribution F is associated with a higher welfare (less unequal) than distribution G. To estimate the RLC (or indeed the GLC), a traditional approach consists in replacing in the definition the true distribution by the empirical distribution. Because this approach can lead to welfare conclusions that are reversed just because of a few data points (Cowell and Victoria-Feser, 2002), we propose a semi-parametric model for the Lorenz curve based on the Pareto model for the α percent upper tail and use a robust estimator for the parameter of the Pareto distribution. We also provide the the necessary results for inference with estimated RLC (and indeed GLC). Finally some simulated examples will show how our approach can be beneficial at least as an alternative to a more traditional approach.

References

F. Cowell and M.-P. Victoria-Feser (2002). Welfare rankings in the presence of contaminated data. *Econometrica*, (forthcoming).