Robust Lorenz Curves: A Semi-Parametric Approach

M.-P. Victoria-Feser (presenting author)
University of Geneva, 40, bd du Pont d’Arve,
1211 Geneva, Switzerland.

F.A. Cowell
London School of Economics, Houghton street,
London WC2A 2AE, UK.

May 27, 2004

Abstract

Welfare comparisons based on Lorenz curve is are made using income-distribution data. In practice however these 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. In fact, Cowell and Victoria-Feser (2002) have shown that Lorenz and stochastic dominance results are non-robust, especially when the contamination occurs in the upper tail of the distribution. The purpose of this paper is to provide a method for handling these potential problems.

We explore here an approach based on the specification of parametric models for the upper-tail distribution of the data and use robust estimators of the parameters. This type of approach is actually used in practice for other reasons such as in cases where data are sparse in order to provide better estimates of inequality measures, or in modelling extremal events in order to provide better extreme quantile estimates.

More precisely, let

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

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

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

To estimate the RLC (or indeed the GLC), a traditional approach consists in replacing in the definition the true distribution by the empirical distribution. We propose instead a semi-parametric model for the Lorenz curve based on the Pareto model for the \( \alpha \) percent upper tail and use a robust estimator for the parameter of the Pareto distribution. We also provide the necessary results for inference and through some simulated examples will show how our approach can be beneficial.

References