Modelling Lorenz Curves: robust and semi-parametric issues

Frank A. Cowell

Maria-Pia Victoria-Feser

London School of Economics and Université de Genève

DARP 91
March 2007

The Toyota Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics
Houghton Street
London WC2A 2A

(+44 020) 7955 6674

Abstract

Modelling Lorenz curves (LC) for stochastic dominance comparisons is central to the analysis of income distribution. It is conventional to use non-parametric statistics based on empirical income cumulants which are in the construction of LC and other related second-order dominance criteria. However, although attractive because of its simplicity and its apparent flexibility, this approach suffers from important drawbacks. While no assumptions need to be made regarding the data-generating process (income distribution model), the empirical LC can be very sensitive to data particularities, especially in the upper tail of the distribution. This robustness problem can lead in practice to "wrong" interpretation of dominance orders. A possible remedy for this problem is the use of parametric or semi-parametric models for the data-generating process and robust estimators to obtain parameter estimates. In this paper, we focus on the robust estimation of semi-parametric LC and investigate issues such as sensitivity of LC estimators to data contamination (Cowell and Victoria-Feser 2002), trimmed LC (Cowell and Victoria-Feser 2006) and inference for trimmed LC (Cowell and Victoria-Feser 2003), robust semi-parametric estimation for LC (Cowell and Victoria-Feser 2007) selection of optimal thresholds for (robust) semi-parametric modelling (Dupuis and Victoria-Feser 2006) and use both simulations and real data to illustrate these points.

Distributional Analysis Research Programme

The Distributional Analysis Research Programme was established in 1993 with funding from the Economic and Social Research Council. It is located within the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science. The programme is directed by Frank Cowell. The Discussion Paper series is available free of charge. To subscribe to the DARP paper series, or for further information on the work of the Programme, please contact our Research Secretary, Leila Alberici on:

Telephone: UK+20 7955 6674
Fax: UK+20 7955 6951
Email: l.alberici@lse.ac.uk
Web site: http://sticerd.lse.ac.uk/DARP

1 Introduction

The Lorenz curve is central to the analysis of income distributions, embodying fundamental intuition about inequality comparisons (Dagum 1985, Cowell 2007). Ranking theorems based on Lorenz dominance and the associated concept of stochastic dominance are fundamental to the theoretical welfare economics of distributions. But formal welfare propositions can only be satisfactorily invoked for empirical constructs if sample data can be taken as a reasonable representation of the underlying income distributions under consideration. In practice 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. Indeed, it has been formally 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 ranking orders: the "wrong place" usually means in the upper tail of the distribution. This is of particular interest in view of a burgeoning recent literature that has focused on empirical issues concerning the upper tail of both income distributions and wealth distributions (Atkinson 2004, Kopczuk and Saez 2004, Moriguchi and Saez 1991, Piketty 2001, Piketty and Saez 2003, Saez and Veall 2005). 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. This paper addresses the problem by introducing a robust method of estimating Lorenz curves and implementing stochastic dominance criteria.

Our approach is organized as follows. We begin, in section 2, by setting out the formal background to the Lorenz curve and the estimation problems associated with extreme values. Section 3 develops the semi-parametric approach to modelling Lorenz curves and section 4 discusses the practical problem of parameter choice in implementing the method. Section 5 applies the method to UK data and section 6 concludes.

2 Background

We may set out the formal representation of the Lorenz Curve using the following simple framework. Let $\mathfrak{F}$ be the set of all univariate probability distributions and X be a random variable with probability distribution $F \in \mathfrak{F}$

and support $\mathfrak{X} \subseteq \mathbb{R}$. F can be thought of as a parametric model F_θ. We shall write statistics of any distribution $F \in \mathfrak{F}$ as a functional $T(F)$; in particular we write the mean as $\mu(F) := \int x dF(x)$. A key distributional concept derived from F is given by the q^{th} *cumulative functional* $C : \mathfrak{F} \times [0,1] \mapsto \mathfrak{X}$:

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

where $\underline{x} := \inf \mathfrak{X}$ and

$$Q(F;q) = \inf\{x | F(x) \geq q\} = x_q \tag{2}$$

is the quantile functional. The importance of this concept is considerable in the practical analysis of income distributions: for a given $F \in \mathfrak{F}$, the graph of $C(F,q)$ against q describes the *generalized Lorenz curve* (GLC); normalizing by the mean functional $\mu(F) = C(F,1)$ one has the *Relative Lorenz curve* (RLC) (Lorenz 1905):

$$L(F;q) := \frac{C(F;q)}{\mu(F)} \tag{3}$$

The GLC and RLC are fundamental to a number of theorems drawing welfare-conclusions from income-distribution data and other types of data.

Now consider the problem of estimating Lorenz curves. There are broadly three approaches.

1. *Nonparametric methods.* Cumulative functionals can obviously be estimated by replacing F in (1) by the empirical distribution of a sample of incomes $x_1, \ldots, x_n$

$$F^{(n)}(y) = \frac{1}{n} \sum_{i=1}^{n} \iota (y \leq x_i)$$

where $\iota(\cdot)$ is the *indicator function*. However, this can lead to misleading conclusions when it comes to comparing distributions in terms of their cumulative functionals when there is data contamination (Cowell and Victoria-Feser 2002). One way of avoiding the potential bias induced by extreme data in the tails is to rely on the concept of trimmed

Lorenz curves: basically, F in (1) is replaced by the trimmed distribution $\tilde{F}_\alpha$ given by:

$$\tilde{F}_\alpha(x) := \begin{cases} 0 & \text{if } x < Q(F, \underline{\alpha}) \\ \frac{F(x)-\underline{\alpha}}{1-\alpha} & \text{if } Q(F, \underline{\alpha}) \leq x < Q(F, \overline{\alpha}) \\ 1 & \text{if } x \geq Q(F, \overline{\alpha}) \end{cases} .$$

with $\underline{\alpha} + \overline{\alpha} = \alpha$. Using $\tilde{F}_\alpha$ instead of $F^{(n)}$ amounts to trimming the sample data below $Q(F, \underline{\alpha})$ and above $Q(F, \overline{\alpha})$, and then compute empirical cumulants. The theoretical aspects are handled in Cowell and Victoria-Feser (2006).

2. *Parametric modelling.* Alternatively, one can estimate F using a model (a functional form) such as the one proposed by Dagum (1977).[1] The parameters should obviously be estimated in a robust fashion (see e.g. Victoria-Feser and Ronchetti 1994, Victoria-Feser 1995), but as has been discussed in Cowell and Victoria-Feser (2007), a full parametric estimation forces the data into the mould of a functional form that may not be suitable for comparisons.

3. *Semi-parametric approach.* The problem that a single, tractable functional form may not be appropriate for the data motivates the use of an approach in which the data above a threshold x_0 are (robustly) fitted to a parametric distribution, while the rest of the data are treated non-parametrically. The semi parametric approach is of particular interest because of its *ad hoc* use in practical treatment of problems associated with the upper tails of distributions. For example a Pareto tail is sometimes fitted to data in cases where data are sparse in order to provide better estimates of upper tail probabilities or higher quantiles.

It is this third estimation method, the semiparametric approach, that forms the focus of the present paper.

[1] Other models can be found in Dagum (1980), Dagum (1983) and McDonald (1984) and an excellent overview is provided by Kleiber and Kotz (2003).

3 Semi-parametric robust estimation of Lorenz curves

If the range of X is bounded below – 0 is a typical value – the problems with contaminated data occur in the upper tail of the distribution (Cowell and Victoria-Feser 2002). A case can therefore be made for using parametric modelling only in the upper tail and estimating the parameter of the upper-tail model robustly. The rest of the distribution is estimated using the empirical distribution function. If no restriction is imposed on the range of the random variable of interest, then the results below can easily be extended accordingly.

Cowell and Victoria-Feser (2007) proposed an approach which is suitable for any parametric model for the upper tail of the distribution. They however choose a model that is of special relevance empirically, that is the Pareto distribution given by

$$F_\theta(x) = 1 - \left[\frac{x}{x_0}\right]^{-\theta}, x > x_0 \tag{4}$$

with density $f(x;\theta) = \theta x^{-(\theta+1)} x_0^\theta$. The parameter of interest is θ[2]. A semi-parametric approach will combine a non-parametric RLC for say the $(1-\alpha)\%$ lower incomes and a parametric RLC based on the Pareto distribution for the $\alpha\%$ upper incomes. Therefore x_0 is determined by the $1-\alpha$ quantile $Q(F;1-\alpha)$ defined in (2). The method for a suitable choice for x_0 is given in section 4. The full semi-parametric distribution $\widetilde{F}$ of the income variable X is

$$\widetilde{F}(x) = \begin{cases} F(x) & x \leq x_0 \\ F(x_0) + (1 - F(x_0))F_{\theta,x_0}(x) & x > x_0 \end{cases}$$

where F could be in principle any suitable parametric distribution, but in our case will be estimated by the empirical distribution. With $x_0 = Q(F;1-\alpha)$, we have

$$\widetilde{F}(x) = \begin{cases} F(x) & x \leq Q(F;1-\alpha) \\ 1 - \alpha\left[\frac{x}{Q(F;1-\alpha)}\right]^{-\theta} = & x > Q(F;1-\alpha) \end{cases} . \tag{5}$$

For $x > Q(F;1-\alpha)$, the density $\widetilde{f}$ is

$$\widetilde{f}(x;\theta) = \alpha\theta Q(F;1-\alpha)^\theta x^{-\theta-1} .$$

[2] θ is assumed to be greater than 2 for the variance to exist.

In particular

$$\widetilde{f}(x_{1-\alpha};\theta) = \frac{\alpha\theta}{x_{1-\alpha}} \; . \tag{6}$$

The quantile functional is then obtained using (5) and is given by

$$Q(\widetilde{F},q) = \begin{cases} Q(F,q) & q \leq 1-\alpha \\ Q(F;1-\alpha)\left(\frac{1-q}{\alpha}\right)^{-1/\theta} & q > 1-\alpha \end{cases}$$

Hence the cumulative income functional defining the semi-parametric GLC becomes

$$\begin{aligned} C(\widetilde{F};q) &= \int_{\underline{x}}^{Q(\widetilde{F},q)} x d\widetilde{F}(x) \\ &= \begin{cases} \int_{\underline{x}}^{Q(F,q)} x dF(x) & q \leq 1-\alpha \\ \int_{\underline{x}}^{Q(F,1-\alpha)} x dF(x) & \\ +\alpha \int_{Q(F,1-\alpha)}^{Q(F;1-\alpha)\left(\frac{1-q}{\alpha}\right)^{-1/\theta}} x dF_\theta & q > 1-\alpha \end{cases} \\ &= \begin{cases} \int_{\underline{x}}^{Q(F,q)} x dF(x) & q \leq 1-\alpha \\ \int_{\underline{x}}^{Q(F,1-\alpha)} x dF(x) & \\ +\alpha\frac{\theta}{1-\theta} Q(F;1-\alpha)\left[\left(\frac{1-q}{\alpha}\right)^{\frac{\theta-1}{\theta}} - 1\right] & q > 1-\alpha \end{cases} \end{aligned} \tag{7}$$

where $\underline{x} := \inf \mathfrak{X}$. The mean of the semi-parametric distribution is given by:

$$\begin{aligned} C(\widetilde{F};1) &= \int_{\underline{x}}^{Q(F,1-\alpha)} x dF(x) - \alpha Q(F;1-\alpha)\frac{\theta}{1-\theta} \\ &= c_{1-\alpha} - \alpha x_{1-\alpha}\frac{\theta}{1-\theta} \\ &= \mu(\widetilde{F}) \end{aligned} \tag{8}$$

The semi-parametric RLC is simply

$$L(\widetilde{F};q) = \frac{C(\widetilde{F};q))}{\mu(\widetilde{F})} \tag{9}$$

(7) needs obviously to be estimated. F is replaced by the empirical distribution $F^{(n)}$ and an estimate for α will be discussed in Section 4. To estimate

the Pareto model, hence θ, for the upper tail of the distribution, one can use the maximum likelihood estimator (MLE). Unfortunately, the MLE for the Pareto model is known to be very sensitive to data contamination (Victoria-Feser and Ronchetti 1994). This is also the case for other models such as Dagum (1977) model (see Victoria-Feser 1995). Cowell and Victoria-Feser (2007) propose to use a robust estimator in the class of M-estimators (Huber 1981). For a sample of n observations x_i, a general M-estimator is defined as the solution in θ of

$$\frac{1}{n}\sum_{i=1}^{k}\psi\left(x_i;\theta\right)=0$$

with some (mild) conditions on the function ψ. This function is chosen so that the resulting estimator is consistent at the model F_θ and also that it is robust to slight model deviations (for a discussion, see e.g. Hampel, Ronchetti, Rousseeuw, and Stahel 1986). The latter condition is satisfied if the ψ-function is bounded, which is the case for so-called weighted MLE (WMLE), i.e.

$$\frac{1}{n}\sum_{i=1}^{k}w\left(x_i;\theta\right)\left[s\left(z_i;\theta\right)-a(\theta)\right]=0 \tag{10}$$

where $w\left(x;\theta\right)$ is a weight function with value in $[0,1]$ insuring the robustness of the estimator, $s\left(x;\theta\right)=\partial/\partial\theta\log f(x;\theta)$ is the score function and $a(\theta)$ is a consistency correction factor[3]. Cowell and Victoria-Feser (2007) choose the *optimal B-robust estimators* (OBRE) (Hampel et al. 1986), a robust estimator with minimal asymptotic covariance matrix (see e.g. for Cowell and Victoria-Feser 2007 details).

The resulting semi-parametric GLC (and RLC) estimates are hence robust to data contamination. They can be used for robust welfare comparison. Cowell and Victoria-Feser (2007) also provide the asymptotic covariances of the estimators for inference. In section 5, an example will illustrate the performance of robust semi-parametric estimators of RLC and GLC.

4 Choosing α

The choice of the proportion α of data in the upper tail to be fitted to the Pareto model, or equivalently the threshold x_0 above which the data are fitted

[3]The correction factor does not need to be estimated simultaneously, see below.

to a Pareto model, is not a problem specific to income distribution analysis. It has attracted and still attracts the attention of researchers in domains such as finance, insurance, engineering, or environmental sciences. This problem falls within the general heading of extreme value distributions (for a general reference, see e.g. Embrechts et al. 1997). To estimate the threshold, a compromise should be sought between bias and variance: choosing a threshold too close to the central data will cause bias in the Pareto model estimator since only the tail can be assumed to be Pareto distributed, and selecting too extreme a threshold will yield large variances for the estimator since it will be based on a small sample. A common practice is to use the Pareto quantile plot (see e.g. Beirlant, Vynckier, and Teugels 1996). Indeed, rearranging (4) one gets

$$\log\left(\frac{x}{x_0}\right) = -\frac{1}{\theta}\log\left(1 - F_\theta(x)\right), \quad x > x_0 \tag{11}$$

showing that there is a linear relationship between the log of the $x > x_0$ and the log of the survival function. This relationship was actually found empirically by Pareto (1896) and led him to the construction of his model (see also Dagum 1983). Let $x^*_{[i]}$, $i = 1, \ldots, k$, be the ordered largest k observations, so that $x^*_{[i]} = Q(F^{*(n)}; i/(k+1))$, with $F^{*(n)}$ the empirical distribution of $x^*_{[i]}$. The empirical counterpart of (11) is the Pareto quantile plot

$$\log\left(\frac{Q(F^{*(n)}; i/(k+1))}{x_0}\right) = -\frac{1}{\theta}\log\left(\frac{k+1-i}{k+1}\right), \quad i = 1, \ldots, k. \tag{12}$$

Therefore, given a sample of n income data $x_i, 1, \ldots, n$ and by letting $x_{[i]}$ denote the ith order statistic, the plot of $\log\left(x_{[i]}\right)$ versus $-\log\left((n+1-i)/(n+1)\right)$, $i = 1, \ldots, n$ is the Pareto quantile plot that is used to detect graphically the quantile $x_{[i]}$ above which the Pareto relationship is valid, i.e. the point above which the plot yields a straight line. We note that there is a clear relationship between x_0 and k in that $k = \sum_{i=1}^{n} \iota(x_{[i]} \geq x_0)$.

More formally, a general approach in determining k is the minimization of an estimate of the asymptotic mean squared error (AMSE) of the estimator of θ. If a classical estimator such as the MLE is chosen, then the determination of k can be influenced by extreme data in the upper tail (see Dupuis and Victoria-Feser 2006). Note that here extreme is used relatively to the Pareto model: if it is assumed to fit the upper tail, then extreme data represent deviations for this assumption that can appear in the Pareto quantile plot as data that do not fit the straight line.

In order to choose k, or equivalently x_0 in a robust fashion, Dupuis and Victoria-Feser (2006) use another criterion, namely a prediction error criterion that is estimated robustly (see also Ronchetti and Staudte 1994), named the RC-criterion. Let $Y_i = \log\left(x^*_{[i]}/x_0\right), i = 1, \ldots, k$, $\hat{Y}_i = -1/\hat{\theta}\log\left[(k+1-i)/(k+1)\right], i = 1, \ldots, k$ where $\hat{\theta}$ is an estimator of θ, and

$$\widehat{\sigma}_i^2 = \text{var}(Y_{.i}) = \sum_{j=1}^{i} \frac{1}{\widehat{\theta}^2 (k-i+j)^2}$$

the (estimated) RC-criterion is given by

$$C_R(x_0) = \frac{1}{n}\sum_{i=1}^{n} \hat{w}_i^2 \left(\frac{Y_i - \hat{Y}_i}{\sigma_i}\right)^2 + \frac{2}{n}\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\text{cov}\left[\hat{w}_i Y_i, \hat{w}_i \hat{Y}_i\right] - \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\text{var}\left[\hat{w}_i Y_i\right] \tag{13}$$

where each $\hat{w}_i, 0 \leq \hat{w}_i \leq 1$, is the fitted weight of the i^{th} observation, provided by a robust fit of the Pareto model, using a WMLE given in (10). For suitable estimates of $\text{cov}\left[\hat{w}_i Y_i, \hat{w}_i \hat{Y}_i\right]$ and $\text{var}[\hat{w}_i Y_i]$, see Dupuis and Victoria-Feser (2006). The effect of extreme observations on the calculation of $C_R(x_0)$ is controlled by the weights $\hat{w}_i$. The criterion is minimized over possible values for x_0. Obviously, at the minimum, we have that $Y_i \approx \hat{Y}_i$, hence $\log\left(x^*_{[i]}/x_0\right) \approx -1/\hat{\theta}\log\left[(k+1-i)/(k+1)\right]$.

For the choice of the WMLE, Dupuis and Victoria-Feser (2006) propose an estimator which downweights observations that are "far" from the Pareto model in terms of the size of the residuals with respect to the Pareto regression model, i.e.

$$w(x^*_{[i]}; \theta) = \begin{cases} 1 & \text{if } |r_i| \leq c \\ c/|r_i| & \text{if } |r_i| > c \end{cases} \tag{14}$$

with $r_i = (Y_i - \hat{Y}_i)/\sigma_i$ and c is a constant regulating the amount of robustness (for more details, see Dupuis and Victoria-Feser 2006).

In the following section, an empirical example will illustrate the method.

5 Data analysis

Let us put the semiparametric method into practice using a typical income distribution. The data for our illustration are for household disposable in-

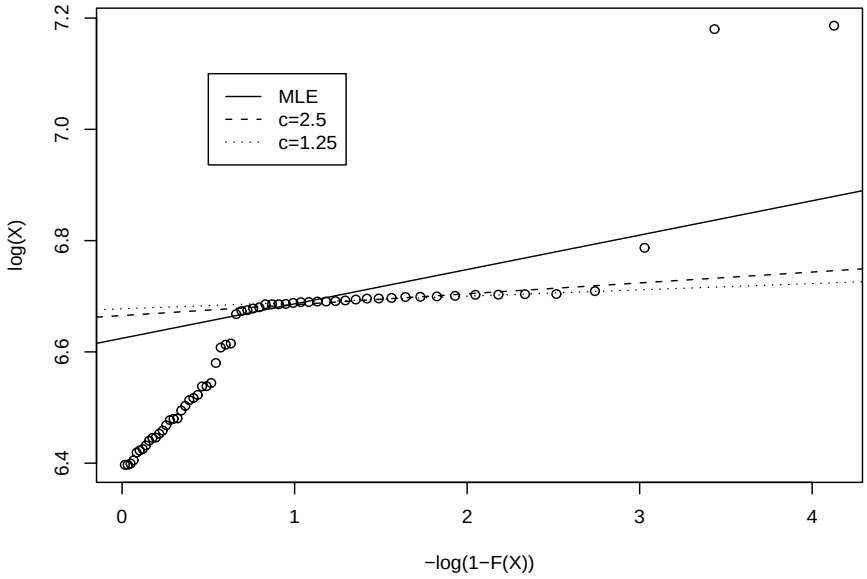

Figure 1: Pareto regression plot. Fitted regression line based on classical and robust $RC-$ criteria added. Only incomes above 600 are shown for clarity.

comes in the UK, 1981 ($n = 7470$)[4].

A Pareto quantile plot of the data together with fitted regression lines are given in Figure 1. The fits are provided by WMLE estimates with residual weights (14) for two values of c as well as the classical MLE. The optimal values for x_0 are obtained using $C_R(x_0)$ in which the weights $\hat{w}_i$ and $\hat{Y}_i$ are obtained using the different estimators. For the MLE, $\hat{w}_i = 1, \forall i$. The fit for the MLE (and hence the corresponding optimal value for x_0) are not adequate, probably because of a few very extreme observations. Both robust fits seem on the other hand appropriate. For the latter, the optimal value of x_0 corresponds to $k = 22$ selected upper incomes ($k = 32$ for the MLE). Figure 2 shows observations above the robustly selected threshold $x_0 = 803.3$ and arrows indicate the downweighted observations. The striking feature

[4]The data set is Households Below Average Income which, despite its name, actually provides a representative sample of households over the whole income range – see Department of Social Security 1992 for details.

is that not only the largest observations are downweighted, but also the smallest.

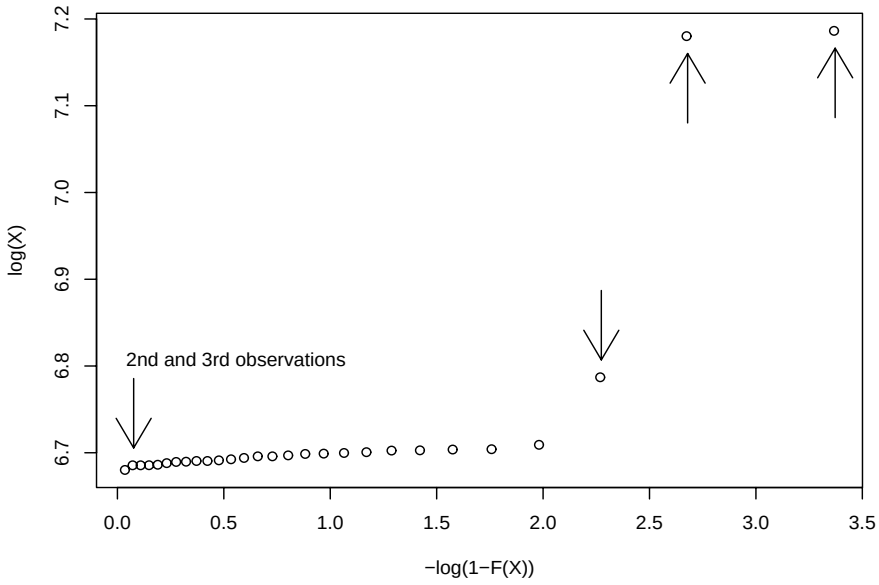

Figure 2: Pareto quantile plot of income data above robustly chosen threshold. Downweighted observations (with WMLE, $c = 1.25$) are identified.

To estimate the Pareto parameter, we hence choose $k = 22$. The value for the MLE is $\hat{\theta} = 17.5$ (with standard error 3.73) and the one for the OBRE with $c = 2^5$ is $\hat{\theta} = 76.65$ (17.62). We use these two estimates to build estimated RLC (see (7) and (8)). These curves (corresponding to the 0.5% top incomes) are presented in Figure 3 together with the empirical RLC estimate. Even if it is small, one can see a difference between the three estimates, in that the MLE follows the empirical RLC up to roughly the 0.1% of the top distribution, while the OBRE leads to an estimated RLC showing less inequality on the entire 0.5% top range.

[5] One can note that a different robust estimator is used to estimate the Pareto parameter. For the choice of k a WMLE based on residual weigths is a reasonnable choice, whereas the more efficient robust estimator (OBRE) for the Pareto parameter given a value for k is also a reasonnable choice.

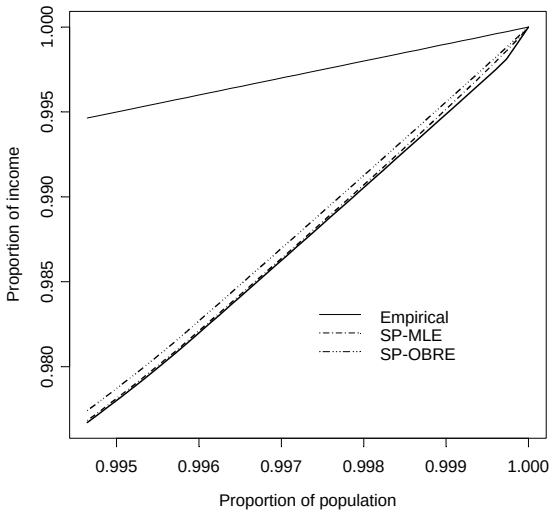

Figure 3: RLC (top 0.5%) estimates (empirical and semi-parametric with MLE and OBRE with $c = 2$) of the UK income data

6 Conclusion

Using ranking criteria to compare distributions is of immense theoretical advantage and practical convenience. In welfare economics they provide a connection between the philosophical basis of welfare judgments and elementary statistical tools for describing distributions. In practical applications they suggest useful ways in which simple computational procedures may be used to draw inferences from collections of empirical distributions. However, since it has been shown that second order rankings are not robust to data contamination, especially in the upper tail of the distribution, it is important to provide the empirical researcher with computational devices which can be used to draw inferences about the properties of distributional comparisons in a robust fashion.

One way forward might be to estimate Lorenz curves through an appropriately specified parametric model and to estimate the model parameters robustly. However, this approach is too restrictive because tractable parametric models are unlikely to be sufficiently flexible to capture some of the

essential nuances of Lorenz comparisons. For example, in order for Lorenz curves to be able to cross, a parametric model would usually need to incorporate at least three parameters, which itself may lead to serious estimation complications.

The method proposed here is a semi-parametric approach in that the upper tail of the distribution is robustly fitted using the Pareto model and a semi-parametric Lorenz curve is then built which combines non-parametric cumulative functionals and estimated ones. Simulated examples have proved not only that a few extreme data can reverse the ranking order, but also that the robust parametric Lorenz curve restores the initial ordering. Inference can be made for comparing two distributions even in the semi-parametric setting, by extending the general setting provided in Cowell and Victoria-Feser (2003). For variances too, a robust approach provides reasonable estimates when there is contamination.

Finally note that although we took the Pareto distribution as a suitable parametric model for the upper tail, and although we considered the (most common) case of a range of definition for the variable bounded below, our results can be extended to other models and/or to a two-tail modelling in a relatively straightforward manner.

References

Atkinson, A. B. (2004). Income tax and top incomes over the twentieth century. *Hacienda Pública Española 168*, 123–141.

Beirlant, J., P. Vynckier, and J. L. Teugels (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association 91*, 1651–1667.

Cowell, F. A. (2007). Inequality: Measurement. In L. Blume and S. Durlauf (Eds.), *The New Palgrave*. Basingstoke, Hampshire, UK: Palgrave Macmillan.

Cowell, F. A. and M.-P. Victoria-Feser (2002). Welfare rankings in the presence of contaminated data. *Econometrica 70*, 1221–1233.

Cowell, F. A. and M.-P. Victoria-Feser (2003). Distribution-free inference for welfare indices under complete and incomplete information. *Journal of Economic Inequality 1*, 191–219.

Cowell, F. A. and M.-P. Victoria-Feser (2006). Distributional dominance with trimmed data. *Journal of Business and Economics Statistics 24*, 291–300.

Cowell, F. A. and M.-P. Victoria-Feser (2007). Robust stochastic dominance: A semi-parametric approach. *Journal of Economic Inequality 5*, 21–37.

Dagum, C. (1977). A new model of personal income distribution: Specification and estimation. *Economie Appliquée 30*, 413–436.

Dagum, C. (1980). Generating systems and properties of income distribution models. *Metron 38*(3-4), 3–26.

Dagum, C. (1983). Income distribution models. In D. L. Banks, C. B. Read, and S. Kotz (Eds.), *Encyclopedia of Statistical Sciences*, Volume 4, pp. 27–34.

Dagum, C. (1985). Lorenz curve. In D. L. Banks, C. B. Read, and S. Kotz (Eds.), *Encyclopedia of Statistical Sciences*, Volume 5, pp. 156–161.

Department of Social Security (1992). *Households Below Average Income: A Statistical Analysis, 1979-1988/9*. London: HMSO.

Dupuis, D. J. and M.-P. Victoria-Feser (2006). A robust prediction error criterion for Pareto modeling of upper tails. *Canadian Journal of Statistics*. To appear.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events*. Applications of Mathematics: Stochastic Modelling and Applied Probability. Berlin, Heidelberg: Springer-Verlag.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley.

Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken. N.J.: John Wiley.

Kopczuk, W. and E. Saez (2004). Top wealth shares in the United States, 1916-2000: Evidence from estate tax returns. *National Tax Journal 57*, 445–487.

Lorenz, M. O. (1905). Methods for measuring concentration of wealth. *Journal of the American Statistical Association 9*, 209–219.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica 52*, 647–664.

Moriguchi, C. and E. Saez (1991). The evolution of income concentration in Japan, 1886-2002: Evidence from income tax statistics. NBER Working Paper 12558, National Bureau of Economic Research, Cambridge, Massachusetts.

Pareto, V. (1896). La courbe de la répartition de la richesse. In C. Viret-Genton (Ed.), *Recueil publié par la Faculté de Droit à l'occasion de l'Exposition nationalesuisse, Geneva 1896*, pp. 373–387. Lausanne: Université de Lausanne.

Piketty, T. (2001). *Les hauts revenus en France au 20eme siècle - Inégalités et redistributions,1901-1998*. Paris: Editions Grasset.

Piketty, T. and E. Saez (2003). Income inequality in the United States, 1913-1998. *Quarterly Journal of Economics 118*, 1–39.

Ronchetti, E. and R. G. Staudte (1994). A robust version of Mallows's C_p. *Journal of the American Statistical Association 89*, 550–559.

Saez, E. and M. Veall (2005). The evolution of high incomes in Northern America: Lessons from Canadian evidence. *American Economic Review 95*, 831–849.

Victoria-Feser, M.-P. (1995). Robust methods for personal income distribution models with application to Dagum's model. In C. Dagum and A. Lemmi (Eds.), *Research on Economic Inequality, Volume 6: Income Distribution, Social Welfare, Inequality and Poverty*, pp. 225–239. Greenwich: JAI Press.

Victoria-Feser, M.-P. and E. Ronchetti (1994). Robust methods for personal income distribution models. *Canadian Journal of Statistics 22*, 247–258.