INTERNATIONAL CENTRE FOR ECONOMIC RESEARCH
ICER

WORKING PAPER SERIES

Rolf Aaberge

ASYMPTOTIC DISTRIBUTION THEORY OF EMPIRICAL RANK-DEPENDENT MEASURES OF INEQUALITY

Working Paper No. 12/2006

Published in V. Nair (Ed.): Advances in Statistical Modeling and Inference – Essay in Honor of Kjell A. Doksum, World Scientific, 2006

Asymptotic Distribution Theory of Empirical Rank-dependent Measures of Inequality

by

Rolf Aaberge

Research Department, Statistics Norway

Abstract

A major aim of most income distribution studies is to make comparisons of income inequality across time for a given country and/or compare and rank different countries according to the level of income inequality. However, most of these studies lack information on sampling errors, which makes it difficult to judge the significance of the attained rankings.

The purpose of this paper is to derive the asymptotic properties of the empirical rank-dependent family of inequality measures. A favourable feature of this family of inequality measures is that it includes the Gini coefficients, and that any member of this family can be given an explicit and simple expression in terms of the Lorenz curve. By relying on a result of Doksum [14] it is easily demonstrated that the empirical Lorenz curve, regarded as a stochastic process, converges to a Gaussian process. Moreover, this result forms the basis of the derivation of the asymptotic properties of the empirical rank-dependent measures of inequality.

Acknowledgement

Parts of this paper were written when the author was visiting ICER in Torino. ICER is gratefully acknowledged for providing financial support and excellent working conditions. The paper is accepted for publication in a festschrift for professor Kjell Doksum. I would like to thank Dorota M. Dabrowska and an anonymous referee for helpful comments.

1. Introduction

The standard practice in empirical analyses of income distributions is to make separate comparisons of the overall level of income (the size of the cake) and the distribution of income shares (division of the cake), and to use the Lorenz curve as a basis for analysing the distribution of income shares[1]. By displaying the deviation of each individual income share from the income share that corresponds to perfect equality, the Lorenz curve captures the essential descriptive features of the concept of inequality[2].

When Lorenz curves do not intersect it is universally acknowledged that the higher Lorenz curve displays less inequality than the lower Lorenz curve. This is due to the fact that the higher of two non-intersecting Lorenz curves can be obtained from the lower Lorenz curve by means of rank-preserving income transfers from richer to poorer individuals. However, since observed Lorenz curves normally intersect weaker ranking criteria than the dominance criterion of non-intersecting Lorenz curves are required. In this case one may either search for weaker dominance criteria, see e.g. Shorrocks and Foster [31], Dardanoni and Lambert [11], Lambert [25] and Aaberge [3], or one may apply summary measures of inequality. The latter approach also offers a method for quantifying the extent of inequality in income distributions, which may explain why numerous alternative measures of inequality are introduced in the literature. The most well-known and widely used measure of inequality is the Gini coefficient, which is equal to twice the area between the Lorenz curve and its equality reference. However, to get a broader picture of inequality than what is captured by the Gini coefficient the use of alternative measures of inequality is required.

By making explicit use of the Lorenz curve Mehran [25], Donaldson and Weymark [15,16], Weymark [34], Yitzhaki [35] and Aaberge [2, 4] introduce various "generalized" Gini families of inequality measures. Moreover, Aaberge [2] demonstrates that one of these families, called the Lorenz family of inequality measures, can be considered as the moments of the Lorenz curve and thus provides a complete characterization of the Lorenz curve. This means that the Lorenz curve can be uniquely recovered from the knowledge of the corresponding Lorenz measures of inequality, i.e. without loss of information examination of inequality in an income distribution can be restricted to application of the Lorenz measures of inequality. Note that a subclass of the extended Gini family

[1] See e.g. Atkinson et al. [6] who make cross-country comparisons of Lorenz curves allowing for differences between countries in level of income and Lambert [25] for a discussion of applying Lorenz dominance criteria as basis for evaluating distributional effects of tax reforms.

[2] For a discussion of the normative aspects of Lorenz curve orderings see Kolm [23, 24, 25], Atkinson [5], Yaari [36, 37] and Aaberge [4].

introduced by Donaldson and Weymark [15,16] is uniquely determined by the Lorenz family of inequality measures[3].

Since the different alternative "generalized" families of inequality measures can be considered as subfamilies of Mehran's [26] general family of rank-dependent measures of inequality it appears useful to consider the asymptotic properties of the empirical version of the general family of rank-dependent measures of inequality rather than to restrict to the empirical version of the Lorenz family of inequality measures.

The plan of the paper is as follows. Section 2 provides formal definitions of the Lorenz curve and the family of rank-dependent measures of inequality and the corresponding non-parametric estimators. By relying on a result of Doksum [14] it is demonstrated in Section 3.1 that the empirical Lorenz curve (regarded as a stochastic process) converges to a Gaussian process. This result forms the basis of the derivation of the asymptotic properties of the empirical rank-dependent measures of inequality that are presented in Section 3.2.

2. Definition and estimation of the Lorenz curve and rank-dependent measures of inequality

Let X be an income variable with cumulative distribution function F and mean m. Let $[0,\infty\rangle$ be the domain of F where F^{-1} is the left inverse of F and $F^{-1}(0)\equiv 0$. The Lorenz curve L for F is defined by

$$L(u)=\frac{1}{m}\int_0^u F^{-1}(t)\,dt, \quad 0\le u\le 1. \tag{2.1}$$

Thus, the Lorenz curve $L(u)$ shows the share of total income received by the 100u per poorest of the population. By introducing the conditional mean function $H(\cdot)$ defined by

$$H(u)=E\left(X\,\middle|\,X\le F^{-1}(u)\right)=\frac{1}{u}\int_0^u F^{-1}(t)\,dt, \quad 0\le u\le 1, \tag{2.2}$$

Aaberge [1] found that the Lorenz curve can be written on the following form

$$L(u)=u\frac{H(u)}{H(1)} \quad 0\le u\le 1. \tag{2.3}$$

[3] See Aaberge [2] for a proof.

Let $X_1, X_2,..., X_n$ be independent random variables with common distribution function F and let F_n be the corresponding empirical distribution function. Since the parametric form of F is not known, it is natural to use the empirical distribution function F_n to estimate F and to use

$$H_n(u) = \frac{1}{u}\int_0^u F_n^{-1}(t)\,dt, \quad 0 \le u \le 1 \tag{2.4}$$

to estimate $H(u)$, where F_n^{-1} is the left inverse of F_n. Now replacing $H(u)$ by $H_n(u)$ in the expression (2.3) for $L(u)$, we get the empirical Lorenz curve

$$L_n(u) = u\frac{H_n(u)}{H_n(1)}, \quad 0 \le u \le 1. \tag{2.5}$$

To obtain an explicit expression for $H_n(u)$ and the empirical Lorenz curve, let $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$ denote the ordered $X_1, X_2,..., X_n$. For $u = i/n$ we have

$$H_n\left(\frac{i}{n}\right) = \frac{1}{i}\sum_{j=1}^{i} X_{(j)}, \quad i = 1,2,...,n \tag{2.6}$$

and

$$L_n\left(\frac{i}{n}\right) = \frac{\sum_{j=1}^{i} X_{(j)}}{\sum_{j=1}^{n} X_j}, \quad i = 1,2,...,n \tag{2.7}$$

which is the familiar estimate formula of the empirical Lorenz curve.

As mentioned in Section 1 the ranking of Lorenz curves becomes problematic when the Lorenz curves in question intersect. For this reason and to be able to quantify the inequality in distributions of income it is common to apply summary measures of inequality. As justified in Section 1 it appears attractive to consider the family of rank-dependent measures of inequality introduced by Mehran [26] and defined by

$$J_R(L) = 1 - \int_0^1 R(u)L(u)du \tag{2.8}$$

where R is a non-negative weight-function[4].

By inserting for the following two alternative subclasses R_1 and R_2 of R,

$$R_{1k}(u) = k(k+1)(1-u)^{k-1}, \quad k > 0 \tag{2.9}$$

and

$$R_{2k}(u) = (k+1)\, u^{k-1}, \quad k > 0 \tag{2.10}$$

we get the following subfamilies of the general rank-dependent family of inequality measures J_R,

$$G_k \equiv J_{R_{1k}}(L) = 1 - k(k+1)\int_0^1 (1-u)^{k-1} L(u)du, \quad k > 0 \tag{2.11}$$

and

$$D_k \equiv J_{R_{2k}}(L) = 1 - (k+1)\int_0^1 u^{k-1} L(u)du, \quad k > 0. \tag{2.12}$$

Note that $\{G_k : k > 0\}$ was denoted the extended Gini family and $\{D_k : k > 0\}$ the "illfare-ranked single series Ginis" by Donaldson and Weymark [15][5]. However, as mentioned in Section 1 Aaberge [2] proved that each of the subfamilies $\{D_k : k = 1,2,...\}$ (denoted the Lorenz family of inequality measures) and $\{G_k : k = 1,2,...\}$ provides a complete characterization of the Lorenz curve, independent of whether the distribution function F is defined on a bounded interval or not. Thus, any distribution function F defined on P^+ can be specified by its mean and Lorenz measures of inequality even if some of the conventional moments do not exist.

It follows directly from expressions (2.11) and (2.12) that the Gini coefficient defined by

$$G = 1 - 2\int_0^1 L(u)du \tag{2.13}$$

is included in the extended Gini family as well as in the Lorenz family of inequality measures.

By replacing L by L_n in the expression (2.8) for J_R, we get the following estimator of J_R,

---

[4] A slightly different version of J_R was introduced by Piesch [27], whereas Giaccardi [18] considered a discrete version of J_R. For alternative normative motivations of the J_R-family and various subfamilies of the J_R-family we refer to Donaldson and Weymark [16], Yaari [36,37], Ben Porath and Gilboa [7] and Aaberge [4]. See also Zitikis [39] and Tarsitano [33] for a discussion on related families of inequality measures.

[5] See Zitikis and Gastwirth [41] for a derivation of the asymptotic distribution of the empirical extended Gini family of inequality measures.

$$\hat{J}_R \equiv J_R\left(L_n\right) = 1 - \int_0^1 R(u)\, L_n(u)du \,. \qquad (2.14)$$

For $R(u) = 2$, (2.14) gives the estimator[6] of *G*,

$$\hat{G} = 1 - 2\int_0^1 L_n(u)du = 1 - \frac{2\sum_{i=1}^{n}\sum_{j=1}^{i} X_{(j)}}{(n+1)\sum_{j=1}^{n} X_j}\,. \qquad (2.15)$$

3. Asymptotic distribution theory of the empirical Lorenz curve and empirical rank-dependent measures of inequality

As demonstrated by expressions (2.8) and (2.14), the rank-dependent measures of inequality and their empirical counterparts are explicitly defined in terms of the Lorenz curve and its empirical counterpart, respectively. Thus, in order to derive the asymptotic distribution of the empirical rank-dependent measures of inequality it is convenient to firstly derive the asymptotic properties of the empirical Lorenz curve. To this end we utilize the close formal connection between the shift function of Doksum [13] and the Lorenz curve.

As an alternative to the approach chosen in this paper we can follow Zitikis [39] by expressing the rank-dependent measures of inequality in terms of L-statistics and rely on asymptotic distribution results for L-statitcs[7]. Note that Csörgø, Gastwirth and Zitikis [10] have derived asymptotic confidence bands for the Lorenz and the Bonferroni curves without requiring the existence of the density f. Moreover, Davydov and Zitikis [12,13] have considered the case where observations are allowed to be dependent. As demonstarted by Zitikis [38] note that the Vervaat process proves to be a particularly helpful device in deriving asymptotic properties of various aggregates of empirical quantiles.

3.1. Asymptotic properties of the empirical Lorenz curve

Since F_n is a consistent estimate of *F*, $H_n(u)$ and $L_n(u)$ are consistent estimates of $H(u)$ and $L(u)$, respectively.

---

[6] The asymptotic properties of the empirical Gini coefficient has been considered by Hoeffding [21], Goldie [19], Aaberge [1], Zitikis [39, 40] and Zitikis and Gastwirth [41].

Approximations to the variance of L_n and the asymptotic properties of L_n can be obtained by considering the limiting distribution of the process $Z_n(u)$ defined by

$$Z_n(u) = n^{\frac{1}{2}}\left[L_n(u) - L(u)\right]. \tag{3.1}$$

In order to study the asymptotic behavior of $Z_n(u)$ we find it useful to start with the process $Y_n(u)$ defined by

$$Y_n(u) = n^{\frac{1}{2}}\left[H_n(u) - H(u)\right] = \frac{1}{u}\int_0^u n^{\frac{1}{2}}\left(F_n^{-1}(t) - F^{-1}(t)\right)dt. \tag{3.2}$$

Assume that the support of F is a non-empty finite interval $[a,b]$. (When F is an income distribution, a is commonly equal to zero.) Then $Y_n(u)$ and $Z_n(u)$ are members of the space D of functions on [0,1] which are right continuous and have left hand limits. On this space we use the Skorokhod topology and the associated σ-field (e.g. Billingsley [8], page 111). We let $W_0(t)$ denote a Brownian Bridge on [0,1], that is, a Gaussian process with mean zero and covariance function $s(1-t)$, $0 \leq s \leq t \leq 1$.

THEOREM 3.1: Suppose that F has a continuous nonzero derivate f on $[a,b]$. Then $Y_n(u)$ converges in distribution to the process

$$Y(u) = \frac{1}{u}\int_0^u \frac{W_0(t)}{f\left(F^{-1}(t)\right)}dt. \tag{3.3}$$

PROOF: It follows directly from Theorem 4.1 of Doksum [14] that

$$n^{\frac{1}{2}}\left(F_n^{-1}(t) - F^{-1}(t)\right)$$

converges in distribution to the Gaussian process $W_0(t)/f\left(F^{-1}(t)\right)$.

[7] On general results for L-statistics see e.g. Chernoff et al. [9], Shorack [30], Stigler [32] and Serfling [29].

Using the arguments of Durbin ([17], section 4.4), we find that $Y(u)$ as a function of $\left(W_0(t) \big/ f\left(F^{-1}(t)\right)\right)$ is continuous in the Skorokhod topology. The result then follows from Billingsley ([8], Theorem 5.1).

Q.E.D.

The following result states that $Y(u)$ is a Gaussian process and thus that $Y_n(u)$ is asymptotically normally distributed, both when considered as a process, and for fixed u.

THEOREM 3.2: Suppose the conditions of Theorem 1 are satisfied. Then the process $uY(u)$ has the same probability distribution as the Gaussian process

$$\sum_{j=1}^{\infty} q_j(u) Z_j$$

where $q_j(u)$ is given by

$$q_j(u) = \frac{2^{\frac{1}{2}}}{j\pi} \int_0^u \frac{\sin(j\pi t)}{f\left(F^{-1}(t)\right)} dt \tag{3.4}$$

and $Z_1, Z_2, \ldots$ are independent $N(0,1)$ variables.

PROOF: Put

$$V_N(t) = \frac{2^{\frac{1}{2}}}{f\left(F^{-1}(t)\right)} \sum_{j=1}^{N} \frac{\sin(j\pi t)}{j\pi} Z_j$$

and note that

$$2\sum_{j=1}^{\infty} \frac{\sin(j\pi s)\sin(j\pi t)}{(j\pi)^2} = s(1-t), \quad 0 \le s \le t \le 1. \tag{3.5}$$

Thus, the process $V_N(t)$ is Gaussian with mean zero and covariance function

$$\operatorname{cov}\left(V_N(s), V_N(t)\right) = \frac{2}{f\left(F^{-1}(s)\right) f\left(F^{-1}(t)\right)} \sum_{j=1}^{N} \frac{\sin(j\pi s)\sin(j\pi t)}{(j\pi)^2} \to \operatorname{cov}\left(V(s), V(t)\right),$$

where

$$V(t)=\frac{W_0(t)}{f\left(F^{-1}(t)\right)}.$$

In order to prove that $V_N(t)$ converges in distribution to the Gaussian process $V(t)$, it is, according to Hajek and Sidak ([20], Theorem 3.1.a, Theorem 3.1.b, Theorem 3.2) enough to show that

$$E\left[V_N(t)-V_N(s)\right]^4 \le M\left(t-s\right)^2,\quad 0\le s,t\le 1,$$

where M is independent of N.

Since for normally distributed random variables with mean 0,

$$EX^4=3\left[EX^2\right]^2,$$

we have

$$\begin{aligned}
&E\left[V_N(t)-V_N(s)\right]^4=3\left[\operatorname{var}\left(V_N(t)-V_N(s)\right)\right]^2\\
&=3\left\{2\operatorname{var}\left[\sum_{j=1}^{N}\frac{1}{jp}\left(\frac{\sin\left(jp\,t\right)}{f\left(F^{-1}(t)\right)}-\frac{\sin\left(jp\,s\right)}{f\left(F^{-1}(s)\right)}\right)Z_j\right]\right\}^2\\
&=3\left\{2\sum_{j=1}^{N}\left[\frac{1}{jp}\left(\frac{\sin\left(jpt\right)}{f\left(F^{-1}(t)\right)}-\frac{\sin\left(jp\,s\right)}{f\left(F^{-1}(s)\right)}\right)\right]^2\right\}^2\le 3\left\{2\sum_{j=1}^{\infty}\left[\frac{1}{jp}\left(\frac{\sin\left(jp\,t\right)}{f\left(F^{-1}(t)\right)}-\frac{\sin\left(jp\,s\right)}{f\left(F^{-1}(s)\right)}\right)\right]^2\right\}^2\\
&=3\left\{\frac{t\left(1-t\right)}{f^2\left(F^{-1}(t)\right)}+\frac{s\left(1-s\right)}{f^2\left(F^{-1}(s)\right)}-2\frac{\operatorname{cov}\left(W_0(s),W_0(t)\right)}{f\left(F^{-1}(s)\right)f\left(F^{-1}(t)\right)}\right\}^2.
\end{aligned}$$

Since $0< f(x)<\infty$ on $[a,b]$, there exists a constant M such that

$$f\left(F^{-1}(t)\right)\ge M^{-\frac{1}{4}}\ \text{ forall }\ t\in[0,1].$$

Then

$$E\left[V_N(t)-V_N(s)\right]^2\le 3M\left(t-s\right)^2\left(1-\left|t-s\right|\right)^2\le 3M\left(t-s\right)^2.$$

Hence $V_N(t)$ converges in distribution to the process $V(t)$. Thus, according to Billingsley ([8], Theorem 5.1)

$$\int_0^u V_N(t)\,dt = \sum_{j=1}^{N} q_j(u) Z_j$$

converges in distribution to the process

$$\int_0^u V(t)\,dt = \int_0^u \frac{W_0(t)}{f\left(F^{-1}(t)\right)}\,dt = uY(u)\,.$$

Q.E.D.

Now, let h_j be a function defined by

$$h_j(u) = \frac{1}{m}\left[q_j(u) - q_j(1) L(u) \right] \tag{3.6}$$

where $q_j(u)$ is given by (3.4).

THEOREM 3.3: Suppose the conditions of Theorem 3.1 are satisfied. Then $Z_n(u)$ given by (3.1) converges in distribution to the Gaussian process

$$Z(u) = \sum_{j=1}^{\infty} h_j(u) Z_j \tag{3.7}$$

where $Z_1, Z_2, \ldots$ are independent $N\left(0,1\right)$ variables and $h_j(u)$ is given by (3.6).

PROOF: By combining (2.5), (3.1) and (3.2) we see that

$$Z_n(u) = \frac{1}{H_n(1)}\left[uY_n(u) - L(u) Y_n(1) \right]$$

where $Y_n(u)$ is given by (3.2).

Now, Theorem 3.1 implies that the process

$$uY_n(u) - L(u) Y_n(1)$$

converges in distribution to the process

$$uY(u) - L(u) Y(1)$$

where $Y(u)$ is given by (3.3). Then, since $H_n(1)$ converges in probability to m, Cramer-Slutsky's theorem gives that $Z_n(u)$ converges in distribution to the process

$$\frac{1}{m}\left[uY(u)-L(u)Y(1)\right].$$

Thus, by applying Theorem 3.2 the proof is completed.

Q.E.D.

In order to derive the asymptotic covariance functions of the processes $Y_n(u)$ and $Z_n(u)$, the following lemma is needed.

LEMMA 3.1: Suppose the conditions of Theorem 1 are satisfied. Then

$$\sum_{i=1}^{\infty} q_i(u)q_i(v)=t^2(u)+l\left(u,v\right),\quad 0\le u\le v\le 1, \tag{3.8}$$

where $q_i(u)$ is defined by (3.4) and $t^2(u)$ and $l\left(u,v\right)$ are given by

$$t^2(u)=2\int_a^{F^{-1}(u)}\int_a^{y} F(x)\left(1-F(y)\right)dxdy,\quad 0\le u\le 1 \tag{3.9}$$

and

$$l\left(u,v\right)=\int_{F^{-1}(u)}^{F^{-1}(v)}\int_a^{F^{-1}(u)} F(x)\left(1-F(y)\right)dxdy,\quad 0\le u\le v\le 1. \tag{3.10}$$

PROOF: Assume that $0\le u\le v\le 1$. From the definition of $q_i(u)$ we have that

$$\sum_{i=1}^{\infty} q_i(u)q_i(v)=\sum_{i=1}^{\infty}\int_0^v\int_0^u\left[\frac{2}{f\left(F^{-1}(t)\right)f\left(F^{-1}(s)\right)}\frac{\sin\left(ipt\right)\sin\left(ip\,s\right)}{\left(ip\right)^2}\right]dtds\,.$$

By applying Fubini's theorem (e.g. Royden [28]) and the identity (3.5) we get

$$\sum_{i=1}^{\infty} q_i(u)q_i(v) = \int_0^v \int_0^u \left[\frac{2}{f\left(F^{-1}(t)\right) f\left(F^{-1}(s)\right)} \sum_{i=1}^{\infty} \frac{\sin\left(ip\,t\right)\sin\left(ip\,s\right)}{\left(ip\right)^2} \right] dtds$$

$$= 2\int_0^u \int_0^s \frac{t\left(1-s\right)}{f\left(F^{-1}(t)\right) f\left(F^{-1}(s)\right)} dtds + \int_u^v \int_0^u \frac{t\left(1-s\right)}{f\left(F^{-1}(t)\right) f\left(F^{-1}(s)\right)} dtds$$

$$= 2\int_a^{F^{-1}(u)} \int_a^{y} \left(F(x)\left(1-F(y)\right) dxdy\right) + \int_{F^{-1}(u)}^{F^{-1}(v)} \int_a^{F^{-1}(u)} F(x)\left(1-F(y)\right) dxdy = t^2(u) + l\left(u,v\right).$$

Q.E.D.

As an immediate consequence of Theorem 3.1, Theorem 3.2 and Lemma 3.1 we have the following corollary.

COROLLARY 3.1: Under the conditions of Theorem 1, $Y_n(u)$ has asymptotic covariance function $q^2\left(u,v\right)$ given by

$$q^2\left(u,v\right) = \frac{1}{uv}\left[t^2(u) + l\left(u,v\right)\right], \quad 0 < u \le v \le 1. \tag{3.11}$$

From Theorem 3.3 and Lemma 3.1 we get the next corollary.

COROLLARY 3.2: Under the conditions of Theorem 3.1, $Z_n(u)$ has asymptotic covariance function $n^2\left(u,v\right)$ given by

$$\begin{aligned} n^2\left(u,v\right) = \frac{1}{m^2}\Big[& t^2(u) + l\left(u,v\right) - L(u)\left(t^2(v) + l\left(v,1\right)\right) \\ & - L(v)\left(t^2(u) + l\left(u,1\right)\right) + L(u)L(v)t^2(1)\Big], \quad 0 < u \le v \le 1. \end{aligned} \tag{3.12}$$

In order to construct confidence intervals for the Lorenz curve at fixed points, we apply the results of Theorem 3.3 and Corollary 3.2 which imply that the distribution of

$$n^{\frac{1}{2}} \frac{L_n(u) - L(u)}{n\left(u,u\right)}$$

tends to the $N\left(0,1\right)$ distribution for fixed u, where $n^2\left(u,u\right)$ is given by

$$n^2\left(u,u\right) = \frac{1}{m^2}\left[t^2(u) - 2L(u)\left(t^2(u) + l\left(u,1\right)\right) + L^2(u)t^2(1)\right], \quad 0 < u \le 1. \tag{3.13}$$

Before this result can be applied, we must estimate the asymptotic variance $n^2(u,u)$, i.e., we must estimate m, L, t^2 and l. The estimates of m and L are given by $\overline{X}$ and (2.7), respectively. Now, by introducing the statistics a_k and b_k defined by

$$a_k = \left(1 - \frac{k}{n}\right)\left(X_{(k+1)} - X_{(k)}\right) \tag{3.14}$$

and

$$b_k = \frac{k}{n}\left(X_{(k+1)} - X_{(k)}\right), \tag{3.15}$$

we obtain the following consistent estimates of t^2 and l,

$$\hat{t}^2\left(\frac{i}{n}\right) = 2\sum_{k=1}^{i-1}\left(a_k \sum_{l=1}^{k} b_l\right), \quad i = 2,3,\ldots, n \tag{3.16}$$

and

$$\hat{l}\left(\frac{i}{n},\frac{j}{n}\right) = \left(\sum_{k=i}^{j-1} a_k\right)\left(\sum_{l=1}^{i-1} b_l\right), \quad i = 2,3,\ldots,n-1;\ j \geq i+1 . \tag{3.17}$$

Thus, replacing m, L, t^2 and l by their respective estimates in the expression (3.13) for n^2 we obtain a consistent estimate of n^2.

To get an idea of how reliable $L_n(u)$ is as an estimate for $L(u)$, we have to construct a confidence band based on $L_n(u)$ and $L(u)$. Such a confidence band can be obtained from statistics of the type

$$K_n = n^{\frac{1}{2}} \sup_{0 \leq u \leq 1} \frac{\left|L_n(u) - L(u)\right|}{y\left(L_n(u)\right)} \tag{3.18}$$

where y is a continuous nonnegative weight function. By applying Theorem 3.3 and Billingsley ([8], Theorem 5.1), we find that K_n converges in distribution to

$$K = \sup_{0 \leq u \leq 1}\left|\sum_{j=1}^{\infty} \frac{h_j(u)}{y\left(L(u)\right)} Z_j\right| . \tag{3.19}$$

Let

$$T_m(u) = \sum_{j=1}^{m} \frac{h_j(u)}{y\left(L(u)\right)} Z_j \,, \tag{3.20}$$

$$T(u) = \sum_{j=1}^{\infty} \frac{h_j(u)}{y\left(L(u)\right)} Z_j \tag{3.21}$$

and

$$K'_m = \sup_{0 \le u \le} \left| T_m(u) \right| . \tag{3.22}$$

Since T_m converges in distribution to T, we find by applying Billingsley ([8], Theorem 5.1) that K'_m converges in distribution to K. Hence, for a suitable choice of m and y , for instance $y = 1$, simulation methods may be used to obtain the distribution of K'_m and thus an approximation for the distribution of K.

3.2. Asymptotic properties of the empirical rank-dependent family of inequality measures

We shall now study the asymptotic distribution of the statistics $\hat{J}_R$ given by (2.14). Mehran [26] states without proof that $n^{\frac{1}{2}}\left(\hat{J}_R - J_R\right)$ is asymptotically normally distributed with mean zero. The asymptotic variance, however, cannot be derived, as maintained by Mehran [26], from Stigler [32], Theorem 3.1)[8]. However, as will be demonstrated below Theorem 3.3 forms a helpful basis for deriving the asymptotic variance of $\hat{J}_R$.

Let w^2 be a parameter defined by

$$\begin{aligned} w^2 = \frac{1}{m^2}\Bigg\{ & 2\int_0^1 \int_0^v \left[t^2(u) + l\left(u,v\right) \right] R(u)R(v)\,du\,dv \\ & - 2\left[\int_0^1 uR(u)du - J_R \right]\left[\int_0^1 \left(t^2(u) + l\left(u,1\right)\right) R(u)du \right] + t^2(1)\left[\int_0^1 uR(u)du - J_R \right]^2 \Bigg\}. \end{aligned} \tag{3.23}$$

THEOREM 3.4: Suppose the conditions of Theorem 1 are satisfied and $w^2 < \infty$. Then the distribution of

[8] See also Zitikis and Gastwirth [41] on the asymptotic estimation of the S-Ginis, Zitikis [40] on the asymptotic estimation of the E-Gini index and a more general discussion in Davydov and Zitikis [13].

$$n^{\frac{1}{2}}\left(\hat{J}_R - J_R\right)$$

tends to the normal distribution with zero mean and variance w^2.

PROOF: From (2.8), (2.14) and (3.1) we see that

$$n^{\frac{1}{2}}\left(\hat{J}_R - J_R\right) = -\int_0^1 R(u) Z_n(u) du .$$

By Theorem 3.3 we have that $Z_n(u)$ converges in distribution to the Gaussian process $Z(u)$ defined by (3.7). By applying Billingsley ([8], Theorem 5.1) and Fubini's theorem we get that $n^{\frac{1}{2}}\left(\hat{J}_R - J_R\right)$ converges in distribution to

$$-\int_0^1 R(u)Z(u)du = -\int_0^1 R(u)\left(\sum_{j=1}^{\infty} h_j(u) Z_j\right)du = -\sum_{j=1}^{\infty}\left[\int_0^1 R(u) h_j(u)\, du\right] Z_j$$

where $Z_1, Z_2, \ldots$ are independent $N\left(0,1\right)$ variables and $h_j(u)$ is given by (3.6), i.e., the asymptotic distribution of $n^{\frac{1}{2}}\left(\hat{J}_R - J_R\right)$ is normal with mean zero and variance

$$\sum_{j=1}^{\infty}\left[\int_0^1 R(u) h_j(u) du\right]^2 . \tag{3.24}$$

Then it remains to show that the asymptotic variance is equal to w^2.

Inserting (3.6) in (3.24), we get

$$\sum_{j=1}^{\infty}\left[\int_0^1 R(u)\, h_j(u) du\right]^2 = \frac{1}{m^2}\sum_{j=1}^{\infty}\left[\int_0^1 R(u)\left(q_j(u) - q_j(1) L(u)\right) du\right]^2$$

$$= \frac{1}{m^2}\left\{\sum_{j=1}^{\infty}\left[\int_0^1 R(u) q_j(u) du\right]^2 - 2\left[\int_0^1 R(u) L(u) du\right]\left[\sum_{j=1}^{\infty} q_j(1)\int_0^1 R(u) q_j(u) du\right]\right.$$

$$\left. + \left[\sum_{j=1}^{\infty} q_j^2(1)\right]\left[\int_0^1 R(u) L(u) du\right]^2\right\}.$$

In the following derivation we apply Fubini's theorem and the identity (3.5),

$$\sum_{j=1}^{\infty}\left[\int_0^1 R(u)q_j(u)du\right]^2 = \sum_{j=1}^{\infty}\int_0^1\int_0^1 R(u)q_j(u)R(v)q_j(v)dudv$$

$$= \int_0^1\int_0^1\left[\int_0^v\int_0^u \frac{2}{f\left(F^{-1}(t)\right)f\left(F^{-1}(s)\right)}\left(\sum_{j=1}^{\infty}\frac{\sin(jp\,t)\sin(jp\,s)}{(jp)^2}\right)dtds\right]R(u)R(v)dudv$$

$$= 2\int_0^1\int_0^v\left[2\int_0^u\int_0^s \frac{t(1-s)}{f\left(F^{-1}(t)\right)f\left(F^{-1}(s)\right)}dtds + \int_u^v\int_0^u \frac{t(1-s)}{f\left(F^{-1}(t)\right)f\left(F^{-1}(s)\right)}dtds\right]R(u)R(v)dudv$$

$$= 2\int_0^1\int_0^v\left[2\int_a^{F^{-1}(u)}\int_a^y F(x)(1-F(y))dxdy + \int_{F^{-1}(u)}^{F^{-1}(v)}\int_a^{F^{-1}(u)} F(x)(1-F(y))dxdy\right]R(u)R(v)dudv$$

$$= 2\int_0^1\int_0^v \left[t^2(u) + l(u,v)\right]R(u)R(v)dudv$$

where $t^2(u)$ and $l(u,v)$ are given by (3.9) and (3.10), respectively. Similarly, we find that

$$\sum_{j=1}^{\infty} q_j(1)\int_0^1 R(u)q_j(u)du = \int_0^1 \left[t^2(u) + l(u,1)\right]R(u)du .$$

From Lemma 3.1 it follows that

$$\sum_{j=1}^{\infty} q_j(1) = t^2(1) .$$

Finally, by noting that

$$\int_0^1 R(u)L(u)du = \int_0^1 uR(u)du - J_R ,$$

the proof is completed.

Q.E.D.

For $R(u) = 2$, Theorem 3.4 states that $w^2 = g^2$, where g^2 is defined by

$$g^2 = \frac{4}{m^2}\left\{2\int_0^1\int_0^v \left[t^2(u) + l(u,v)\right]dudv - (1-G)\int_0^1 \left[t^2(u) + l(u,1)\right]du + \frac{1}{4}(1-G)^2 t^2(1)\right\}, \quad (3.25)$$

is the asymptotic variance of $n^{\frac{1}{2}}\hat{G}$.

The estimation of g^2 is straightforward. As in Section 2 we assume that the parametric form of F is not known. Thus, replacing F by the empirical distribution function F_n in expression (4.1) for g^2, we obtain a consistent nonparametric estimator for g^2. The current estimator is given by

$$\hat{g}^2 = \frac{4}{\bar{X}^2}\left\{\frac{2}{n^2}\sum_{j=2}^{n}\sum_{i=2}^{j}\hat{t}^2\left(\frac{i}{n}\right)+\frac{2}{n^2}\sum_{j=3}^{n}\sum_{i=2}^{j-1}\hat{l}\left(\frac{i}{n},\frac{j}{n}\right)\right.$$
$$\left.-\frac{1}{n}\left(1-\hat{G}\right)\left[\sum_{i=2}^{n}\hat{t}^2\left(\frac{i}{n}\right)+\sum_{i=2}^{n-1}\hat{l}\left(\frac{i}{n},1\right)\right]+\frac{1}{4}\left(1-\hat{G}\right)^2\hat{t}^2(1)\right\} \quad (3.26)$$

where $\hat{t}^2$, $\hat{l}$ and $\hat{G}$ are given by (3.16), (3.17) and (2.15), respectively.

Similarly, a consistent estimator for w^2 is obtained by replacing t^2, l, m and I by their respective estimates in the expression (3.23) for w^2.

References

[1] Aaberge, R. (1982): On the problem of measuring inequality. (In Norwegian). Rapporter 82/9, Statistics Norway.

[2] Aaberge, R. (2000): Characterizations of Lorenz curves and income distributions. *Social Choice and Welfare* **17**, 639-653.

[3] Aaberge, R. (2000): "Ranking intersecting Lorenz curves", Discussion Paper 271, Statistics Norway.

[4] Aaberge, R. (2001): Axiomatic characterization of the Gini coefficient and Lorenz curve orderings. *Journal of Economic Theory* **101**, 115-132.

[5] Atkinson, A.B. (1970): On the measurement of inequality, *Journal of Economic Theory*, **2**, 244-263.

[6] Atkinson, A.B., L. Rainwater and T. Smeeding (1995): Income distribution in OECD countries: The evidence from the Luxembourg Income Study (LIS), Social Policy Studies No. 18, OECD, Paris.

[7] Ben Porath, E. and I. Gilboa (1994): Linear measures, the Gini index, and the income-equality trade-off. *Journal of Economic Theory* **64**, 443-467.

[8] Billingsley, P. (1968): *Convergence of Probability Measures*. John Wiley & Sons, Inc., New York.

[9] Chernoff, H., J. L. Gastwirth and M. V. Johns (1967): Asymptotic distribution of linear combinations of functions of order statistics with application to estimation. *Annals of Mathematical Statistics*, **38**, 52-72.

[10] Csörgø, M., J. L. Gastwith and R. Zitikis (1998): Asymptotic confidence bands for the Lorenz and Bonferroni curves based on the empirical Lorenz curve, *Journal of Statistical Planning and Inference*, **74**, 65 - 91.

[11] Dardanoni, V. and P.J. Lambert (1988): Welfare rankings of income distributions: A role for the variance and some insights for tax reforms. *Social Choice and Welfare* **5**, 1-17.

[12] Davydov, Y. and R. Zitikis (2003): Generalized Lorenz curves and convexifications of stochastic processes, *Journal of Applied Probability*, **40**, 906 - 925.

[13] Davydov, Y. and R. Zitikis (2004): Convex rearrangements of random elements, *Asymptotic Methods in Stochastics*, Fields Institute communications, American Mathematical Society, **44**, 141 - 171.

[14] Doksum, K. (1974): Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The Annals of Statistics* **2** (2), 267-277.

[15] Donaldson, D. and J.A. Weymark (1980): A single parameter generalization of the Gini indices of inequality. *Journal of Economic Theory* **22**, 67-86.

[16] Donaldson, D. and J.A. Weymark (1983): Ethically flexible indices for income distribution in the continuum. *Journal of Economic Theory* **29**, 353-358.

[17] Durbin, J. (1973): *Distribution Theory for Tests Based on the Sample Distribution Function*. Society for Industrial and Applied Mathematics, Philadelphia.

[18] Giaccardi, F. (1950): Un criterio per la construzione di indici di conzentrazione. *Rivista Italiana di Demografia e Statistica* **4**, 527-538.

[19] Goldie, C.M. (1977).Convergence theorems for empirical Lorenz curves and their inverses. *Advances in Applied Probability 9*, 765-791.

[20] Hajek, J. and Z. Sidak (1967): *Theory of Rank Tests*. Academic Press, New York.

[21] Hoeffding, W. (1948): A Class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19**, 293-325.

[22] Kolm, S.Ch. (1969): The optimal production of social justice. In J. Margolis and H. Guitton (eds.): *Public Economics*. Macmillan, New York/London.

[23] Kolm, S.Ch. (1976): Unequal inequalities I, *Journal of Economic Theory* **12**, 416-442.

[24] Kolm, S.Ch. (1976): Unequal inequalities II, *Journal of Economic Theory* **13**, 82-111.

[25] Lambert, P.J. (1993): *The Distribution and Redistribution of Income: A Mathematical Analysis*. Manchester University Press, Manchester.

[26] Mehran, F. (1976): Linear measures of inequality, *Econometrica* **44**, 805-809.

[27] Piesch, W. (1975): *Statistische Konzentrationsmasse*. Mohr (Siebeck), Tübingen.

[28] Royden, H. (1963): *Real Analysis*. Macmillan, New York.

[29] Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc., New York.

[30] Shorack, G.R. (1972): Functions of order statistics. *The Annals of Mathematical Statistics* **43** (2), 412-427.

[31] Shorrocks, A.F. and J.E. Foster (1978): Transfer sensitive inequality measures. *Review of Economic Studies* **14**, 485-497.

[32] Stigler, S.M. (1974): Linear functions of order statistics with smooth weight functions. *The Annals of Statistics* **2** (4), 676-693.

[33] Tarsitano, A. (2004): A new class of inequality measures based on a ratio of L- Statistics, *Metron*, LXII, 137 - 160.

[34] Weymark, J. (1981): Generalized Gini inequality indices. *Mathematical Social Science* **1**, 409-430.

[35] Yitzhaki, S. (1983): On an extension of the Gini inequality index. *International Economic Review* **24**, 617-628.

[36] Yaari, M.E. (1987): The dual theory of choice under risk. *Econometrica* **55**, 95-115.

[37] Yaari, M.E. (1988): A controversial proposal concerning inequality measurement. *Journal of Economic Theory* **44**, 381-397.

[38] Zitikis, R. (1998): The Vervaat process. In B. Szyszkowicz (ed): *Asymptotic Methods in Probability and Statistics*. North-Holland, Amsterdam.

[39] Zitikis, R. (2002): Large Sample Estimation of a Family of Economic Inequality Indices, *Pakistan Journal of Statistics* Special issue in honour of Dr. S. Ejaz Ahmad, **18**, 225 - 248.

[40] Zitikis, R. (2003): Asymptotic estimation of the E - Gini index, *Econometric Theory*, **19**, 587-601.

[41] Zitikis, R. and J. T. Gastwirth (2002): The asymptotic distribution of the S-Gini index, *Autralian and New Zealand Journal of Statistics*, **44**, 439 - 446.