

Discussion Papers No. 402, January 2005
Statistics Norway, Research Department

Rolf Aaberge

Asymptotic Distribution Theory of Empirical Rank-dependent Measures of Inequality

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 (1974) 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.

Keywords: The Lorenz curve, the Gini coefficient, rank-dependent measures of inequality, nonparametric estimation methods, asymptotic distribution theory.

JEL classification: C14, D 63

Acknowledgement: I would like to thank Anne Skoglund for technical assistance and word processing.

Address: Rolf Aaberge, Statistics Norway, Research Department. E-mail: rolf.aaberge@ssb.no

Discussion Papers

comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:

<http://www.ssb.no>

<http://ideas.repec.org/s/ssb/dispap.html>

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
NO-2225 Kongsvinger

Telephone: +47 62 88 55 00

Telefax: +47 62 88 55 95

E-mail: Salg-abonnement@ssb.no

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. 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 inequality¹.

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 (1987), Dardanoni and Lambert (1988), Lambert (1993) and Aaberge (2000b), 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 (1976), Donaldson and Weymark (1980, 1983), Weymark (1981), Yitzhaki (1983) and Aaberge (2000a, 2001) introduce various “generalized” Gini families of inequality measures. Moreover, Aaberge (2000a) 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 introduced by Donaldson and Weymark (1980, 1983) is uniquely determined by the Lorenz family of inequality measures².

¹ For a discussion of the normative aspects of Lorenz curve orderings see Kolm (1969, 1976a, 1976b), Atkinson (1970), Yaari (1987, 1988) and Aaberge (2001).

² See Aaberge (2000a) for a proof.

Since the different alternative “generalized” families of inequality measures can be considered as subfamilies of Mehran’s (1976) 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 (1974) 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 μ . Let $[0, \infty)$ 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

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

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

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

Aaberge (1982) found that the Lorenz curve can be written on the following form

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

Let X_1, X_2, \dots, 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

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

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

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

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

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

and

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

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 (1976) and defined by

$$(2.8) \quad J_R(L) = 1 - \int_0^1 R(u) L(u) du$$

where R is a non-negative weight-function³.

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

³ A slightly different version of J_R was introduced by Piesch (1975), whereas Giaccardi (1950) 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 (1981), Yaari (1987, 1988), Ben Porath and Gilboa (1984) and Aaberge (2001).

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

and

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

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

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

and

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

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 (1980). However, as mentioned in Section 1 Aaberge (2000a) proved that each of the subfamilies $\{D_k : k = 1, 2, \dots\}$ (denoted the Lorenz family of inequality measures) and $\{G_k : k = 1, 2, \dots\}$ 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 \mathbb{R}^+ 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

$$(2.13) \quad G = 1 - 2 \int_0^1 L(u) du$$

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 ,

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

For $R(u) = 2$, (2.14) gives the estimator of G ,

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

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 (1974) and the Lorenz curve⁴.

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.

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

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

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

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

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

⁴ We refer to Goldie (1977) for an alternative proof of the asymptotic properties of the empirical Lorenz curve. Note, however, that Goldie did not observe the close connection between the Lorenz curve and the shift function of Doksum (1974) and thus use a more complex approach for proving the large sample properties of the empirical Lorenz curve.

Skorokhod topology and the associated σ -field (e.g. Billingsley (1968), 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 1: Suppose that F has a continuous nonzero derivate f on $[a,b]$. Then $Y_n(u)$ converges in distribution to the process

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

PROOF: It follows directly from Theorem 4.1 of Doksum (1974) that

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

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

Using the arguments of Durbin (1973, section 4.4), we find that $Y(u)$ as a function of $(W_0(t)/f(F^{-1}(t)))$ is continuous in the Skorokhod topology. The result then follows from Billingsley (1968, 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 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

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

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

PROOF: Put

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

and note that

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

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

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

where

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

In order to prove that $V_N(t)$ converges in distribution to the Gaussian process $V(t)$, it is, according to Hajek and Sidak (1967, Theorem 3.1.a, Theorem 3.1.b, Theorem 3.2) enough to show that

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

where M is independent of N .

Since for normally distributed random variables with mean 0,

$$EX^4 = 3[EX^2]^2,$$

we have

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

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

$$f(F^{-1}(t)) \geq M^{-\frac{1}{4}} \text{ for all } t \in [0, 1].$$

Then

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

Hence $V_N(t)$ converges in distribution to the process $V(t)$. Thus, according to Billingsley (1968, 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

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

Q.E.D.

Now, let h_j be a function defined by

$$(3.6) \quad h_j(u) = \frac{1}{\mu} [q_j(u) - q_j(1)L(u)]$$

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

THEOREM 3: Suppose the conditions of Theorem 1 are satisfied. Then $Z_n(u)$ given by (3.1) converges in distribution to the Gaussian process

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

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

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

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

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

Now, Theorem 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 μ , Cramer-Slutsky's theorem gives that $Z_n(u)$ converges in distribution to the process

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

Thus, by applying Theorem 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 1: Suppose the conditions of Theorem 1 are satisfied. Then

$$(3.8) \quad \sum_{i=1}^{\infty} q_i(u)q_i(v) = \tau^2(u) + \lambda(u, v), \quad 0 \leq u \leq v \leq 1,$$

where $q_i(u)$ is defined by (3.4) and $\tau^2(u)$ and $\lambda(u, v)$ are given by

$$(3.9) \quad \tau^2(u) = 2 \int_a^{F^{-1}(u)} \int_a^y F(x)(1-F(y)) dx dy, \quad 0 \leq u \leq 1$$

and

$$(3.10) \quad \lambda(u, v) = \int_{F^{-1}(u)}^{F^{-1}(v)} \int_a^{F^{-1}(u)} F(x)(1-F(y)) dx dy, \quad 0 \leq u \leq v \leq 1.$$

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

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

By applying Fubini's theorem (e.g. Royden (1963)) and the identity (3.5) we get

$$\begin{aligned} \sum_{i=1}^{\infty} q_i(u)q_i(v) &= \int_0^v \int_0^u \left[\frac{2}{f(F^{-1}(t))f(F^{-1}(s))} \sum_{i=1}^{\infty} \frac{\sin(i\pi t)\sin(i\pi s)}{(i\pi)^2} \right] dt ds \\ &= 2 \int_0^u \int_0^s \frac{t(1-s)}{f(F^{-1}(t))f(F^{-1}(s))} dt ds + \int_u^v \int_0^u \frac{t(1-s)}{f(F^{-1}(t))f(F^{-1}(s))} dt ds \\ &= 2 \int_a^{F^{-1}(u)} \int_a^y (F(x)(1-F(y)) dx dy) + \int_{F^{-1}(u)}^{F^{-1}(v)} \int_a^{F^{-1}(u)} F(x)(1-F(y)) dx dy = \tau^2(u) + \lambda(u, v). \end{aligned}$$

Q.E.D.

As an immediate consequence of Theorem 1, Theorem 2 and Lemma 1 we have the following corollary.

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

$$(3.11) \quad \theta^2(u, v) = \frac{1}{uv} [\tau^2(u) + \lambda(u, v)], \quad 0 < u \leq v \leq 1.$$

From Theorem 3 and Lemma 1 we get the next corollary.

COROLLARY 2: Under the conditions of Theorem 1, $Z_n(u)$ has asymptotic covariance function $v^2(u, v)$ given by

$$(3.12) \quad v^2(u, v) = \frac{1}{\mu^2} \left[\tau^2(u) + \lambda(u, v) - L(u)(\tau^2(v) + \lambda(v, 1)) - L(v)(\tau^2(u) + \lambda(u, 1)) + L(u)L(v)\tau^2(1) \right], \quad 0 < u \leq v \leq 1.$$

In order to construct confidence intervals for the Lorenz curve at fixed points, we apply the results of Theorem 3 and Corollary 2 which imply that the distribution of

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

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

$$(3.13) \quad v^2(u, u) = \frac{1}{\mu^2} \left[\tau^2(u) - 2L(u)(\tau^2(u) + \lambda(u, 1)) + L^2(u)\tau^2(1) \right], \quad 0 < u \leq 1.$$

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

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

and

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

we obtain the following consistent estimates of τ^2 and λ ,

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

and

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

Thus, replacing μ , L , τ^2 and λ by their respective estimates in the expression (3.13) for ν^2 we obtain a consistent estimate of ν^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

$$(3.18) \quad K_n = n^{\frac{1}{2}} \sup_{0 \leq u \leq 1} \frac{|L_n(u) - L(u)|}{\psi(L_n(u))}$$

where ψ is a continuous nonnegative weight function. By applying Theorem 3 and Billingsley (1968, Theorem 5.1), we find that K_n converges in distribution to

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

Let

$$(3.20) \quad T_m(u) = \sum_{j=1}^m \frac{h_j(u)}{\psi(L(u))} Z_j,$$

$$(3.21) \quad T(u) = \sum_{j=1}^{\infty} \frac{h_j(u)}{\psi(L(u))} Z_j$$

and

$$(3.22) \quad K'_m = \sup_{0 \leq u \leq 1} |T_m(u)|.$$

Since T_m converges in distribution to T , we find by applying Billingsley (1968, Theorem 5.1) that K'_m converges in distribution to K . Hence, for a suitable choice of m and ψ , for instance $\psi = 1$, 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 (1976) states without proof that $n^{\frac{1}{2}}(\hat{J}_R - J_R)$ is asymptotically normally distributed with mean zero. The asymptotic variance, however, cannot be derived, as maintained by Mehran (1976), from Stigler (1974, Theorem 1). However, as will be demonstrated below Theorem 3 forms a helpful basis for deriving the asymptotic variance of \hat{J}_R .

Let ω^2 be a parameter defined by

$$(3.23) \quad \omega^2 = \frac{1}{\mu^2} \left\{ 2 \int_0^1 \int_0^v [\tau^2(u) + \lambda(u, v)] R(u) R(v) du dv \right. \\ \left. - 2 \left[\int_0^1 u R(u) du - J_R \right] \left[\int_0^1 (\tau^2(u) + \lambda(u, 1)) R(u) du \right] + \tau^2(1) \left[\int_0^1 u R(u) du - J_R \right]^2 \right\}.$$

THEOREM 4: Suppose the conditions of Theorem 1 are satisfied and $\omega^2 < \infty$. Then the distribution of

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

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

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

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

By Theorem 3 we have that $Z_n(u)$ converges in distribution to the Gaussian process $Z(u)$ defined by (3.7). By applying Billingsley (1968, Theorem 5.1) and Fubini's theorem we get that $n^{\frac{1}{2}}(\hat{J}_R - J_R)$ 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, \dots are independent $N(0,1)$ variables and $h_j(u)$ is given by (3.6), i.e., the asymptotic distribution of $n^{\frac{1}{2}}(\hat{J}_R - J_R)$ is normal with mean zero and variance

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

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

Inserting (3.6) in (3.24), we get

$$\begin{aligned} \sum_{j=1}^{\infty} \left[\int_0^1 R(u)h_j(u)du \right]^2 &= \frac{1}{\mu^2} \sum_{j=1}^{\infty} \left[\int_0^1 R(u)(q_j(u) - q_j(1)L(u))du \right]^2 \\ &= \frac{1}{\mu^2} \left\{ \sum_{j=1}^{\infty} \left[\int_0^1 R(u)q_j(u)du \right]^2 - 2 \left[\int_0^1 R(u)L(u)du \right] \left[\sum_{j=1}^{\infty} q_j(1) \int_0^1 R(u)q_j(u)du \right] \right. \\ &\quad \left. + \left[\sum_{j=1}^{\infty} q_j^2(1) \right] \left[\int_0^1 R(u)L(u)du \right]^2 \right\}. \end{aligned}$$

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

$$\begin{aligned} \sum_{j=1}^{\infty} \left[\int_0^1 R(u)q_j(u)du \right]^2 &= \sum_{j=1}^{\infty} \int_0^1 \int_0^1 R(u)q_j(u)R(v)q_j(v)du dv \\ &= \int_0^1 \int_0^1 \left[\int_0^u \int_0^v \frac{2}{f(F^{-1}(t))f(F^{-1}(s))} \left(\sum_{j=1}^{\infty} \frac{\sin(j\pi t)\sin(j\pi s)}{(j\pi)^2} \right) dt ds \right] R(u)R(v)du dv \\ &= 2 \int_0^1 \int_0^v \left[2 \int_0^u \int_0^s \frac{t(1-s)}{f(F^{-1}(t))f(F^{-1}(s))} dt ds + \int_u^v \int_0^u \frac{t(1-s)}{f(F^{-1}(t))f(F^{-1}(s))} dt ds \right] R(u)R(v)du dv \\ &= 2 \int_0^1 \int_0^v \left[2 \int_a^{F^{-1}(u)} \int_a^y F(x)(1-F(y))dx dy + \int_{F^{-1}(u)}^{F^{-1}(v)} \int_a^{F^{-1}(u)} F(x)(1-F(y))dx dy \right] R(u)R(v)du dv \\ &= 2 \int_0^1 \int_0^v \left[\tau^2(u) + \lambda(u, v) \right] R(u)R(v)du dv \end{aligned}$$

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

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

From Lemma 1 it follows that

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

Finally, by noting that

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

the proof is completed.

Q.E.D.

For $R(u) = 2$, Theorem 4 states that $\omega^2 = \gamma^2$, where γ^2 is defined by

$$(3.25) \quad \gamma^2 = \frac{4}{\mu^2} \left\{ 2 \int_0^1 \int_0^v [\tau^2(u) + \lambda(u,v)] du dv - (1-G) \int_0^1 [\tau^2(u) + \lambda(u,1)] du + \frac{1}{4}(1-G)^2 \tau^2(1) \right\},$$

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

An alternative version of (3.25) is given by Hoeffding (1948).

The estimation of γ^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 γ^2 , we obtain a consistent nonparametric estimator for γ^2 . The current estimator is given by

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

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

Similarly, a consistent estimator for ω^2 is obtained by replacing τ^2 , λ , μ and I by their respective estimates in the expression (3.23) for ω^2 .

References

- Aaberge, R. (1982): On the Problem of Measuring Inequality. (In Norwegian). Rapp 82/9, Statistics Norway.
- Aaberge, R. (2000a): Characterizations of Lorenz Curves and Income Distributions. *Social Choice and Welfare* **17**, 639-653.
- Aaberge, R. (2000b): "Ranking Intersecting Lorenz Curves", Discussion Paper 271, Statistics Norway.
- Aaberge, R. (2001): Axiomatic Characterization of the Gini Coefficient and Lorenz Curve Orderings. *Journal of Economic Theory* **101**, 115-132.
- Atkinson, A.B. (1970): On the measurement of inequality, *Journal of Economic Theory*, **2**, 244-263.
- 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.
- Billingsley, P. (1968): *Convergence of Probability Measures*. John Wiley & Sons, Inc., New York.
- 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.
- 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.
- Donaldson, D. and J.A. Weymark (1980): A Single Parameter Generalization of the Gini Indices of Inequality. *Journal of Economic Theory* **22**, 67-86.
- Donaldson, D. and J.A. Weymark (1983): Ethically Flexible Indices for Income Distribution in the Continuum. *Journal of Economic Theory* **29**, 353-358.
- Durbin, J. (1973): *Distribution Theory for Tests Based on the Sample Distribution Function*. Society for Industrial and Applied Mathematics, Philadelphia.
- Giaccardi, F. (1950): Un criterio per la costruzione di indici di concentrazione. *Rivista Italiana di Demografia e Statistica* **4**, 527-538.
- Goldie, C.M. (1977). Convergence theorems for empirical Lorenz curves and their inverses. *Advances in Applied Probability* **9**, 765-791.
- Hajek, J. and Z. Sidak (1967): *Theory of Rank Tests*. Academic Press, New York.
- Hoeffding, W. (1948): A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics* **19**, 293-325.
- Kolm, S.Ch. (1969): "The optimal production of social justice". In J. Margolis and H. Guitton (eds.): *Public Economics*. Macmillan, New York/London.
- Kolm, S.Ch. (1976a): Unequal inequalities I, *Journal of Economic Theory* **12**, 416-442.
- Kolm, S.Ch. (1976b): Unequal inequalities II, *Journal of Economic Theory* **13**, 82-111.

- Lambert, P.J. (1993): *The Distribution and Redistribution of Income: A Mathematical Analysis*. Manchester University Press, Manchester.
- Mehran, F. (1976): Linear Measures of Shorrocks, A.F. and J.E. Foster (1978): Transfer sensitive inequality measures. *Review of Economic Studies* **14**, 485-497.
- Piesch, W. (1975): *Statistische Konzentrationsmasse*. Mohr (Siebeck), Tübingen.
- Royden, H. (1963): *Real Analysis*. Macmillan, New York.
- Shorack, G.R. (1972): Functions of Order Statistics. *The Annals of Mathematical Statistics* **43** (2), 412-427.
- Shorrocks, A.F. and J.E. Foster (1978): Transfer sensitive inequality measures. *Review of Economic Studies* **14**, 485-497.
- Stigler, S.M. (1974): Linear Functions of Order Statistics with Smooth Weight Functions. *The Annals of Statistics* **2** (4), 676-693.
- Weymark, J. (1981): Generalized Gini Inequality Indices. *Mathematical Social Science* **1**, 409-430.
- Yitzhaki, S. (1983): On an extension of the Gini inequality index. *International Economic Review* **24**, 617-628.
- Yaari, M.E. (1987): The Dual Theory of Choice under Risk. *Econometrica* **55**, 95-115.
- Yaari, M.E. (1988): A Controversial Proposal Concerning Inequality Measurement. *Journal of Economic Theory* **44**, 381-397.

Recent publications in the series Discussion Papers

- 312 K. R. Wangen and E. Biørn (2001): Prevalence and substitution effects in tobacco consumption. A discrete choice analysis of panel data
- 313 G.H. Bjertnær (2001): Optimal Combinations of Income Tax and Subsidies for Education
- 314 K. E. Rosendahl (2002): Cost-effective environmental policy: Implications of induced technological change
- 315 T. Kornstad and T.O. Thoresen (2002): A Discrete Choice Model for Labor Supply and Child Care
- 316 A. Bruvoll and K. Nyborg (2002): On the value of households' recycling efforts
- 317 E. Biørn and T. Skjerpen (2002): Aggregation and Aggregation Biases in Production Functions: A Panel Data Analysis of Translog Models
- 318 Ø. Døhl (2002): Energy Flexibility and Technological Progress with Multioutput Production. Application on Norwegian Pulp and Paper Industries
- 319 R. Aaberge (2002): Characterization and Measurement of Duration Dependence in Hazard Rate Models
- 320 T. J. Klette and A. Raknerud (2002): How and why do Firms differ?
- 321 J. Aasness and E. Røed Larsen (2002): Distributional and Environmental Effects of Taxes on Transportation
- 322 E. Røed Larsen (2002): The Political Economy of Global Warming: From Data to Decisions
- 323 E. Røed Larsen (2002): Searching for Basic Consumption Patterns: Is the Engel Elasticity of Housing Unity?
- 324 E. Røed Larsen (2002): Estimating Latent Total Consumption in a Household.
- 325 E. Røed Larsen (2002): Consumption Inequality in Norway in the 80s and 90s.
- 326 H.C. Bjørnland and H. Hungnes (2002): Fundamental determinants of the long run real exchange rate: The case of Norway.
- 327 M. Søberg (2002): A laboratory stress-test of bid, double and offer auctions.
- 328 M. Søberg (2002): Voting rules and endogenous trading institutions: An experimental study.
- 329 M. Søberg (2002): The Duhem-Quine thesis and experimental economics: A reinterpretation.
- 330 A. Raknerud (2002): Identification, Estimation and Testing in Panel Data Models with Attrition: The Role of the Missing at Random Assumption
- 331 M.W. Arneberg, J.K. Dagsvik and Z. Jia (2002): Labor Market Modeling Recognizing Latent Job Attributes and Opportunity Constraints. An Empirical Analysis of Labor Market Behavior of Eritrean Women
- 332 M. Greaker (2002): Eco-labels, Production Related Externalities and Trade
- 333 J. T. Lind (2002): Small continuous surveys and the Kalman filter
- 334 B. Halvorsen and T. Willumsen (2002): Willingness to Pay for Dental Fear Treatment. Is Supplying Fear Treatment Social Beneficial?
- 335 T. O. Thoresen (2002): Reduced Tax Progressivity in Norway in the Nineties. The Effect from Tax Changes
- 336 M. Søberg (2002): Price formation in monopolistic markets with endogenous diffusion of trading information: An experimental approach
- 337 A. Bruvoll og B.M. Larsen (2002): Greenhouse gas emissions in Norway. Do carbon taxes work?
- 338 B. Halvorsen and R. Nesbakken (2002): A conflict of interests in electricity taxation? A micro econometric analysis of household behaviour
- 339 R. Aaberge and A. Langørgen (2003): Measuring the Benefits from Public Services: The Effects of Local Government Spending on the Distribution of Income in Norway
- 340 H. C. Bjørnland and H. Hungnes (2003): The importance of interest rates for forecasting the exchange rate
- 341 A. Bruvoll, T.Fæhn and Birger Strøm (2003): Quantifying Central Hypotheses on Environmental Kuznets Curves for a Rich Economy: A Computable General Equilibrium Study
- 342 E. Biørn, T. Skjerpen and K.R. Wangen (2003): Parametric Aggregation of Random Coefficient Cobb-Douglas Production Functions: Evidence from Manufacturing Industries
- 343 B. Bye, B. Strøm and T. Åvitsland (2003): Welfare effects of VAT reforms: A general equilibrium analysis
- 344 J.K. Dagsvik and S. Strøm (2003): Analyzing Labor Supply Behavior with Latent Job Opportunity Sets and Institutional Choice Constraints
- 345 A. Raknerud, T. Skjerpen and A. Rygh Swensen (2003): A linear demand system within a Seemingly Unrelated Time Series Equation framework
- 346 B.M. Larsen and R.Nesbakken (2003): How to quantify household electricity end-use consumption
- 347 B. Halvorsen, B. M. Larsen and R. Nesbakken (2003): Possibility for hedging from price increases in residential energy demand
- 348 S. Johansen and A. R. Swensen (2003): More on Testing Exact Rational Expectations in Cointegrated Vector Autoregressive Models: Restricted Drift Terms
- 349 B. Holtmark (2003): The Kyoto Protocol without USA and Australia - with the Russian Federation as a strategic permit seller
- 350 J. Larsson (2003): Testing the Multiproduct Hypothesis on Norwegian Aluminium Industry Plants
- 351 T. Bye (2003): On the Price and Volume Effects from Green Certificates in the Energy Market
- 352 E. Holmøy (2003): Aggregate Industry Behaviour in a Monopolistic Competition Model with Heterogeneous Firms
- 353 A. O. Ervik, E.Holmøy and T. Hægeland (2003): A Theory-Based Measure of the Output of the Education Sector
- 354 E. Halvorsen (2003): A Cohort Analysis of Household Saving in Norway
- 355 I. Aslaksen and T. Synnestvedt (2003): Corporate environmental protection under uncertainty
- 356 S. Glomsrød and W. Taoyuan (2003): Coal cleaning: A viable strategy for reduced carbon emissions and improved environment in China?
- 357 A. Bruvoll T. Bye, J. Larsson og K. Telle (2003): Technological changes in the pulp and paper industry

- and the role of uniform versus selective environmental policy.
- 358 J.K. Dagsvik, S. Strøm and Z. Jia (2003): A Stochastic Model for the Utility of Income.
- 359 M. Rege and K. Telle (2003): Indirect Social Sanctions from Monetarily Unaffected Strangers in a Public Good Game.
- 360 R. Aaberge (2003): Mean-Spread-Preserving Transformation.
- 361 E. Halvorsen (2003): Financial Deregulation and Household Saving. The Norwegian Experience Revisited
- 362 E. Røed Larsen (2003): Are Rich Countries Immune to the Resource Curse? Evidence from Norway's Management of Its Oil Riches
- 363 E. Røed Larsen and Dag Einar Sommervoll (2003): Rising Inequality of Housing? Evidence from Segmented Housing Price Indices
- 364 R. Bjørnstad and T. Skjerpen (2003): Technology, Trade and Inequality
- 365 A. Raknerud, D. Rønningen and T. Skjerpen (2003): A method for improved capital measurement by combining accounts and firm investment data
- 366 B.J. Holtmark and K.H. Alfsen (2004): PPP-correction of the IPCC emission scenarios - does it matter?
- 367 R. Aaberge, U. Colombino, E. Holmøy, B. Strøm and T. Wennemo (2004): Population ageing and fiscal sustainability: An integrated micro-macro analysis of required tax changes
- 368 E. Røed Larsen (2004): Does the CPI Mirror Costs of Living? Engel's Law Suggests Not in Norway
- 369 T. Skjerpen (2004): The dynamic factor model revisited: the identification problem remains
- 370 J.K. Dagsvik and A.L. Mathiassen (2004): Agricultural Production with Uncertain Water Supply
- 371 M. Greaker (2004): Industrial Competitiveness and Diffusion of New Pollution Abatement Technology – a new look at the Porter-hypothesis
- 372 G. Børnes Ringlund, K.E. Rosendahl and T. Skjerpen (2004): Does oilrig activity react to oil price changes? An empirical investigation
- 373 G. Liu (2004) Estimating Energy Demand Elasticities for OECD Countries. A Dynamic Panel Data Approach
- 374 K. Telle and J. Larsson (2004): Do environmental regulations hamper productivity growth? How accounting for improvements of firms' environmental performance can change the conclusion
- 375 K.R. Wangen (2004): Some Fundamental Problems in Becker, Grossman and Murphy's Implementation of Rational Addiction Theory
- 376 B.J. Holtmark and K.H. Alfsen (2004): Implementation of the Kyoto Protocol without Russian participation
- 377 E. Røed Larsen (2004): Escaping the Resource Curse and the Dutch Disease? When and Why Norway Caught up with and Forged ahead of Its Neighbors
- 378 L. Andreassen (2004): Mortality, fertility and old age care in a two-sex growth model
- 379 E. Lund Sagen and F. R. Aune (2004): The Future European Natural Gas Market - are lower gas prices attainable?
- 380 A. Langørgen and D. Rønningen (2004): Local government preferences, individual needs, and the allocation of social assistance
- 381 K. Telle (2004): Effects of inspections on plants' regulatory and environmental performance - evidence from Norwegian manufacturing industries
- 382 T. A. Galloway (2004): To What Extent Is a Transition into Employment Associated with an Exit from Poverty
- 383 J. F. Bjørnstad and E. Ytterstad (2004): Two-Stage Sampling from a Prediction Point of View
- 384 A. Bruvoll and T. Fæhn (2004): Transboundary environmental policy effects: Markets and emission leakages
- 385 P.V. Hansen and L. Lindholt (2004): The market power of OPEC 1973-2001
- 386 N. Keilman and D. Q. Pham (2004): Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area
- 387 G. H. Bjertnæs and T. Fæhn (2004): Energy Taxation in a Small, Open Economy: Efficiency Gains under Political Restraints
- 388 J.K. Dagsvik and S. Strøm (2004): Sectoral Labor Supply, Choice Restrictions and Functional Form
- 389 B. Halvorsen (2004): Effects of norms, warm-glow and time use on household recycling
- 390 I. Aslaksen and T. Synnøve (2004): Are the Dixit-Pindyck and the Arrow-Fisher-Henry-Hanemann Option Values Equivalent?
- 391 G. H. Bjørnes, D. Rime and H. O.Aa. Solheim (2004): Liquidity provision in the overnight foreign exchange market
- 392 T. Åvitsland and J. Aasness (2004): Combining CGE and microsimulation models: Effects on equality of VAT reforms
- 393 M. Greaker and Eirik. Sagen (2004): Explaining experience curves for LNG liquefaction costs: Competition matter more than learning
- 394 K. Telle, I. Aslaksen and T. Synnøve (2004): "It pays to be green" - a premature conclusion?
- 395 T. Harding, H. O. Aa. Solheim and A. Benedictow (2004). House ownership and taxes
- 396 E. Holmøy and B. Strøm (2004): The Social Cost of Government Spending in an Economy with Large Tax Distortions: A CGE Decomposition for Norway
- 397 T. Hægeland, O. Raaum and K.G. Salvanes (2004): Pupil achievement, school resources and family background
- 398 I. Aslaksen, B. Natvig and I. Nordal (2004): Environmental risk and the precautionary principle: "Late lessons from early warnings" applied to genetically modified plants
- 399 J. Møen (2004): When subsidized R&D-firms fail, do they still stimulate growth? Tracing knowledge by following employees across firms
- 400 B. Halvorsen and Runa Nesbakken (2004): Accounting for differences in choice opportunities in analyses of energy expenditure data
- 401 T.J. Klette and A. Raknerud (2004): Heterogeneity, productivity and selection: An empirical study of Norwegian manufacturing firms
- 402 R. Aaberge (2005): Asymptotic Distribution Theory of Empirical Rank-dependent Measures of Inequality