Econometric Theory, **19**, 2003, 587–601. Printed in the United States of America.
DOI: 10.1017/S0266466603194042

ASYMPTOTIC ESTIMATION OF THE *E*-GINI INDEX

RIČARDAS ZITIKIS
University of Western Ontario

Under minimal assumptions on the distribution of income, we demonstrate that Chakravarty's empirical (1988, *International Economic Review* 29, 147–156) *E*-Gini index is consistent and asymptotically normal. We also derive an explicit formula for the asymptotic variance of the index and then construct a consistent and computationally straightforward estimator for it.

1. INTRODUCTION

Measuring economic inequality, and, indeed, finding ways to properly measure it, is a challenging and important problem. A large number of indices of economic inequality, compatible with various axioms of fairness, have been proposed in the literature. For their definitions, related discussions, and further references, we refer, for example, to the books by Kakwani (1980a), Nygård and Sandström (1981), Sen (1997), Champernowne and Cowell (1998), Cowell (1998), Amiel and Cowell (1999), Silber (1999), and Atkinson and Bourguignon (2000).

In one way or another, many indices are based on the Lorenz curve, the notion of which was introduced by Lorenz (1905). Later, the notion was mathematically formalized by several authors, including Kendall and Stuart (1958), assuming that the (cumulative) distribution function F of "income" $X \geq 0$ has a probability density function. Finally, assuming only that the mean $\mu := \mathbf{E}(X)$ is finite and positive, Gastwirth (1971) gave the most general definition of the Lorenz curve as follows:

$$L_F(t) := \frac{1}{\mu} \int_0^t F^{-1}(s)\, ds, \qquad 0 \leq t \leq 1, \tag{1.1}$$

Sincere thanks are due to the co-editor Oliver B. Linton and two anonymous referees whose constructive criticism, advice, and queries helped me in reshaping the paper considerably. As advised by a referee, I had the great pleasure of communicating with Garry F. Barrett and Stephen G. Donald and learning about their interesting and closely related results. The starting point of my work on the project was correspondence with Joseph L. Gastwirth in the spring of 2000 that resulted in our joint work on the *S*-Gini index. I am grateful to Joseph L. Gastwirth for his time, his advice, and his numerous suggestions that followed. The help, in addition to interest in the project, by Ying Zhang of the Statistical Laboratory at the University of Western Ontario is greatly appreciated; the analysis of the dependence of $\sigma^2_{F,\alpha}$ on parameters presented in Table 1 is due to her. This research was partially supported by an NSERC of Canada individual research grant at the University of Western Ontario. Address correspondence to: Ričardas Zitikis, University of Western Ontario, Department of Statistical and Actuarial Sciences, London, Ontario, Canada, N6A 5B7; e-mail zitikis@stats.uwo.ca.

© 2003 Cambridge University Press 0266-4666/03 $12.00

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

where

$$F^{-1}(s) := \inf\{u : F(u) \geq s\}, \qquad 0 < s < 1, \tag{1.2}$$

is the quantile function of F. Note that formula (1.2) defines the quantile function F^{-1} equally well for all distribution functions F. Thus, Gastwirth's definition (1.1) of L_F is also valid for all distribution functions F, implying the outmost generality of it. For more information concerning quantiles and further references on the subject, we refer, for example, to Parzen (1979), Csörgő (1983, 1986), and Csörgő and Horváth (1993).

The value of the Lorenz curve L_F does not depend on the measurement units of X. Furthermore, L_F is always below the diagonal line $I(t) := t$. The diagonal I, on the other hand, has a special interpretation. Namely, if the income X is constant, meaning the absolute equality of incomes, then the corresponding Lorenz curve is identically I. This suggests that the income inequality in a society can be measured by the area between the diagonal I and the Lorenz curve L_F corresponding to the society. Thus, the classical Gini index:

$$G_F = 2\int_0^1 (t - L_F(t))\, dt, \tag{1.3}$$

twice the area between I and L_F. The Gini index G_F has played a central role in measuring economic inequality since the works by Corrado Gini at the beginning of the 20th century. For historical and bibliographical notes on the subject we refer, for example, to David (1968) and Giorgi (1990, 1993).

The Gini index G_F satisfies a number of axioms or, in other words, principles such as those of transfers, scale independence, and others; we refer to Dagum (1983) for an account of some of them. In particular, the Gini index satisfies (cf., e.g., Dagum, 1983) the Pigou–Dalton principle of transfers, which postulates that progressive (i.e., from rich to poor), rank order of incomes and also mean preserving transfers should decrease the values of inequality measures. Having verified the Pigou–Dalton principle, one "may go one step further and value more such a transfer between persons with given income difference if these incomes are lower than if they are higher" (Kolm, 1976b, p. 87). This describes the main idea of the principle of transfer sensitivity that has been discussed in great detail by Atkinson (1970), Mehran (1976), Kolm (1976a, 1976b), Kakwani (1980b), Shorrocks and Foster (1987), Chakravarty (1988), Sen (1997), Zheng (1997), Cowell (1998), Amiel and Cowell (1999), Silber (1999), and others.

Chakravarty (1988) notes that the Gini index G_F fails to satisfy the so-called diminishing transfers axiom, whose meaning can be summarized as follows:

> For two persons having incomes y_i and y_j with $y_i < y_j$, the reduction of inequality attributable to a rank preserving transfer from the person with income y_i to another person with income $y_i - c$ $(0 < c < y_i)$ is greater than the reduction attrib-

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

> utable to an equal rank preserving transfer from the person with income y_j to someone with income $y_j - c$. (Chakravarty, 1988, p. 147)

To rectify the situation, Chakravarty (1988) proposes a route that, in mathematical terms, can be described as follows. Let $\mathcal{F}_\mu$ denote the collection of the quantile functions F^{-1} corresponding to those distribution functions F whose mean is μ. Define the functional $\mathcal{G}$ on $\mathcal{F}_\mu$ by the formula

$$\mathcal{G}(h) := 2\int_0^1 \left(t - \frac{1}{\mu}\int_0^t h(s)\,ds\right) dt. \tag{1.4}$$

Note that the Gini index G_F equals $\mathcal{G}(F^{-1})$. The functional $\mathcal{G}$ is linear, which appears (cf. Chakravarty, 1988, p. 148) to be the reason for the absence of the desired transfer sensitivity of the Gini index G_F. Therefore, Chakravarty (1988) suggests modifying the functional $\mathcal{G}$ in such a way that the newly obtained functional would be strictly convex, that is, the quantity $\mathcal{G}(\gamma h_1 + (1-\gamma)h_2)$ would be strictly less than the sum of $\gamma\mathcal{G}(h_1)$ and $(1-\gamma)\mathcal{G}(h_2)$, for any $\gamma \in (0,1)$ and $h_1, h_2 \in \mathcal{F}_\mu$. As an example of such a functional, Chakravarty (1988) suggests the following expression:

$$\mathcal{G}_\alpha(h) := 2\left(\int_0^1 \left(t - \frac{1}{\mu}\int_0^t h(s)\,ds\right)^\alpha dt\right)^{1/\alpha}. \tag{1.5}$$

The functional $\mathcal{G}_\alpha$ is strictly convex on $\mathcal{F}_\mu$ for any $\alpha > 1$. Thus, we have Chakravarty's (1988) E(xtended)-Gini index $G_{F,\alpha} := \mathcal{G}_\alpha(F^{-1})$, the explicit formula of which is

$$G_{F,\alpha} := 2\left(\int_0^1 (t - L_F(t))^\alpha\,dt\right)^{1/\alpha}. \tag{1.6}$$

We note in passing that in Zitikis (2002a) we argue that using $(\alpha+1)^{1/\alpha}$ instead of 2 in definition (1.6) would be more natural from the mathematical point of view. Whether this modification is desirable from the econometric point of view still remains to be determined. We note here only that the asymptotic results presented in the following sections can easily be adapted to cover the proposed modification of $G_{F,\alpha}$.

We conclude this section with the note that Chakravarty (1988) suggests a functional even more general than $\mathcal{G}_\alpha$, which is

$$\mathcal{G}_\phi(h) := 2\phi^{-1}\left(\int_0^1 \phi\left(t - \frac{1}{\mu}\int_0^t h(s)\,ds\right) dt\right), \tag{1.7}$$

where $\phi : [0,1] \to [0,\infty)$ is a continuous, strictly increasing, and strictly convex function such that $\phi(0) = 0$. As far as we are aware, the importance of functions ϕ other than $\phi(x) = x^\alpha$ has not yet been clarified in the econometric literature. In this paper we therefore restrict ourselves exclusively to the functional $\mathcal{G}_\alpha$, that is, to the E-Gini index $G_{F,\alpha}$ defined in (1.6).

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

2. *E*-GINI ESTIMATOR

The distribution function F is generally unknown, and thus estimation theory for the E-Gini index $G_{F,\alpha}$ is desirable. We shall now construct an empirical estimator for $G_{F,\alpha}$. Let $X_1,\ldots,X_n$ be independent copies of X and let F_n be the empirical distribution function based on $X_1,\ldots,X_n$. Furthermore, let F_n^{-1} be the corresponding quantile function, which is defined by formula (1.2) with F_n instead of F. Replacing μ in definition (1.1) by the sample mean $\overline{X}$ of $X_1,\ldots,X_n$, and also replacing F^{-1} by its empirical counterpart F_n^{-1}, we obtain the empirical Lorenz curve (Gastwirth, 1971):

$$L_n(t) := \frac{1}{\overline{X}} \int_0^t F_n^{-1}(s)\, ds, \qquad 0 \le t \le 1. \tag{2.1}$$

Chakravarty's (1988) E-Gini extimator is now defined as follows:

$$G_{n,\alpha} := 2\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{i}{n} - L_n\left(\frac{i}{n}\right)\right)^{\alpha}\right)^{1/\alpha}. \tag{2.2}$$

The most recent and complete asymptotic theory for the E-Gini estimator $G_{n,\alpha}$ has been developed by Barrett and Donald (2000). In particular, Barrett and Donald (2000) prove that $G_{n,\alpha}$ is consistent and asymptotically normal under the following three assumptions.

(i) The support of F is in a bounded interval $[y_l, y_u]$ for some $0 < y_l < y_u < \infty$.
(ii) The distribution function F is twice differentiable.
(iii) The density function $f = F'$ is such that $0 < \inf f(y) < \sup f(y) < \infty$, where both inf and sup are taken over all $y \in [y_l, y_u]$.

Naturally, it is of interest, and also of importance, to relax Assumptions (i)–(iii). There are several reasons for this. For example, minimal assumptions would assure the widest applicability of the obtained results and only minimal efforts would be required to verify such assumptions in practical situations. For a discussion of these issues, and further references on the subject, we refer to Beach and Davidson (1983).

The paper by Beach and Davidson (1983) was instrumental in developing the *nonparametric* statistical inference for Lorenz curves and associated indices of income inequality. For example, employing the results of Beach and Davidson (1983), Barrett and Pendakur (1995) investigate, among other things, the asymptotic behavior of the empirical E-Gini estimator. Both aforementioned papers, however, depend on a certain parameter K, which is the number of groups defined by percentiles into which the data under investigation are subdivided. Recently, employing the theory of empirical and quantile processes, Barrett and Donald (2000) have demonstrated that the introduction of the parameter K is unnecessary in the case of complete data. We also refer to the monograph by Csörgő, Csörgő, and Horváth (1986), and references therein,

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

on related topics in the presence of complete data. When only grouped data are available, then the results by Gastwirth (1972), Gastwirth and Glauberman (1976), Aghevli and Mehran (1981), Beach and Davidson (1983), and Barrett and Pendakur (1995) become crucial in the development of the large sample based theory of statistical inference.

3. MAIN RESULTS

As we noted in the previous section, $G_{n,\alpha}$ is a natural estimator of $G_{F,\alpha}$. Obviously, the estimator is consistent under some assumptions on F. To show, however, that this is so under minimal assumptions requires some work, the result of which is the following theorem.

THEOREM 3.1. *Let* $\alpha \geq 1$. *Then* $G_{n,\alpha}$ *is a strongly consistent estimator of* $G_{F,\alpha}$.

Having thus proved the consistency of $G_{n,\alpha}$, we would now like to use the estimator for constructing asymptotic confidence intervals for $G_{F,\alpha}$. For this reason, we need to obtain the sampling distribution of $G_{n,\alpha}$. This we achieve in Theorem 3.2, which follows.

THEOREM 3.2. *Let* $\mathbf{E}(X^2) < \infty$ *and* $\alpha \geq 1$. *Then the asymptotic distribution of* $\sqrt{n}\{G_{n,\alpha} - G_{F,\alpha}\}$ *is normal with the mean* 0 *and the variance*

$$\sigma_{F,\alpha}^2 := \frac{4^\alpha}{G_{F,\alpha}^{2(\alpha-1)}\,\mu^2}\,\Delta_{F,\alpha}^2, \tag{3.1}$$

where the quantity $\Delta_{F,\alpha}^2$ *is defined by the formula*

$$\begin{aligned}\Delta_{F,\alpha}^2 := &\int_0^1\int_0^1 (\min(u,v) - uv)\psi(u)\psi(v)\,dF^{-1}(u)\,dF^{-1}(v)\\ &- 2\mu\left(\int_0^1 L_F(t)(t - L_F(t))^{\alpha-1}\,dt\right)\left(\int_0^1 (u - L_F(u))\psi(u)\,dF^{-1}(u)\right)\\ &+ \sigma_X^2\left(\int_0^1 L_F(t)(t - L_F(t))^{\alpha-1}\,dt\right)^2\end{aligned}$$

with σ_X^2 *denoting the variance of* X *and with* $\psi(u) := \int_u^1 (t - L_F(t))^{\alpha-1}\,dt$.

Several authors obtained slightly different, but equivalent, expressions for the variance $\sigma_{F,\alpha}^2$ in the classical Gini case (cf., e.g., Hoeffding, 1948, p. 314; Shorack and Wellner, 1986, p. 780). Obtaining explicit expressions for the variance $\sigma_{F,\alpha}^2$ has been a challenging problem in the general case. A review of various results along with the corresponding references can be found in Xu (2000). We shall now briefly discuss the source of the main difficulties. We start with a quote from Xu (2000): "the generalized-Gini-index estimators are functions of

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

sample quantiles that are not regular statistics" (Xu, 2000, p. 223). Certainly, estimating quantiles is a challenging problem, the main difficulty arising due to the different tail behaviors of the theoretical and empirical quantile functions. However, it should be noted that when integrating the quantile functions F^{-1} and F_n^{-1}, the differences between the resulting two integrals become less noticeable and, indeed, in the definitions of measures of income inequality the quantile functions play the roles of integrands. Thus, not separating integrals from quantile functions should be one of the main messages when working in the area. This approach in the case of the Lorenz curve L_F was crucially utilized by Csörgő and Zitikis (1996), which resulted in the introduction of the general Vervaat process, and also by Csörgő, Gastwirth, and Zitikis (1998). Related developments in the case of the *S*-Gini index are presented in Zitikis and Gastwirth (2002).

As a result of the complex mathematical expression for $\sigma^2_{F,\alpha}$, determining the dependence of $\sigma^2_{F,\alpha}$ on the parameter $\alpha \geq 1$ is difficult when F is general. For this reason we now restrict ourselves to the Pareto and exponential distributions. We start with the Pareto distribution:

$$F(x) := 1 - \left(\frac{x_0}{x}\right)^{\lambda}, \qquad x \geq x_0,$$

where $x_0 > 0$ and $\lambda > 0$ are fixed parameters. To assure the existence of the second moment, we consider the values $\lambda > 2$ only. The values of $\sigma^2_{F,\alpha}$ for various λ and α are presented in Table 1. The asymptotic variance $\sigma^2_{F,\alpha}$ does not depend on x_0.

Next we consider the exponential distribution

$$F(x) = 1 - e^{-\gamma x}, \qquad x \geq 0,$$

where $\gamma > 0$ is a fixed parameter. The corresponding values of $\sigma^2_{F,\alpha}$ are presented in the bottom row of Table 1. The asymptotic variance $\sigma^2_{F,\alpha}$ does not depend on γ.

TABLE 1. The true values of the asymptotic variance $\sigma^2_{F,\alpha}$ in the Pareto case

	α					
Pareto (λ)	1	2	3	4	5	6
3	0.137143	0.163836	0.182008	0.195384	0.205743	0.214057
5	0.025324	0.030307	0.033714	0.036228	0.038179	0.039746
10	0.004451	0.005344	0.005956	0.006408	0.006759	0.007041
20	0.000957	0.001151	0.001284	0.001383	0.001459	0.001521
Exponential	0.083333	0.105481	0.121196	0.133034	0.142337	0.149879

Note: The first four rows correspond to $\lambda = 3, 5, 10$, and 20, respectively. The exponential case is presented in the last row. In both cases, $\alpha = 1,2,\ldots,6$.

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

4. ESTIMATION OF THE ASYMPTOTIC VARIANCE

Theorem 3.2 can be used for deriving confidence intervals for the *E*-Gini index $G_{F,\alpha}$. A standard argument based on Theorem 3.2 implies the following asymptotic $100(1-p)\%$ confidence interval:

$$G_{n,\alpha} \pm z_{p/2} \frac{\sigma_{F,\alpha}}{\sqrt{n}}, \tag{4.1}$$

where $z_{p/2}$ is the $p/2$-percentile of the standard normal distribution and $p \in (0,1)$ is any fixed number. The confidence interval of (4.1) is not yet readily applicable in practice because $\sigma^2_{F,\alpha}$ depends on the (unknown) distribution function F and, therefore, needs to be estimated. This can be achieved by replacing F everywhere in the definition of $\sigma^2_{F,\alpha}$ by the empirical distribution function F_n. The obtained estimator $\sigma^2_{F_n,\alpha}$ is consistent, a statement that can be rigorously proved along the lines of the proof of Theorem 3.1. In Theorem 4.1, which follows, we use, however, a slight modification of this estimator so that practical implementation of it would be easier.

Let $\overline{X}$ and S_n^2 denote, respectively, the sample mean and the sample variance of $X_1,\dots,X_n$. Furthermore, let

$$A_{n,\alpha}(i) := \frac{1}{n}\sum_{k=i}^{n-1}\left(\frac{k}{n} - L_n\left(\frac{k}{n}\right)\right)^{\alpha-1}$$

for any $i = 1,\dots,n-1$, and let

$$B_{n,\alpha} := \frac{1}{n}\sum_{k=1}^{n-1} L_n\left(\frac{k}{n}\right)\left(\frac{k}{n} - L_n\left(\frac{k}{n}\right)\right)^{\alpha-1}.$$

With these notations we are now in the position to formulate the following theorem.

THEOREM 4.1. *Let* $\mathbf{E}(X^2) < \infty$ *and* $\alpha \geq 1$. *Then*

$$\sigma^2_{n,\alpha} := \frac{4^\alpha}{G^{2(\alpha-1)}_{n,\alpha}\overline{X}^2}\,\Delta^2_{n,\alpha}$$

is a strongly consistent estimator of $\sigma^2_{F,\alpha}$, *where*

$$\begin{aligned}\Delta^2_{n,\alpha} := \sum_{l=1}^{n-1}\sum_{m=1}^{n-1}&\left(\min\left(\frac{l}{n},\frac{m}{n}\right) - \frac{l}{n}\cdot\frac{m}{n}\right)\\ &\times A_{n,\alpha}(l)A_{n,\alpha}(m)(X_{l+1:n} - X_{l:n})(X_{m+1:n} - X_{m:n})\\ &- 2\overline{X}B_{n,\alpha}\sum_{l=1}^{n-1}\left(\frac{l}{n} - L_n\left(\frac{l}{n}\right)\right)A_{n,\alpha}(l)(X_{l+1:n} - X_{l:n}) + S_n^2 B^2_{n,\alpha}.\end{aligned}$$

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

We conclude this section with the note that when $\alpha = 1$, which is the classical Gini case, then constructing estimators for $\sigma^2_{F,\alpha}$ is a central theme in a number of papers. For more details on this topic, we refer to Yitzhaki (1991) and references therein.

5. CONCLUDING REMARKS

As we already noted earlier, one of the main aims of this paper is to describe minimal assumptions under which the empirical *E*-Gini index is consistent and asymptotically normal. In view of this, we have restricted ourselves to the simplest but important case of independent and identically distributed random variables. It is of practical and theoretical interest to extend these results to dependent random variables and also to various other sampling schemes. Some work has been done in these two directions, though a large number of contributions are outside the area of the *E*-Gini index and, more generally, of measuring income inequality. We first mention contributions related to econometrics: Shao (1994), Zheng and Cushing (2001), Zheng (2001), and references therein. Contributions pertinent to both econometrics and stochastic processes are Yu (1993), Shao and Yu (1996), Csörgő and Yu (1996, 1999), Davydov and Zitikis (2001, 2002), and references therein. For related contributions in the context of stochastic processes, we refer to Davydov and Vershik (1998), Davydov and Thilly (1999), Thilly (1999), and references therein.

REFERENCES

Aghevli, B.B. & F. Mehran (1981) Optimal grouping of income distribution data. *Journal of the American Statistical Association* 76, 22–26.

Amiel, Y. & F.A. Cowell (1999) *Thinking about Inequality.* Cambridge: Cambridge University Press.

Atkinson, A.B. (1970) On the measurement of inequality. *Journal of Economic Theory* 2, 244–263.

Atkinson, A.B. & F. Bourguignon (eds.) (2000) *Handbook of Income Distribution.* Amsterdam: Elsevier Science.

Barrett, G.F. & S.G. Donald (2000) Statistical Inference with Generalized Gini Indices of Inequality and Poverty. Discussion paper 2002/01, School of Economics, University of New South Wales.

Barrett, G.F. & K. Pendakur (1995) The asymptotic distribution of the generalized Gini indices of inequality. *Canadian Journal of Economics* 28, 1042–1055.

Beach, C.M. & R. Davidson (1983) Distribution-free statistical inference with Lorenz curves and income shares. *Review of Economic Studies* 50, 723–735.

Chakravarty, S.R. (1988) Extended Gini indices of inequality. *International Economic Review* 29, 147–156.

Champernowne, D.G. & F.A. Cowell (1998) *Economic Inequality and Income Distribution.* Cambridge: Cambridge University Press.

Cowell, F.A. (1998) *Measuring Inequality*. 2nd ed. London: Prentice-Hall/Harvester.

Csörgő, M. (1983) *Quantile Processes with Statistical Applications*. Philadelphia: SIAM.

Csörgő, M. (1986) Quantile processes. In S. Kotz, N.L. Johnson, & C.B. Read (eds.), *Encyclopedia of Statistical Sciences,* vol. 7, pp. 412–424. New York: Wiley.

Csörgő, M., S. Csörgő, & L. Horváth (1986) *An Asymptotic Theory for Empirical Reliability and Concentration Processes.* Berlin: Springer-Verlag.

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

Csörgő, M., J.L. Gastwirth, & 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.
Csörgő, M. & L. Horváth (1993) *Weighted Approximations in Probability and Statistics*. Chichester: Wiley.
Csörgő, M. & H. Yu (1996) Weak approximations for quantile processes of stationary sequences. *Canadian Journal of Statististics* 24, 403–430.
Csörgő, M. & H. Yu (1999) Weak approximations for empirical Lorenz curves and their Goldie inverses of stationary observations. *Advances in Applied Probability* 31, 698–719.
Csörgő, M. & R. Zitikis (1996) Strassen's LIL for the Lorenz curve. *Journal of Multivariate Analysis* 59, 1–12.
Dagum, C. (1983) Income inequality measures. In S. Kotz, N.L. Johnson, & C.B. Campbell (eds.), *Encyclopedia of Statistical Sciences,* vol. 4, pp. 34–40. New York: Wiley.
David, H.A. (1968) Gini's mean difference rediscovered. *Biometrika* 55, 573–575.
Davydov, Y. & E. Thilly (1999) Réarrangements convexes de processus stochastiques. *Comptes Rendus de l'Académie des Sciences: Série Mathématique* 329, 1087–1090.
Davydov, Y. & A.M. Vershik (1998) Réarrangements convexes des marches aléatoires. *Annales de l'Institut Henri Poincaré: Probabilités et Statistiques* 34, 73–95.
Davydov, Y. & R. Zitikis (2001) Generalized Lorenz Curves and Convexifications of Stochastic Processes. Technical Report 359, Laboratory for Research in Statistics and Probability, Carleton University and University of Ottawa, Ottawa.
Davydov, Y. & R. Zitikis (2002) Convergence of generalized Lorenz curves based on stationary ergodic random sequences with deterministic noise. *Statistics and Probability Letters* 59, 329–340.
Dobrushin, R.L. (1970) Describing a system of random variables by conditional distributions. *Theory of Probability and Its Applications* 15, 458–486.
Gastwirth, J.L. (1971) A general definition of the Lorenz curve. *Econometrica* 39, 1037–1039.
Gastwirth, J.L. (1972) The estimation of the Lorenz curve and Gini index. *Review of Economics and Statistics* 54, 306–316.
Gastwirth, J.L. & M. Glauberman (1976) The interpolation of the Lorenz curve and Gini index from grouped data. *Econometrica* 44, 479–483.
Giorgi, G.-M. (1990) Bibliographic portrait of the Gini concentration ratio. *Metron* 48, 183–221.
Giorgi, G.-M. (1993) A fresh look at the topical interest of the Gini concentration ratio. *Metron* 51, 83–98.
Hoeffding, W. (1948) A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* 19, 293–325.
Kakwani, N.C. (1980a) *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. New York: Oxford University Press.
Kakwani, N. (1980b) On a class of poverty measures. *Econometrica* 48, 437–446.
Kendall, M.G. & A. Stuart (1958) *The Advanced Theory of Statistics*, vol. 1: *Distribution theory*. New York: Hafner Publishing.
Kolm, S.C. (1976a) Unequal inequalities, part I. *Journal of Economic Theory* 12, 416–442.
Kolm, S.C. (1976b) Unequal inequalities, part II. *Journal of Economic Theory* 13, 82–111.
Lorenz, M.C. (1905) Methods of measuring the concentration of wealth. *Journal of the American Statistical Association* 9, 209–219.
Mehran, F. (1976) Linear measures of income inequality. *Econometrica* 44, 805–809.
Nygård, F. & A. Sandström (1981) *Measuring Income Inequality.* Stockholm: Almqvist & Wiksell.
Parzen, E. (1979) Nonparametric statistical data modeling. With comments by John W. Tukey, Roy E. Welsch, William F. Eddy, D.V. Lindley, Michael E. Tarter, and Edwin L. Crow, and a rejoinder by the author. *Journal of the American Statistical Association* 74, 105–131.
Sen, A. (1997) *On Economic Inequality.* (Expanded edition with a substantial annexe by J.E. Foster & A. Sen.) Oxford: Clarendon Press.

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

Shao, J. (1994) *L*-statistics in complex survey problems. *Annals of Statistics* 22, 946–967.
Shao, Q.-M. & H. Yu (1996) Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* 24, 2098–2127.
Shorack, G.R. (2000) *Probability for Statisticians*. New York: Springer-Verlag.
Shorack, G.R. & J.A. Wellner (1986) *Empirical Processes with Applications to Statistics*. New York: Wiley.
Shorrocks, A.F. & J.E. Foster (1987) Transfer sensitive inequality measures. *Review of Economic Studies* 54, 485–497.
Silber, J. (ed.) (1999) *Handbook on Income Inequality Measurement*. Boston: Kluwer Academic Publishers.
Thilly, E. (1999) *Réarrangements Convexes des Trajectoires de Processus Stochastiques*. Ph.D. Thesis, University of Science and Technology of Lille, Lille.
Xu, K. (2000) Inference for generalized Gini indices using the iterated-bootstrap method. *Journal of Business and Economic Statistics* 18, 223–227.
Yitzhaki, S. (1991) Calculating jackknife variance estimators for parameters of the Gini method. *Journal of Business and Economic Statistics* 9, 235–239.
Yu, H. (1993) A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences. *Probabability Theory and Related Fields* 95, 357–370.
Zheng, B. (1997) Aggregate poverty measures. *Journal of Economic Surveys* 11, 123–162.
Zheng, B. & B.J. Cushing (2001) Statistical inference for testing inequality indices with dependent samples. *Journal of Econometrics* 101, 315–335.
Zheng, B. (2001) Statistical inference for poverty measures with relative poverty lines. *Journal of Econometrics* 10, 337–356.
Zitikis, R. (1998) The Vervaat process. In B. Szyszkowicz (ed.), *Asymptotic Methods in Probability and Statistics: A Volume in Honour of Miklós Csörgő*, pp. 667–694. Amsterdam: Elsevier.
Zitikis, R. (2002a). Analysis of indices of economic inequality from a mathematical point of view. (Invited Plenary Lecture at the 11th Indonesian Mathematics Conference, State University of Malang, Indonesia.) *Matematika* 8, 772–782.
Zitikis, R. (2002b) Large sample estimation of a family of economic inequality indices. *Pakistan Journal of Statistics* (Special Issue in Honour of Dr. S. Ejaz Ahmed) 18, 225–248.
Zitikis, R. & J.L. Gastwirth (2002) Asymptotic distribution of the S-Gini index. *Australian and New Zealand Journal of Statistics* 44, 439–446.

APPENDIX: PROOFS OF THEOREMS

We start with the note that the estimator $G_{n,\alpha}$ is a discretization of the following well-known and thoroughly investigated (cf., e.g., Barrett and Donald, 2000) *E*-Gini estimator:

$$G^*_{n,\alpha} := 2\left(\int_0^1 (t - L_n(t))^\alpha \, dt\right)^{1/\alpha}. \qquad \textbf{(A.1)}$$

Obviously, $G^*_{n,\alpha}$ is $G_{F,\alpha}$ with L_F replaced by L_n. We shall find $G^*_{n,\alpha}$ convenient in the proofs that follow.

We assume throughout the rest of the paper that the distribution function F is not degenerate. Otherwise, some technical details would arise, and one would have to deal

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

with them in special ways, whereas the desired asymptotic results are always trivial in the degenerate case.

Proof of Theorem 3.1. We shall first prove that

$$G_{n,\alpha} - G^*_{n,\alpha} \to_{a.s.} 0. \tag{A.2}$$

We start with the formula

$$y^p - x^p = (y-x)p \int_0^1 \{x + \tau(y-x)\}^{p-1}\, d\tau, \tag{A.3}$$

where p, x, and y are some positive real numbers to be specified subsequently. An application of formula (A.3) with $y = (1/n)\sum_{i=1}^n (i/n - L_n(i/n))^\alpha$, $x = \int_0^1 (t - L_n(t))^\alpha\, dt$, and $p = 1/\alpha$ shows that statement (A.2) follows if

$$\int_0^1 (t - L_n(t))^\alpha\, dt - \frac{1}{n}\sum_{i=1}^n \left(\frac{i}{n} - L_n\left(\frac{i}{n}\right)\right)^\alpha \to_{a.s.} 0. \tag{A.4}$$

To prove (A.4) we use formula (A.3) with $y = i/n - L_n(i/n)$, $x = t - L_n(t)$, and $p = \alpha$. In this way we obtain that, for any $t \in [(i-1)/n, i/n]$,

$$\begin{aligned}\left|\{t - L_n(t)\}^\alpha - \left\{\frac{i}{n} - L_n\left(\frac{i}{n}\right)\right\}^\alpha\right| &\le c\left|\{t - L_n(t)\} - \left\{\frac{i}{n} - L_n\left(\frac{i}{n}\right)\right\}\right| \\ &\le c\,\frac{1}{n} + c \sup_{i=1,\ldots,n}\left(L_n\left(\frac{i}{n}\right) - L_n\left(\frac{i-1}{n}\right)\right).\end{aligned} \tag{A.5}$$

From bound (A.5) we see that statement (A.4) follows if the supremum on the right-hand side of bound (A.5) converges to 0 almost surely. To prove this, we use the following bound:

$$\begin{aligned}\overline{X}\left(L_n\left(\frac{i}{n}\right) - L_n\left(\frac{i-1}{n}\right)\right) &= \int_{(i-1)/n}^{i/n} F_n^{-1}(s)\, ds \\ &\le \int_0^1 |F_n^{-1}(s) - F^{-1}(s)|\, ds + \sup_{i=1,\ldots,n}\int_{(i-1)/n}^{i/n} F^{-1}(s)\, ds.\end{aligned} \tag{A.6}$$

The first summand on the right-hand side of (A.6) is known in the literature as the Dobrushin distance between F_n and F (cf. Dobrushin, 1970). Thus,

$$d_1(F_n, F) := \int_0^1 |F_n^{-1}(s) - F^{-1}(s)|\, ds \to_{a.s} 0 \tag{A.7}$$

(cf., e.g., Shorack and Wellner, 1986, exercise 9, p. 66). To prove that the second summand on the right-hand side of (A.6) converges to 0, we first note that, as a result of $\mathbf{E}(X) < \infty$, the function $GL_F(t) = \int_0^t F^{-1}(s)\, ds$ is continuous and thus is uniformly continuous on [0,1]. This implies the convergence of the second summand on the right-hand side of (A.6) to 0. The proof of statement (A.2) is complete.

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

In view of (A.2), Theorem 3.1 follows if

$$G^*_{n,\alpha} \to_{a.s.} G_{F,\alpha}. \tag{A.8}$$

Choosing $y = \int_0^1 \{t - L_n(t)\}^\alpha \, dt$, $x = \int_0^1 \{t - L_F(t)\}^\alpha \, dt$, and $p = 1/\alpha$ in formula (A.3), we obtain that statement (A.8) is a consequence of the following statement:

$$\int_0^1 \{t - L_n(t)\}^\alpha \, dt - \int_0^1 \{t - L_F(t)\}^\alpha \, dt \to_{a.s.} 0. \tag{A.9}$$

Applying equality (A.3) once more but this time with $y = t - L_n(t)$, $x = t - L_F(t)$, and $p = \alpha$, we obtain (A.9) provided that

$$\int_0^1 |L_n(t) - L_F(t)| \, dt \to_{a.s.} 0. \tag{A.10}$$

To prove (A.10), we proceed with the bound

$$\begin{aligned}\int_0^1 |L_n(t) - L_F(t)| \, dt \le{}& \frac{1}{\overline{X}} \sup_{t\in(0,1)} \left| \int_0^t \{F_n^{-1}(s) - F^{-1}(s)\} \, ds \right| \\ &+ \left| \frac{1}{\overline{X}} - \frac{1}{\mu} \right| \sup_{t\in(0,1)} \left(\int_0^t F^{-1}(s) \, ds \right).\end{aligned} \tag{A.11}$$

The first summand on the right-hand side of (A.11) converges to 0 almost surely as a result of (A.7). The second summand converges to 0 because $\overline{X}$ is a strongly consistent estimator of μ, and the integral $\int_0^t F^{-1}(s) \, ds$ does not exceed $\mathbf{E}(X) < \infty$. The proof of statement (A.10) is now complete and thus that of (A.8) also. This proves Theorem 3.1. ■

Proof of Theorem 3.2. We shall first prove that

$$\sqrt{n}\{G^*_{n,\alpha} - G_{n,\alpha}\} \to_P 0. \tag{A.12}$$

The proof of (A.12) closely resembles that of statement (A.2). Therefore, we omit most of the details and only discuss the proof of the statement

$$\sqrt{n} \sup_{i=1,\dots,n} \left(L_n\left(\frac{i}{n}\right) - L_n\left(\frac{i-1}{n}\right) \right) \to_P 0. \tag{A.13}$$

We start the proof of (A.13) with the bound

$$\overline{X}\sqrt{n} \left(L_n\left(\frac{i}{n}\right) - L_n\left(\frac{i-1}{n}\right) \right) \le \left(\int_{(i-1)/n}^{i/n} \{F_n^{-1}(s)\}^2 \, ds \right)^{1/2}. \tag{A.14}$$

Next, we estimate the integral on the right-hand side of (A.14) as follows:

$$\begin{aligned}\int_{(i-1)/n}^{i/n} \{F_n^{-1}(s)\}^2 \, ds \le{}& 2 \int_0^1 \{F_n^{-1}(s) - F^{-1}(s)\}^2 \, ds \\ &+ 2 \sup_{i=1,\dots,n} \int_{(i-1)/n}^{i/n} \{F^{-1}(s)\}^2 \, ds.\end{aligned} \tag{A.15}$$

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

Because $\mathbf{E}(X^2) < \infty$, the function $\lambda(t) := \int_0^t \{F^{-1}(s)\}^2\, ds$ is continuous and, therefore, uniformly continuous on the interval [0,1]. This implies that the second summand on the right-hand side of (A.15) converges to 0. The first summand

$$d_2(F_n, F) := \int_0^1 \{F_n^{-1}(s) - F^{-1}(s)\}^2\, ds \tag{A.16}$$

is known in the literature as the Wasserstein distance between F_n and F. It is also known (cf., e.g., Shorack and Wellner, 1986, Corollary 1, p. 63) that $d_2(F_n, F) \to_{\text{a.s.}} 0$ under the assumption $\mathbf{E}(X^2) < \infty$. This proves (A.13) and thus (A.12) also.

In view of (A.12), Theorem 3.2 follows if

$$\sqrt{n}\{G^*_{n,\alpha} - G_{F,\alpha}\} \to_d \mathcal{N}(0, \sigma^2_{F,\alpha}), \tag{A.17}$$

where $\mathcal{N}(0, \sigma^2_{F,\alpha})$ denotes a mean zero normal random variable with the variance $\sigma^2_{F,\alpha}$. We proceed with the proof of (A.17) by rewriting formula (A.3) in the following form:

$$y^p - x^p = (y-x)px^{p-1} + (y-x)p\int_0^1 (\{x + \tau(y-x)\}^{p-1} - x^{p-1})\, d\tau. \tag{A.18}$$

Using (A.18) with $y = \int_0^1 \{t - L_n(t)\}^\alpha\, dt$, $x = \int_0^1 \{t - L_F(t)\}^\alpha\, dt$, and $p = 1/\alpha$, we obtain statement (A.17) if $2\sqrt{n}(y-x)px^{p-1} \to_d \mathcal{N}(0, \sigma^2_{F,\alpha})$. The latter can explicitly be written as follows:

$$2^\alpha G_{F,\alpha}^{1-\alpha}\frac{1}{\alpha}\sqrt{n}\int_0^1 (\{t - L_n(t)\}^\alpha - \{t - L_F(t)\}^\alpha)\, dt \to_d \mathcal{N}(0, \sigma^2_{F,\alpha}). \tag{A.19}$$

To prove (A.19), we first apply formula (A.18) with $y = t - L_n(t)$, $x = t - L_F(t)$, and $p = \alpha$. We obtain the representation

$$\{t - L_n(t)\}^\alpha - \{t - L_F(t)\}^\alpha = -\{L_n(t) - L_F(t)\}\alpha\{t - L_F(t)\}^{\alpha-1} - Q_{n,\alpha}(t). \tag{A.20}$$

The remainder term $Q_{n,\alpha}(t)$ on the right-hand side of (A.20) is defined by the formula

$$Q_{n,\alpha}(t) := \alpha\{L_n(t) - L_F(t)\}\int_0^1 ((\{t - L_F(t)\} - \tau\{L_n(t) - L_F(t)\})^{\alpha-1} - \{t - L_F(t)\}^{\alpha-1})\, d\tau.$$

Note that $Q_{n,\alpha}(t)$ is asymptotically smaller than the main (i.e., the first) term on the right-hand side of (A.20). Therefore, using (A.20) we obtain (A.19) if

$$2^\alpha G_{F,\alpha}^{1-\alpha}\sqrt{n}\int_0^1 \{L_n(t) - L_F(t)\}\{t - L_F(t)\}^{\alpha-1}\, dt \to_d \mathcal{N}(0, \sigma^2_{F,\alpha}). \tag{A.21}$$

To prove (A.21), we start with the equality

$$L_n(t) - L_F(t) = \frac{1}{\mu}\int_0^t \{F_n^{-1}(s) - F^{-1}(s)\}\, ds - \frac{\overline{X} - \mu}{\mu}L_F(t) + R_n(t), \tag{A.22}$$

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

where

$$R_n(t) := \left(\frac{1}{\overline{\overline{X}}} - \frac{1}{\mu}\right)\int_0^t \{F_n^{-1}(s) - F^{-1}(s)\}\,ds - (\overline{X} - \mu)\left(\frac{1}{\overline{\overline{X}}} - \frac{1}{\mu}\right)L_F(t).$$

Obviously,

$$|R_n(t)| \le \frac{1}{\overline{\overline{X}}\mu}\,|\overline{X} - \mu| \int_0^1 |F_n^{-1}(s) - F^{-1}(s)|\,ds + \frac{1}{\overline{\overline{X}}\mu}\,(\overline{X} - \mu)^2. \tag{A.23}$$

Using statement (A.7) and the fact that $\sqrt{n}\{\overline{X} - \mu\}$ is asymptotically bounded in probability, we obtain from (A.23) that $\sqrt{n}\int_0^1 |R_n(t)|\,dt$ converges to 0 in probability. Therefore, (A.22) implies statement (A.21) provided that

$$\sqrt{n}\int_0^1 \left(\int_0^t \{F_n^{-1}(s) - F^{-1}(s)\}\,ds + (-\overline{X} + \mu)L_F(t)\right)\{t - L_F(t)\}^{\alpha-1}\,dt \to_d \mathcal{N}(0, \Delta_{F,\alpha}^2). \tag{A.24}$$

To prove statement (A.24), we use the representation

$$\int_0^t \{F_n^{-1}(s) - F^{-1}(s)\}\,ds = -\int_{-\infty}^{F^{-1}(t)} \{F_n(x) - F(x)\}\,dx + V_n(t), \tag{A.25}$$

where V_n is the general Vervaat process (cf. Csörgő and Zitikis, 1996) defined by the formula

$$V_n(t) := \int_0^t \{F_n^{-1}(s) - F^{-1}(s)\}\,ds + \int_{-\infty}^{F^{-1}(t)} \{F_n(x) - F(x)\}\,dx.$$

It is well known (cf., e.g., Zitikis, 1998) that the Vervaat process V_n is nonnegative and does not exceed $-\{F_n(F^{-1}(t)) - t\}\{F_n^{-1}(t) - F^{-1}(t)\}$. Note that if t is a continuity point of F, then the latter statement can be written as the following inequality:

$$V_n(t) \le -\{F_n(F^{-1}(t)) - F(F^{-1}(t))\}\{F_n^{-1}(t) - F^{-1}(t)\}. \tag{A.26}$$

This implies the bound

$$\sqrt{n}\int_0^1 |V_n(t)|\{t - L_F(t)\}^{\alpha-1}\,dt \le \sqrt{n}\,\sup_{x\in\mathbf{R}}|F_n(x) - F(x)| \int_0^1 |F_n^{-1}(t) - F^{-1}(t)|\,dt. \tag{A.27}$$

Note that the use of (A.26) for proving (A.27) was legitimate because of the fact that the set of all discontinuity points of F has the Lebesgue measure 0. Thus, the value of the integral is not affected by removing the points. Now, by the classical Kolmogorov–Smirnov theorem, $\sqrt{n}\,\sup_{x\in\mathbf{R}}|F_n(x) - F(x)|$ is asymptotically bounded in probability. Furthermore, statement (A.7) implies that the integral on the right-hand side of (A.27) converges to 0 in probability. These facts taken together with (A.25) imply statement (A.24) provided that

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press

$$\sqrt{n}\int_0^1\left(\int_{-\infty}^{F^{-1}(t)}\{F_n(x)-F(x)\}\,dx+(-\overline{X}+\mu)L_F(t)\right)\{t-L_F(t)\}^{\alpha-1}\,dt\to_d\mathcal{N}(0,\Delta_{F,\alpha}^2). \tag{A.28}$$

The proof of (A.28) is straightforward. Note first that (cf., e.g., Shorack, 2000, formula (6), p. 116)

$$-\overline{X}+\mu=\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbf{R}}\{\mathbf{I}\{X_i\le x\}-F(x)\}\,dx. \tag{A.29}$$

Thus, the integral on the left-hand side of (A.28) is the sum $n^{-1/2}\sum_{i=1}^n Y_i$ of independent and identically distributed random variables $Y_1, Y_2, \ldots, Y_n$ defined by the formula

$$\begin{aligned}Y_i:=&\int_0^1\left(\int_{-\infty}^{F^{-1}(t)}\{\mathbf{I}\{X_i\le x\}-F(x)\}\,dx\right)\{t-L_F(t)\}^{\alpha-1}\,dt\\&-\left(\int_0^1\{\mathbf{I}\{X_i\le x\}-F(x)\}\,dx\right)\left(\int_{-\infty}^{\infty}L_F(t)\{t-L_F(t)\}^{\alpha-1}\,dt\right).\end{aligned}$$

By the classical central limit theorem, statement (A.28) holds if $\Delta_{F,\alpha}^2$ is the variance $\sigma_{Y_1}^2$ of Y_1. Straightforward calculations show that this is the case. The proof of (A.21) is complete. ■

Proof of Theorem 4.1. The proof of Theorem 4.1 follows the lines of the proof of Theorem 3.1. Naturally, the therein used assumption $\mathbf{E}(X)<\infty$ should now be replaced by the stronger one $\mathbf{E}(X^2)<\infty$. We omit the technical details, noting only that analogous calculations in the case of the asymptotic variance of the *S*-Gini index are carried out by Zitikis (2002b). ■

https://doi.org/10.1017/S0266466603194042 Published online by Cambridge University Press