



Munich Personal RePEc Archive

Zenga's new index of economic inequality, its estimation, and an analysis of incomes in Italy

Greselin, Francesca and Pasquazzi, Leo and Zitikis, Ricardas

Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali,, Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali,, Department of Statistical and Actuarial Sciences, University of Western Ontario,

August 2009

Online at <https://mpra.ub.uni-muenchen.de/17147/>

MPRA Paper No. 17147, posted 08 Sep 2009 14:03 UTC

ZENGA'S NEW INDEX OF ECONOMIC INEQUALITY, ITS ESTIMATION, AND AN ANALYSIS OF INCOMES IN ITALY

FRANCESCA GRESELIN

*Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali,
Università di Milano Bicocca, Milan, Italy.
E-mail: francesca.greselin@unimib.it*

LEO PASQUAZZI

*Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali,
Università di Milano Bicocca, Milan, Italy. E-mail: leo.pasquazzi@unimib.it*

RIČARDAS ZITIKIS

*Department of Statistical and Actuarial Sciences, University of Western Ontario,
London, Ontario N6A 5B7, Canada. E-mail: zitikis@stats.uwo.ca*

Abstract. For at least a century academics and governmental researchers have been developing measures that would aid them in understanding income distributions, their differences with respect to geographic regions, and changes over time periods. It is a challenging area due to a number of reasons, one of them being the fact that different measures, or indices, are needed to reveal different features of income distributions. Keeping also in mind that the notions of ‘poor’ and ‘rich’ are relative to each other, M. Zenga has recently proposed a new index of economic inequality. The index is remarkably insightful and useful, but deriving statistical inferential results has been a challenge. For example, unlike many other indices, Zenga’s new index does not fall into the classes of L -, U -, and V -statistics. In this paper we derive desired statistical inferential results, explore their performance in a simulation study, and then employ the results to analyze data from the Bank of Italy’s *Survey on Household Income and Wealth*.

Keywords and phrases: Zenga index, lower conditional expectation, upper conditional expectation, confidence interval, Bonferroni curve, Lorenz curve, Vervaat process.

1. INTRODUCTION

Measuring and analyzing incomes, losses, risks and other (non-negative) random outcomes, which we denote by X , has been an active and fruitful research area, particularly in the fields of econometrics and actuarial science. The Gini index has been arguably the most popular measure of inequality, with a number of extensions and generalizations available in the literature. Recently, keeping in mind that the notions of ‘poor’ and ‘rich’ are relative to each other, M. Zenga constructed an new index that reflects this relativity. We shall next introduce the Gini and Zenga indices in such a way that they would be easy to compare and interpret.

To proceed, we need additional notation. Let $F(x)$ denote the cumulative distribution function (cdf) of X , and let $F^{-1}(s)$ denote the corresponding quantile function. Furthermore, let μ_F denote the mean of X . In terms of the Lorenz curve $L_F(p) = \mu_F^{-1} \int_0^p F^{-1}(s) ds$ (see Pietra, 1915), the Gini index can be written as follows:

$$G_F = \int_0^1 \left(1 - \frac{L_F(p)}{p}\right) \psi(p) dp,$$

where $\psi(p) = 2p$, which is a density function on $[0, 1]$. Given the usual econometric interpretation of the Lorenz curve $L_F(p)$, the function

$$G_F(p) = 1 - \frac{L_F(p)}{p}$$

is a relative measure of inequality (see Gini, 1914), called the Gini curve. Indeed, $L_F(p)/p$ is the ratio between 1) the mean income of the poorest $p \times 100\%$ of the population and 2) the mean income of the entire population; the closer to each other these two means are, the lower is the inequality. Zenga’s (2007) index of inequality is defined by the formula

$$Z_F = \int_0^1 Z_F(p) dp, \tag{1.1}$$

where

$$Z_F(p) = 1 - \frac{L_F(p)}{p} \cdot \frac{1-p}{1-L_F(p)}, \tag{1.2}$$

called the Zenga curve, measures the inequality between 1) the poorest $p \times 100\%$ of the population and 2) the richer remaining (i.e., $(1-p) \times 100\%$) part of it by comparing the mean incomes of these two disjoint and exhaustive sub-populations. We shall elaborate on this interpretation later, in Section 5 below.

Both the Gini and Zenga indices are averages of point inequality measures, that is, of the Gini and Zenga curves, respectively, but while in the case of the Gini index the weight (i.e., density) function $\psi(p) = 2p$ is employed, in the case of the Zenga

index the uniform weight (i.e., density) function is used. As a consequence, the Gini index underestimates comparisons between the very poor and the whole population and emphasizes comparisons which involve almost identical population subgroups. From this point of view, the Zenga index is more impartial: it is based on all comparisons between complementary disjoint population subgroups and gives the same weight to each comparison. Hence, with the same sensibility, the index detects all deviations from equality in any part of the distribution.

To illustrate the Gini curve $G_F(p)$ and its weighted version $g_F(p) = G_F(p)\psi(p)$, and to also facilitate their comparisons with the Zenga curve $Z_F(p)$, we choose the Pareto distribution

$$F(x) = 1 - \left(\frac{x_0}{x}\right)^\theta, \quad x > x_0, \quad (1.3)$$

where $x_0 > 0$ and $\theta > 0$ are parameters. (We shall use this distribution in our simulation study later in this paper as well, setting $x_0 = 1$ and $\theta = 2.06$.) Corresponding to this distribution, the Lorenz curve is equal to $L_F(p) = 1 - (1-p)^{1-1/\theta}$ (see Gastwirth, 1971), and so the Gini curve is equal to $G_F(p) = ((1-p)^{1-1/\theta} - (1-p))/p$. In Figure 1.1 (left panel) we have depicted the Gini and weighted Gini curves. The corresponding Zenga curve is equal to $Z_F(p) = (1 - (1-p)^{1/\theta})/p$ and is depicted in Figure 1.1 (right panel) alongside the Gini curve $G_F(p)$ for an easy comparison.

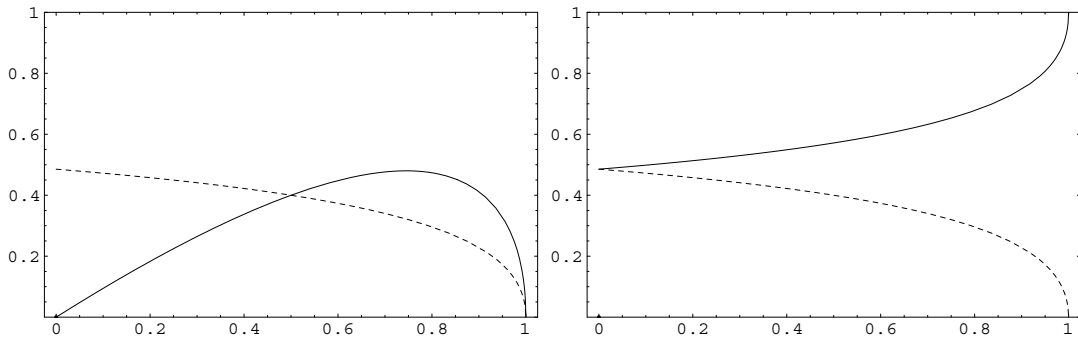


FIGURE 1.1. The Gini curve $G_F(p)$ (dashed; both panels), the weighted Gini curve $g_F(p)$ (solid; left panel), and the Zenga curve $Z_F(p)$ (solid; right panel) in the Pareto case with $x_0 = 1$ and $\theta = 2.06$.

The rest of this paper is organized as follows. In Section 2 we define two estimators of the Zenga index and develop statistical inferential results. In Section 3 we present results of a simulation study, which explores the empirical performance of the two empirical Zenga estimators, including their coverage accuracy and length of several types of confidence intervals. In Section 4 we present an analysis of data from the Bank of Italy's *Survey on Household Income and Wealth*. In Section 5 we further

contribute to understanding of the Zenga index by relating it to lower and upper conditional expectations. In Section 6 we provide a theoretical justification of the aforementioned two empirical Zenga estimators. In Section 7 we justify the definitions of several variance estimators as well as their uses in constructing confidence intervals. In Section 8 we prove Theorem 2.1 of Section 2, which is the main technical result of the present paper. Some technical lemmas and their proofs are relegated to Section 9.

2. ESTIMATORS AND STATISTICAL INFERENCE

Let X_1, \dots, X_n be independent copies of a random variable $X \geq 0$, which may, for example, represent incomes in the context of economic inequality, or risks and losses in the insurance context. We use two non-parametric estimators of the Zenga index. The first one (see Greselin and Pasquazzi, 2009) is given by the formula

$$\widehat{Z}_n = 1 - \frac{1}{n} \sum_{i=1}^{n-1} \frac{i^{-1} \sum_{k=1}^i X_{k:n}}{(n-i)^{-1} \sum_{k=i+1}^n X_{k:n}}, \quad (2.1)$$

where $X_{1:n} \leq \dots \leq X_{n:n}$ are the order statistics of X_1, \dots, X_n . With \bar{X} denoting the sample mean of X_1, \dots, X_n , the second estimator of the Zenga index is given by the formula

$$\begin{aligned} Z_n = & - \sum_{i=2}^n \frac{\sum_{k=1}^{i-1} X_{k:n} - (i-1)X_{i:n}}{\sum_{k=i+1}^n X_{k:n} + iX_{i:n}} \log \left(\frac{i}{i-1} \right) \\ & + \sum_{i=1}^{n-1} \left(\frac{\bar{X}}{X_{i:n}} - 1 - \frac{\sum_{k=1}^{i-1} X_{k:n} - (i-1)X_{i:n}}{\sum_{k=i+1}^n X_{k:n} + iX_{i:n}} \right) \log \left(1 + \frac{X_{i:n}}{\sum_{k=i+1}^n X_{k:n}} \right). \end{aligned} \quad (2.2)$$

The two estimators \widehat{Z}_n and Z_n are asymptotically equivalent. However, despite the fact that the estimator Z_n is obviously more complex, it is more convenient to work with when establishing asymptotic results, as we shall see later in this paper.

Unless explicitly stated otherwise, our following statistical inferential results are derived under the assumption that data are outcomes of independent and identically distributed (i.i.d.) random variables. It should be noted, however, that – as is the case in many surveys concerning income analysis – households are selected using complex sampling designs. In such cases statistical inferential tools and results are quite complex. To alleviate the difficulties, in the present paper we follow the commonly accepted practice and treat income data as if they were i.i.d. Certainly, extensions of our results to complex sampling designs would be an interesting and worthwhile contribution, though it would certainly be considerably more involved than the current one, which is already quite complex as we shall see later in the paper.

Unless explicitly stated otherwise, throughout we assume that the cdf F of X is a continuous function. We note that continuous cdf's are natural choices when modeling income distributions, insurance risks and losses (see, e.g., Kleiber and Kotz, 2003).

Theorem 2.1. *If the moment $\mathbf{E}[X^{2+\alpha}]$ is finite for some $\alpha > 0$, then we have the asymptotic representation*

$$\sqrt{n}(Z_n - Z_F) = \frac{1}{\sqrt{n}} \sum_{i=1}^n h(X_i) + o_{\mathbf{P}}(1), \quad (2.3)$$

where

$$h(X_i) = \int_0^\infty (\mathbf{1}\{X_i \leq x\} - F(x)) w_F(F(x)) dx$$

with the weight function

$$w_F(t) = -\frac{1}{\mu_F} \int_0^t \left(\frac{1}{p} - 1 \right) \frac{L_F(p)}{(1 - L_F(p))^2} dp + \frac{1}{\mu_F} \int_t^1 \left(\frac{1}{p} - 1 \right) \frac{1}{1 - L_F(p)} dp.$$

In view of Theorem 2.1, the asymptotic distribution of $\sqrt{n}(Z_n - Z_F)$ is centered normal with the variance $\sigma_F^2 = \mathbf{E}[h^2(X)]$, which is finite (see Theorem 7.1) and can be rewritten as follows:

$$\sigma_F^2 = \int_0^\infty \int_0^\infty (\min\{F(x), F(y)\} - F(x)F(y)) w_F(F(x)) w_F(F(y)) dx dy \quad (2.4)$$

or, alternatively,

$$\sigma_F^2 = \int_0^1 \left(\int_{[0,u]} t w_F(t) dF^{-1}(t) - \int_{[u,1]} (1-t) w_F(t) dF^{-1}(t) \right)^2 du. \quad (2.5)$$

The latter expression is particularly convenient when working with distributions for which the first derivative (when it exists) of $F^{-1}(t)$ is a simple function, as is the case for a large class of distributions (see, e.g., Karian and Dudewicz, 2000).

Irrespectively of what expression for the variance σ_F^2 we use, it is unknown since the cdf $F(x)$ is unknown. Replacing the cdf $F(x)$ on the right-hand side of equation (2.4) by the empirical cdf $F_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}$ where $\mathbf{1}$ denotes the indicator function, we obtain the following variance estimator (see Theorem 7.2 for details):

$$S_{X,n}^2 = \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} \left(\frac{\min\{k, l\}}{n} - \frac{k}{n} \frac{l}{n} \right) \times w_{X,n} \left(\frac{k}{n} \right) w_{X,n} \left(\frac{l}{n} \right) (X_{k+1:n} - X_{k:n})(X_{l+1:n} - X_{l:n}), \quad (2.6)$$

where

$$w_{X,n}(k/n) = - \sum_{i=1}^k I_{X,n}(i) + \sum_{i=k+1}^n J_{X,n}(i)$$

with the following expressions for the summands $I_{X,n}(i)$ and $J_{X,n}(i)$. First, we have

$$I_{X,n}(1) = - \frac{\sum_{k=2}^n X_{k:n} - (n-1)X_{1:n}}{(\sum_{k=1}^n X_{k:n})(\sum_{k=2}^n X_{k:n})} + \frac{1}{X_{1:n}} \log \left(1 + \frac{X_{1:n}}{\sum_{k=2}^n X_{k:n}} \right). \quad (2.7)$$

Furthermore, for every $i = 2, \dots, n-1$, we have

$$\begin{aligned} I_{X,n}(i) = & n \frac{\sum_{k=1}^{i-1} X_{k:n} - (i-1)X_{i:n}}{(\sum_{k=i+1}^n X_{k:n} + iX_{i:n})^2} \log \left(\frac{i}{i-1} \right) \\ & - \frac{(\sum_{k=i+1}^n X_{k:n} - (n-i)X_{i:n})(\sum_{k=1}^n X_{k:n})}{(\sum_{k=i+1}^n X_{k:n} + iX_{i:n})(\sum_{k=i+1}^n X_{k:n})(\sum_{k=i}^n X_{k:n})} \\ & + \left(\frac{1}{X_{i:n}} + n \frac{\sum_{k=1}^{i-1} X_{k:n} - (i-1)X_{i:n}}{(\sum_{k=i+1}^n X_{k:n} + iX_{i:n})^2} \right) \log \left(1 + \frac{X_{i:n}}{\sum_{k=i+1}^n X_{k:n}} \right) \end{aligned} \quad (2.8)$$

and

$$\begin{aligned} J_{X,n}(i) = & \frac{n}{\sum_{k=i+1}^n X_{k:n} + iX_{i:n}} \log \left(\frac{i}{i-1} \right) \\ & - \frac{\sum_{k=i+1}^n X_{k:n} - (n-i)X_{i:n}}{X_{i:n}(\sum_{k=i+1}^n X_{k:n} + iX_{i:n})} \log \left(1 + \frac{X_{i:n}}{\sum_{k=i+1}^n X_{k:n}} \right). \end{aligned} \quad (2.9)$$

Finally,

$$J_{X,n}(n) = \frac{1}{X_{n,n}} \log \left(\frac{n}{n-1} \right). \quad (2.10)$$

With the just defined estimator $S_{X,n}^2$ of the variance σ_F^2 , we have the asymptotic result

$$\frac{\sqrt{n}(Z_n - Z_F)}{S_{X,n}} \rightarrow_d \mathcal{N}(0, 1). \quad (2.11)$$

We shall next discuss variants of statement (2.11) in the case of two populations, when samples are independent and also when paired.

We start with the independent case. Namely, let the random variables $X_1, \dots, X_n \sim F$ and $Y_1, \dots, Y_m \sim H$ be independent within and between the two samples. Just like in the case of $F(x)$, we assume that the cdf $H(x)$ is continuous and $\mathbf{E}[Y^{2+\alpha}] < \infty$ for some $\alpha > 0$. Furthermore, we assume that the sample sizes n and m are comparable in the sense that there exists $\eta \in (0, 1)$ such that

$$\frac{m}{n+m} \rightarrow \eta \in (0, 1)$$

when n and m tend to infinity. Then from statement (2.3) and its counterpart for $Y_i \sim H$ we have that $\sqrt{nm/(n+m)}((Z_{X,n} - Z_{Y,m}) - (Z_F - Z_H))$ is asymptotically

normal with mean zero and the variance $\eta\sigma_F^2 + (1 - \eta)\sigma_H^2$. To estimate the variances σ_F^2 and σ_H^2 , we use $S_{X,n}^2$ and $S_{Y,n}^2$, respectively, and obtain the following result:

$$\frac{(Z_{X,n} - Z_{Y,m}) - (Z_F - Z_H)}{\sqrt{\frac{1}{n}S_{X,n}^2 + \frac{1}{m}S_{Y,m}^2}} \rightarrow_d \mathcal{N}(0, 1). \quad (2.12)$$

Consider now the case when the two samples $X_1, \dots, X_n \sim F$ and $Y_1, \dots, Y_m \sim H$ are paired. Thus, we have $m = n$ and know that the pairs $(X_1, Y_1), \dots, (X_n, Y_n)$ are independent and identically distributed, but nothing is assumed about the joint distribution of (X, Y) . As before, the cdf's $F(x)$ and $H(y)$ are continuous and have finite moments of the order $2 + \alpha$ for some $\alpha > 0$. From statement (2.3) and its analog for Y we have that $\sqrt{n}((Z_{X,n} - Z_{Y,n}) - (Z_F - Z_H))$ is asymptotically normal with mean zero and the variance $\sigma_{F,H}^2 = \mathbf{E}[(h(X) - h(Y))^2]$. The variance can of course be written as $\sigma_F^2 - 2\mathbf{E}[h(X)h(Y)] + \sigma_H^2$. With the already constructed estimators $S_{X,n}^2$ and $S_{Y,n}^2$, we are only left to construct an estimator for $\mathbf{E}[h(X)h(Y)]$. (Note that when X and Y are independent, then $\mathbf{P}[X \leq x, Y \leq y] = F(x)H(y)$ and the expectation $\mathbf{E}[h(X)h(Y)]$ vanishes.) To this end, we write the equation

$$\mathbf{E}[h(X)h(Y)] = \int_0^\infty \int_0^\infty (\mathbf{P}[X \leq x, Y \leq y] - F(x)H(y))w_F(F(x))w_H(H(y))dxdy.$$

Replacing the cdf's $F(x)$ and $H(y)$ everywhere on the right-hand side of the above equation by their respective estimators $F_n(x)$ and $H_n(y)$, we have (see Theorem 7.3 for details)

$$S_{X,Y,n} = \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} \left(\frac{1}{n} \sum_{i=1}^k \mathbf{1}\{Y_{(i,n)} \leq Y_{l:n}\} - \frac{k}{n} \frac{l}{n} \right) \times w_{X,n} \left(\frac{k}{n} \right) w_{Y,n} \left(\frac{l}{n} \right) (X_{k+1:n} - X_{k:n})(Y_{l+1:n} - Y_{l:n}), \quad (2.13)$$

where $Y_{(1,n)}, \dots, Y_{(n,n)}$ are the induced (by X_1, \dots, X_n) order statistics of Y_1, \dots, Y_n . (Note that when $Y \equiv X$, then $Y_{(i,n)} = Y_{i:n}$ and the sum $\sum_{i=1}^k \mathbf{1}\{Y_{(i,n)} \leq Y_{l:n}\}$ is equal to $\min\{k, l\}$; hence, estimator (2.13) coincides with estimator (2.6) as expected.) Consequently, $S_{X,n}^2 - 2S_{X,Y,n} + S_{Y,n}^2$ is an empirical estimator of $\sigma_{F,H}^2$ and so

$$\frac{\sqrt{n}(Z_{X,n} - Z_{Y,n}) - (Z_F - Z_H)}{\sqrt{S_{X,n}^2 - 2S_{X,Y,n} + S_{Y,n}^2}} \rightarrow_d \mathcal{N}(0, 1). \quad (2.14)$$

We conclude this section with a note that the above established asymptotic results (2.11), (2.12), and (2.14) are what we typically need when dealing with two populations, or two time periods, but extensions to more populations and/or time periods would be a worthwhile contribution.

3. A SIMULATION STUDY

We investigate the numerical performance of the estimators \widehat{Z}_n and Z_n by simulating data from Pareto distribution (1.3) with the parameters $x_0 = 1$ and $\theta = 2.06$, which give the value $Z_F = 0.6000$ that we approximately see in real income distributions. Following Davison and Hinkley (1997, Chapter 5), we compute four types of confidence intervals: normal, percentile, BCa, and t -bootstrap. For normal and studentized bootstrap confidence intervals we estimate the variance using empirical influence values. For the estimator Z_n , the influence values $h(X_i)$ have been obtained from Theorem 2.1, and those for the estimator \widehat{Z}_n using numerical differentiation.

In Table 3.1 we report coverage percentages of 10,000 confidence intervals, for

TABLE 3.1. Coverage proportions of confidence intervals from the Pareto parent distribution with $x_0 = 1$ and $\theta = 2.06$ ($Z_F = 0.6$).

	\widehat{Z}_n				Z_n			
	0.9000	0.9500	0.9750	0.9900	0.9000	0.9500	0.9750	0.9900
n	Normal confidence intervals							
200	0.7915	0.8560	0.8954	0.9281	0.7881	0.8527	0.8926	0.9266
400	0.8059	0.8705	0.9083	0.9409	0.8047	0.8693	0.9078	0.9396
800	0.8256	0.8889	0.9245	0.9514	0.8246	0.8882	0.9237	0.9503
n	Percentile confidence intervals							
200	0.7763	0.8326	0.8684	0.9002	0.7629	0.8190	0.8567	0.8892
400	0.8004	0.8543	0.8919	0.9218	0.7934	0.8487	0.8864	0.9179
800	0.8210	0.8777	0.9138	0.9415	0.8168	0.8751	0.9119	0.9393
n	BCa confidence intervals							
200	0.8082	0.8684	0.9077	0.9383	0.8054	0.867	0.9047	0.9374
400	0.8205	0.8863	0.9226	0.9531	0.8204	0.886	0.9212	0.9523
800	0.8343	0.8987	0.9331	0.9634	0.8338	0.8983	0.9323	0.9634
n	t -bootstrap confidence intervals							
200	0.8475	0.9041	0.9385	0.9658	0.8485	0.9049	0.9400	0.9675
400	0.8535	0.9124	0.9462	0.9708	0.8534	0.9120	0.9463	0.9709
800	0.8580	0.9168	0.9507	0.9758	0.8572	0.9169	0.9504	0.9754

each of the four types: normal, percentile, BCa, and t -bootstrap. Bootstrap based approximations have been obtained from 9999 resamples of the original samples. As suggested by Efron (1987), we have approximated the acceleration constant for the BCa

confidence intervals by one-sixth times the standardized third moment of the influence values. In Table 3.2 we report summary statistics concerning the size of the 10,000 confidence intervals. As expected, the confidence intervals based on \widehat{Z}_n and Z_n exhibit similar characteristics. We observe from Table 3.1 that all the confidence intervals suffer from some undercoverage. For example, about 97.5% of the studentized bootstrap confidence intervals with 0.99 nominal confidence level contain the true value of the Zenga index. It should be noted that the higher coverage accuracy of the studentized bootstrap confidence intervals (when compared to other ones) comes at the cost of their larger sizes, as seen in Table 3.2. Some of the studentized bootstrap confidence

TABLE 3.2. Size of the 95% asymptotic confidence intervals from the Pareto parent distribution with $x_0 = 1$ and $\theta = 2.06$ ($Z_F = 0.6$).

	\widehat{Z}_n			Z_n		
	min	mean	max	min	mean	max
n	Normal confidence intervals					
200	0.0680	0.1493	0.7263	0.0674	0.1500	0.7300
400	0.0564	0.1164	0.7446	0.0563	0.1167	0.7465
800	0.0462	0.0899	0.6528	0.0462	0.0900	0.6535
n	Percentile confidence intervals					
200	0.0673	0.1456	0.4751	0.0667	0.1462	0.4782
400	0.0561	0.1140	0.4712	0.0561	0.1143	0.4721
800	0.0467	0.0883	0.4110	0.0468	0.0884	0.4117
n	BCa confidence intervals					
200	0.0668	0.1491	0.4632	0.0661	0.1497	0.4652
400	0.0561	0.1183	0.4625	0.0558	0.1186	0.4629
800	0.0465	0.0925	0.4083	0.0467	0.0927	0.4085
n	t -bootstrap confidence intervals					
200	0.0677	0.2068	2.4307	0.0680	0.2099	2.5148
400	0.0572	0.1550	2.0851	0.0573	0.1559	2.1009
800	0.0473	0.1159	2.2015	0.0474	0.1162	2.2051

intervals extend beyond the range of the Zenga index, but this can easily be fixed by taking the minimum between the currently recorded upper bounds and 1, which is the upper bound of the Zenga index Z_F for every cdf F . We note that for the BCa confidence intervals, the number of bootstrap replications of the original sample has to

be increased beyond 9,999 if the nominal confidence level is high. Indeed, for samples of size 800, it turns out that the upper bound of 1,598 (out of 10,000) of the BCa confidence intervals based on \widehat{Z}_n and with 0.99 nominal confidence level is given by the largest order statistics of the bootstrap distribution. For the confidence intervals based on Z_n , the corresponding figure is 1,641.

4. AN ANALYSIS OF ITALIAN INCOME DATA

Here we use the Zenga index to analyze data from the Bank of Italy's *Survey on Household Income and Wealth*. The sample of the 2006 wave of this survey contain 7,768 households, with 3,957 of them being panel households. For detailed information on the survey, we refer to the Bank of Italy (2006) publication. In order to treat data correctly in the case of different household sizes, we work with equivalent incomes, which we have obtained by dividing the total household income by an equivalence coefficient, which is the sum of weights assigned to each household member. Following the modified OECD (Organization for Economic Cooperation and Development) equivalence scale, we give weight 1 to the household head, 0.5 to the other adult members of the household, and 0.3 to the members under 14 years of age.

In Table 4.1 we report the values of \widehat{Z}_n and Z_n according to the geographic area of households, and we also report confidence intervals for Z_F based on the two estimators. We note that two households in the sample had negative incomes in 2006 and so we have not included them in our computations. Consequently, the point estimates of Z_F are based on 7,766 equivalent incomes with values $\widehat{Z}_n = 0.6470$ and $Z_n = 0.6464$. As pointed out by Maasoumi (1994), however, good care is needed when comparing point estimates of inequality measures. Indeed, direct comparison of the point estimates corresponding to the five geographical areas of Italy would lead us to the erroneous conclusion that the inequality is higher in the central and southern areas when compared to the northern area and the islands. But as we glean from pairwise comparisons of the confidence intervals, only the differences between the estimates corresponding to the northwestern and southern areas and perhaps to the islands and the southern area may be deemed statistically significant.

Moreover, we have used the 3,957 panel households to check whether the Zenga inequality index has changed from the year 2004 to 2006. Table 4.2 reports the values of Z_n based on the panel households for these two years, and the 95% confidence intervals for the difference between the values of the Zenga index for the years 2006 and 2004. These computations have been based on formula (2.14). Removing the

TABLE 4.1. Confidence intervals for Z_F in the 2006 Italian income distribution

	\widehat{Z}_n estimator				Z_n estimator			
	95%		99%		95%		99%	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Northwest: $n = 1988$, $\widehat{Z}_n = 0.5953$, $Z_n = 0.5948$								
Normal	0.5775	0.6144	0.5717	0.6202	0.5771	0.6138	0.5713	0.6196
Student	0.5786	0.6168	0.5737	0.6240	0.5791	0.6172	0.5748	0.6243
Percent	0.5763	0.6132	0.5710	0.6193	0.5758	0.6124	0.5706	0.6185
BCa	0.5789	0.6160	0.5741	0.6234	0.5785	0.6156	0.5738	0.6226
Northeast: $n = 1723$, $\widehat{Z}_n = 0.6108$, $Z_n = 0.6108$								
Normal	0.5849	0.6393	0.5764	0.6478	0.5849	0.6393	0.5764	0.6479
Student	0.5874	0.6526	0.5796	0.6669	0.5897	0.6538	0.5836	0.6685
Percent	0.5840	0.6379	0.5773	0.6476	0.5839	0.6379	0.5772	0.6475
BCa	0.5894	0.6478	0.5841	0.6616	0.5894	0.6479	0.5842	0.6615
Center: $n = 1574$, $\widehat{Z}_n = 0.6316$, $Z_n = 0.6316$								
Normal	0.5957	0.6708	0.5839	0.6826	0.5956	0.6708	0.5838	0.6827
Student	0.5991	0.6991	0.5897	0.7284	0.6036	0.7016	0.5977	0.7311
Percent	0.5948	0.6689	0.5864	0.6818	0.5948	0.6688	0.5863	0.6818
BCa	0.6024	0.6850	0.5963	0.7021	0.6024	0.6850	0.5963	0.7020
South: $n = 1620$, $\widehat{Z}_n = 0.6557$, $Z_n = 0.6543$								
Normal	0.6358	0.6770	0.6293	0.6834	0.6346	0.6756	0.6282	0.6820
Student	0.6371	0.6805	0.6313	0.6902	0.6371	0.6796	0.6320	0.6900
Percent	0.6351	0.6757	0.6286	0.6828	0.6337	0.6742	0.6276	0.6812
BCa	0.6375	0.6793	0.6325	0.6888	0.6363	0.6778	0.6315	0.6873
Islands: $n = 861$, $\widehat{Z}_n = 0.6109$, $Z_n = 0.6095$								
Normal	0.5918	0.6317	0.5856	0.6380	0.5910	0.6302	0.5848	0.6364
Student	0.5927	0.6339	0.5864	0.6405	0.5928	0.6330	0.5874	0.6401
Percent	0.5897	0.6297	0.5839	0.6360	0.5885	0.6275	0.5831	0.6340
BCa	0.5923	0.6324	0.5868	0.6414	0.5914	0.6307	0.5860	0.6394
Italy (entire population): $n = 7766$, $\widehat{Z}_n = 0.6470$, $Z_n = 0.6464$								
Normal	0.6346	0.6596	0.6307	0.6636	0.6341	0.6591	0.6302	0.6630
Student	0.6359	0.6629	0.6327	0.6686	0.6358	0.6627	0.6331	0.6683
Percent	0.6348	0.6597	0.6314	0.6640	0.6343	0.6592	0.6309	0.6635
BCa	0.6363	0.6619	0.6334	0.6676	0.6358	0.6613	0.6330	0.6669

TABLE 4.2. 95% confidence intervals for the difference of the Zenga indices between 2006 and 2004 in the Italian income distribution

Northwest (926 pairs)		Northeast (841 pairs)		Center (831 pairs)	
$Z_n^{(2006)}$	0.5797	$Z_n^{(2006)}$	0.6199	$Z_n^{(2006)}$	0.5921
$Z_n^{(2004)}$	0.5955	$Z_n^{(2004)}$	0.6474	$Z_n^{(2004)}$	0.5766
Difference	-0.0158	Difference	-0.0275	Difference	0.0155
	Lower Upper		Lower Upper		Lower Upper
Normal	-0.0426 0.0102		-0.0573 0.0003		-0.0183 0.0514
Student	-0.0463 0.0103		-0.0591 0.0017		-0.0156 0.0644
Percent	-0.0421 0.0108		-0.0537 0.0040		-0.0183 0.0505
BCa	-0.0440 0.0087		-0.0551 0.0022		-0.0130 0.0593
South (843 pairs)		Islands (512 pairs)		Italy (3953 pairs)	
$Z_n^{(2006)}$	0.6200	$Z_n^{(2006)}$	0.6179	$Z_n^{(2006)}$	0.6362
$Z_n^{(2004)}$	0.6325	$Z_n^{(2004)}$	0.6239	$Z_n^{(2004)}$	0.6485
Difference	-0.0125	Difference	-0.0060	Difference	-0.0123
	Lower Upper		Lower Upper		Lower Upper
Normal	-0.0372 0.0129		-0.0333 0.0213		-0.0259 0.0007
Student	-0.0365 0.0166		-0.0351 0.0222		-0.0264 0.0013
Percent	-0.0372 0.0131		-0.0333 0.0214		-0.0253 0.0016
BCa	-0.0351 0.0162		-0.0331 0.0216		-0.0255 0.0013

four households with at least one negative income in the paired sample, we are left with a total of 3,953 observations. As before, we see that even though we deal with large sample sizes, the point estimates alone are not reliable. Indeed, for Italy as the whole and for all geographic areas except the center, the point estimates suggest that the Zenga index decreased from the year 2004 to 2006. However, the 95% confidence intervals in Table 4.2 suggest that this change is not significant.

5. AN ALTERNATIVE LOOK AT THE ZENGA INDEX

In various contexts we have notions of rich and poor, large and small, risky and secure. They divide the underlying population into two parts, which we can view as sub-populations. The quantile

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}$$

for some $p \in (0, 1)$ usually serves as a boundary separating the two sub-populations. For example, we may define ‘rich’ if $X > F^{-1}(p)$ and ‘poor’ if $X \leq F^{-1}(p)$. Calculating the

mean value of the former sub-population gives rise to the upper conditional expectation $\mathbf{E}[X|X > F^{-1}(p)]$, which is known in the actuarial risk theory as the conditional tail expectation. Calculating the mean value of the latter sub-population gives rise to the lower conditional expectation $\mathbf{E}[X|X \leq F^{-1}(p)]$, which is known in the econometric literature as the absolute Bonferroni curve, as a function of p . The ratio

$$R_F(p) = \frac{\mathbf{E}[X|X \leq F^{-1}(p)]}{\mathbf{E}[X|X > F^{-1}(p)]}$$

of the lower and upper conditional expectations takes on values in the interval $[0, 1]$, as we show in the next lemma.

Lemma 5.1. *For every $p \in (0, 1)$, we have that $R_F(p) \in [0, 1]$.*

Proof. We rewrite the ratio $R_F(p)$ as follows:

$$R_F(p) = \frac{\mathbf{E}[Xw_1(X)]}{\mathbf{E}[w_1(X)]} \bigg/ \frac{\mathbf{E}[Xw_2(X)]}{\mathbf{E}[w_2(X)]}, \quad (5.1)$$

where $w_1(x) = -\mathbf{1}\{x \leq F^{-1}(p)\}$ and $w_2(x) = \mathbf{1}\{x > F^{-1}(p)\}$. Both functions $w_1(x)$ and $w_2(x)$ are non-decreasing, and so by Lemma 3 on p. 1140 of Lehmann (1966) we have that $\mathbf{E}[Xw_1(X)] \geq \mathbf{E}[X]\mathbf{E}[w_1(X)]$ and $\mathbf{E}[Xw_2(X)] \geq \mathbf{E}[X]\mathbf{E}[w_2(X)]$. Hence, the ratio $\mathbf{E}[Xw_1(X)]/\mathbf{E}[w_1(X)]$ is not larger than $\mathbf{E}[X]$ (note that $\mathbf{E}[w_1(X)]$ is negative) and the ratio $\mathbf{E}[Xw_2(X)]/\mathbf{E}[w_2(X)]$ is not smaller than $\mathbf{E}[X]$. Consequently, the right-hand side of equation (5.1) does not exceed 1. This proves Lemma 5.1. \square

When X is a constant, which can be interpreted as ‘egalitarian’ case, then $R_F(p)$ is equal to 1. The ratio $R_F(p)$ is equal to 0 for all $p \in (0, 1)$ when the lower conditional expectation is equal to 0 for all $p \in (0, 1)$ which means extreme inequality in the sense that, loosely speaking, there is only one individual who possesses the entire wealth. Our wish to associate the egalitarian case with 0 and the extreme inequality with 1 leads to curve $1 - R_F(p)$, which coincides with the Zenga curve (see equation (1.2)) when the cdf F is continuous. The area

$$1 - \int_0^1 \frac{\mathbf{E}[X|X \leq F^{-1}(p)]}{\mathbf{E}[X|X > F^{-1}(p)]} dp \quad \left(= Z_F \text{ when } F \text{ is continuous} \right) \quad (5.2)$$

beneath the curve $1 - R_F(p)$ is always in the interval $[0, 1]$ as follows from Lemma 5.1. Quantity (5.2) is a measure of inequality and coincides with the earlier defined Zenga index when the cdf F is continuous, which we assume throughout the paper. Note that under this assumption, the lower and upper conditional expectations are equal

to the absolute Bonferroni curve $p^{-1}AL_F(p)$ and the dual absolute Bonferroni curve $(1-p)^{-1}(\mu_F - AL_F(p))$, respectively, where

$$AL_F(p) = \int_0^p F^{-1}(t)dt$$

is the absolute Lorenz curve. This leads us to the expression of the Zenga index Z_F given by equation (1.1), which we rewrite in terms of the just introduced *absolute* Lorenz curve as follows:

$$Z_F = 1 - \int_0^1 \left(\frac{1}{p} - 1 \right) \frac{AL_F(p)}{\mu_F - AL_F(p)} dp. \quad (5.3)$$

We shall extensively use expression (5.3) in the proofs below.

6. A CLOSER LOOK AT THE TWO ZENGA ESTIMATORS

Since samples are ‘discrete populations’, equations (5.2) and (5.3) lead to slightly different empirical estimators of Z_F . If we choose equation (1.1), then we arrive at the estimator \widehat{Z}_n , as seen from the proof of the following theorem.

Theorem 6.1. *The empirical Zenga index \widehat{Z}_n is an empirical estimator of Z_F .*

Proof. Let U be a uniform on $[0, 1]$ random variable independent of X . The cdf of $F^{-1}(U)$ is F . Hence, we have the following equations:

$$\begin{aligned} Z_F &= 1 - \mathbf{E}_U \left(\frac{\mathbf{E}_X[X|X \leq F^{-1}(U)]}{\mathbf{E}_X[X|X > F^{-1}(U)]} \right) \\ &= 1 - \int_{(0,\infty)} \frac{1 - F(x)}{F(x)} \frac{\mathbf{E}[X\mathbf{1}\{X \leq x\}]}{\mathbf{E}[X\mathbf{1}\{X > x\}]} dF(x) \\ &= 1 - \int_{(0,\infty)} \frac{1 - F(x)}{F(x)} \frac{\int_{(0,x]} y dF(y)}{\int_{(x,\infty)} y dF(y)} dF(x). \end{aligned} \quad (6.1)$$

Replacing every F on the right-hand side of equation (6.1) by F_n , we obtain

$$1 - \frac{1}{n} \sum_{i=1}^{n-1} \frac{1 - F_n(X_{i:n})}{F_n(X_{i:n})} \frac{\sum_{k=1}^n X_{k:n} \mathbf{1}\{X_{k:n} \leq X_{i:n}\}}{\sum_{k=1}^n X_{k:n} \mathbf{1}\{X_{k:n} > X_{i:n}\}},$$

which simplifies to

$$1 - \frac{1}{n} \sum_{i=1}^{n-1} \frac{1 - i/n}{i/n} \frac{\sum_{k=1}^i X_{k:n}}{\sum_{k=i+1}^n X_{k:n}}.$$

This is the estimator \widehat{Z}_n . □

If we choose equation (5.3) as the starting point for constructing an estimator for Z_F , then we replace the quantile $F^{-1}(p)$ by its empirical counterpart

$$\begin{aligned} F_n^{-1}(p) &= \inf\{x : F_n(x) \geq p\}. \\ &= X_{i:n} \quad \text{when } p \in ((i-1)/n, i/n] \end{aligned}$$

in the definition of $AL_F(p)$, which gives us the empirical absolute Lorenz curve $AL_n(p)$, and then we replace each $AL_F(p)$ on the right-hand side of equation (5.3) by the just constructed $AL_n(p)$. (Note that $\mu_F = AL_F(1) \approx AL_n(1) = \bar{X}$.) This gives us the empirical Zenga index Z_n as seen from the proof of the following theorem.

Theorem 6.2. *The empirical Zenga index Z_n is an estimator of Z_F .*

Proof. By construction, the estimator Z_n is given by the equation:

$$Z_n = 1 - \int_0^1 \left(\frac{1}{p} - 1 \right) \frac{AL_n(p)}{\bar{X} - AL_n(p)} dp. \quad (6.2)$$

Hence, the proof of the lemma reduces to verifying that the right-hand sides of equations (2.2) and (6.2) coincide. For this, we split the integral in equation (6.2) into the sum of integrals over the intervals $((i-1)/n, i/n)$ for $i = 1, \dots, n$. For every $p \in ((i-1)/n, i/n)$, we have $AL_n(p) = C_{i,n} + pX_{i:n}$, where

$$C_{i,n} = \frac{1}{n} \sum_{k=1}^{i-1} X_{k:n} - \frac{i-1}{n} X_{i:n}. \quad (6.3)$$

Hence, equation (6.2) can be rewritten as $Z_n = \sum_{i=1}^n \zeta_{i,n}$, where

$$\zeta_{i,n} = \frac{1}{n} - \int_{(i-1)/n}^{i/n} \left(\frac{1}{p} - 1 \right) \frac{\Lambda_{i,n} + p}{\Psi_{i,n} - p} dp$$

with

$$\Lambda_{i,n} = \frac{C_{i,n}}{X_{i:n}} \quad \text{and} \quad \Psi_{i,n} = \frac{\bar{X} - C_{i,n}}{X_{i:n}}. \quad (6.4)$$

Consider the case $i = 1$. We have $C_{1,n} = 0$ and thus $\Lambda_{1,n} = 0$, which implies

$$\zeta_{1,n} = \left(\frac{\bar{X}}{X_{1:n}} - 1 \right) \log \left(1 + \frac{X_{1:n}}{\sum_{k=2}^n X_{k:n}} \right).$$

Next, consider the case $i = n$. We have $C_{n,n} = \bar{X} - X_{n:n}$ and thus $\Psi_{n,n} = 1$, which implies

$$\zeta_{n,n} = \left(1 - \frac{\bar{X}}{X_{n:n}} \right) \log \left(\frac{n}{n-1} \right).$$

When $2 \leq i \leq n-1$, then the integrand in the definition of $\zeta_{i,n}$ does not have any singularity, since $\Psi_{i,n} > i/n$ due to $\sum_{k=i+1}^n X_{k:n} > 0$ almost surely. Hence, after simple integration we have that, for $i = 2, \dots, n-1$,

$$\begin{aligned} \zeta_{i,n} &= \frac{(i-1)X_{i:n} - \sum_{k=1}^{i-1} X_{k:n}}{\sum_{k=i+1}^n X_{k:n} + iX_{i:n}} \log\left(\frac{i}{i-1}\right) \\ &\quad + \left(\frac{\bar{X}}{X_{i:n}} - 1 + \frac{(i-1)X_{i:n} - \sum_{k=1}^{i-1} X_{k:n}}{\sum_{k=i+1}^n X_{k:n} + iX_{i:n}}\right) \log\left(1 + \frac{X_{i:n}}{\sum_{k=i+1}^n X_{k:n}}\right). \end{aligned}$$

With the above formulas for $\zeta_{i,n}$ we easily check that the sum $\sum_{i=1}^n \zeta_{i,n}$ is equal to the right-hand side of equation (2.2). This completes the proof of Theorem 6.2. \square

7. A CLOSER LOOK AT VARIANCES

Following the formulation of Theorem 2.1 we claimed that the asymptotic distribution of $\sqrt{n}(Z_n - Z_F)$ is centered normal with the *finite* variance $\sigma_F^2 = \mathbf{E}[h^2(X)]$. The following theorem provides a proof of this claim.

Theorem 7.1. *When $\mathbf{E}[X^{2+\alpha}] < \infty$ for some $\alpha > 0$, then $n^{-1/2} \sum_{i=1}^n h(X_i)$ converges in distribution to the centered normal random variable*

$$\Gamma = \int_0^\infty \mathcal{B}(F(x)) w_F(F(x)) dx,$$

where \mathcal{B} is the Brownian bridge on the interval $[0, 1]$. The variance of Γ is finite and equal to σ_F^2 .

Proof. Note that $n^{-1/2} \sum_{i=1}^n h(X_i)$ can be written as $\int_0^\infty e_n(F(x)) w_F(F(x)) dx$, where $e_n(p) = \sqrt{n}(E_n(p) - p)$ is the empirical process based on the uniform on $[0, 1]$ random variables $U_i = F(X_i)$, $i = 1, \dots, n$. We shall next show that

$$\int_0^\infty e_n(F(x)) w_F(F(x)) dx \rightarrow_d \int_0^\infty \mathcal{B}(F(x)) w_F(F(x)) dx. \quad (7.1)$$

The proof is based on the well known fact that, for every $\varepsilon > 0$,

$$\left\{ \frac{e_n(p)}{p^{1/2-\varepsilon}(1-p)^{1/2-\varepsilon}}, 0 \leq p \leq 1 \right\} \Rightarrow \left\{ \frac{\mathcal{B}(p)}{p^{1/2-\varepsilon}(1-p)^{1/2-\varepsilon}}, 0 \leq p \leq 1 \right\}.$$

Hence, in order to prove statement (7.1), we only need to check that the integral

$$\int_0^\infty F(x)^{1/2-\varepsilon}(1-F(x))^{1/2-\varepsilon} w_F(F(x)) dx \quad (7.2)$$

is finite. For this, by considering the two cases $p \leq 1/2$ and $p > 1/2$ separately, we easily show that $|w_F(p)| \leq c + c \log(1/p) + c \log(1/(1-p))$. Hence, for every $\varepsilon > 0$, there exists a constant $c < \infty$ such that, for all $p \in (0, 1)$,

$$|w_F(p)| \leq \frac{c}{p^\varepsilon(1-p)^\varepsilon}. \quad (7.3)$$

Bound (7.3) implies that integral (7.2) is finite provided that $\int_0^\infty (1-F(x))^{1/2-2\varepsilon} dx$ is finite, which is true since the moment $\mathbf{E}[X^{2+\alpha}]$ is finite for some $\alpha > 0$ and the parameter $\varepsilon > 0$ can be chosen as small as desired. Hence, $n^{-1/2} \sum_{i=1}^n h(X_i) \rightarrow_d \Gamma$ with Γ denoting the integral on the right-hand side of statement (7.1). The random variable Γ is normal because the Brownian bridge is a Gaussian process. Furthermore, Γ has mean zero because $\mathcal{B}(p)$ has mean zero for every $p \in [0, 1]$. The variance of Γ is equal to σ_F^2 because $\mathbf{E}[\mathcal{B}(p)\mathcal{B}(q)] = \min\{p, q\} - pq$ for all $p, q \in [0, 1]$. We are left to show that $\mathbf{E}[\Gamma^2] < \infty$. For this, we write the bound:

$$\begin{aligned} \mathbf{E}[\Gamma^2] &= \int_0^\infty \int_0^\infty \mathbf{E}[\mathcal{B}(F(x))\mathcal{B}(F(y))] w_F(F(x)) w_F(F(y)) dx dy \\ &\leq \left(\int_0^\infty \sqrt{\mathbf{E}[\mathcal{B}^2(F(x))]} w_F(F(x)) dx \right)^2. \end{aligned} \quad (7.4)$$

Since $\mathbf{E}[\mathcal{B}^2(F(x))] = F(x)(1-F(x))$, the finiteness of the integral on the right-hand side of bound (7.4) follows from the earlier proved statement that integral (7.2) is finite. Hence, $\mathbf{E}[\Gamma^2] < \infty$, which concludes the proof of Theorem 7.1. \square

Theorem 7.2. *The empirical variance $S_{X,n}^2$ is an estimator of σ_F^2 .*

Proof. We construct an empirical estimator for σ_F^2 by replacing every $F(x)$ on the right-hand side of equation (2.4) by the empirical $F_n(x)$. In particular, we replace the function $w_F(t)$ by its empirical version

$$w_{X,n}(t) = - \int_0^t \left(\frac{1}{p} - 1 \right) \frac{AL_n(p)}{(\bar{X} - AL_n(p))^2} dp + \int_t^1 \left(\frac{1}{p} - 1 \right) \frac{1}{\bar{X} - AL_n(p)} dp.$$

We denote the just defined estimator of σ_F^2 by $S_{X,n}^2$, and the rest of the proof consists of showing that the estimator $S_{X,n}^2$ coincides with the one defined by equation (2.6). Note that $\min\{F_n(x), F_n(y)\} - F_n(x)F_n(y) = 0$ when $x \in [0, X_{1:n}) \cup [X_{n:n}, \infty)$ and/or $y \in [0, X_{1:n}) \cup [X_{n:n}, \infty)$. Hence, the just defined $S_{X,n}^2$ is equal to

$$\int_{X_{1:n}}^{X_{n:n}} \int_{X_{1:n}}^{X_{n:n}} (\min\{F_n(x), F_n(y)\} - F_n(x)F_n(y)) w_{X,n}(F_n(x)) w_{X,n}(F_n(y)) dx dy.$$

Since $F_n(x) = k/n$ when $x \in [X_{k:n}, X_{k+1:n})$, we therefore have that

$$S_{X,n}^2 = \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} \left(\frac{\min\{k, l\}}{n} - \frac{k}{n} \frac{l}{n} \right) \\ \times w_{X,n} \left(\frac{k}{n} \right) w_{X,n} \left(\frac{l}{n} \right) (X_{k+1:n} - X_{k:n})(X_{l+1:n} - X_{l:n}).$$

Furthermore,

$$w_{X,n} \left(\frac{k}{n} \right) = - \int_0^{k/n} \left(\frac{1}{p} - 1 \right) \frac{AL_n(p)}{(\bar{X} - AL_n(p))^2} dp + \int_{k/n}^1 \left(\frac{1}{p} - 1 \right) \frac{1}{\bar{X} - AL_n(p)} dp \\ = - \sum_{i=1}^k I_{X,n}(i) + \sum_{i=k+1}^n J_{X,n}(i), \quad (7.5)$$

where, using notations (6.3) and (6.4), the summands on the right-hand side of equation (7.5) are:

$$I_{X,n}(i) = \frac{1}{X_{i:n}} \int_{(i-1)/n}^{i/n} \left(\frac{1}{p} - 1 \right) \frac{\Lambda_{i,n} + p}{(\Psi_{i,n} - p)^2} dp$$

for all $i = 1, \dots, n-1$, and

$$J_{X,n}(i) = \frac{1}{X_{i:n}} \int_{(i-1)/n}^{i/n} \left(\frac{1}{p} - 1 \right) \frac{1}{\Psi_{i,n} - p} dp$$

for all $i = 2, \dots, n$. When $i = 1$, then $\Lambda_{i,n} = 0$, and we easily check the expression for $I_{X,n}(1)$ given by equation (2.7). When $2 \leq i \leq n-1$, then

$$I_{X,n}(i) = \frac{\Lambda_{i,n}}{X_{i:n} \Psi_{i,n}^2} \log \left(\frac{i}{i-1} \right) - \frac{(\Lambda_{i,n} + \Psi_{i,n})(\Psi_{i,n} - 1)}{n X_{i:n} \Psi_{i,n} (\Psi_{i,n} - (i-1)/n) (\Psi_{i,n} - i/n)} \\ + \frac{1}{X_{i:n}} \left(1 + \frac{\Lambda_{i,n}}{\Psi_{i,n}^2} \right) \log \left(\frac{\Psi_{i,n} - (i-1)/n}{\Psi_{i,n} - i/n} \right),$$

and, after some algebra, we arrive at the right-hand side of equation (2.8). When $2 \leq i \leq n-1$, then we have

$$J_{X,n}(i) = \frac{1}{X_{i:n} \Psi_{i,n}} \log \left(\frac{i}{i-1} \right) - \frac{1}{X_{i:n}} \left(1 - \frac{1}{\Psi_{i,n}} \right) \log \left(\frac{\Psi_{i,n} - (i-1)/n}{\Psi_{i,n} - i/n} \right),$$

which, after some algebra, becomes equation (2.9). When $i = n$, then $\Psi_{i,n} = 1$, and we thus easily see that $J_{X,n}(i)$ is given by equation (2.10). This completes the proof of Theorem 7.2. \square

Theorem 7.3. *The empirical mixed moment $S_{X,Y,n}$ is an estimator of $\mathbf{E}[h(X)h(Y)]$.*

Proof. We proceed similarly to the proof of Theorem 7.2. We estimate the integrand $\mathbf{P}[X \leq x, Y \leq y] - F(x)H(y)$ using

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq x, Y_i \leq y\} - \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\} \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{Y_i \leq y\}. \quad (7.6)$$

After some rearrangement of terms, estimator (7.6) becomes

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_{i:n} \leq x, Y_{(i,n)} \leq y\} - \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_{i:n} \leq x\} \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{Y_{i:n} \leq y\}. \quad (7.7)$$

When $x \in [X_{k:n}, X_{k+1:n})$ and $y \in [Y_{l:n}, Y_{l+1:n})$, then estimator (7.7) is $n^{-1} \sum_{i=1}^k \mathbf{1}\{Y_{(i,n)} \leq Y_{l:n}\} - (k/n)(l/n)$, which leads us to the estimator $S_{X,Y,n}$ and thus completes the proof of Theorem 7.3. \square

8. PROOF OF THEOREM 2.1

Throughout the proof we conveniently use the notation $AL_F^*(p)$ for the dual absolute Lorenz curve $\int_p^1 F^{-1}(t)dt$, which is equal to $\mu_F - AL_F(p)$. Likewise, we use the notation $AL_n^*(p)$ for the empirical dual absolute Lorenz curve. Hence,

$$\sqrt{n}(Z_n - Z_F) = -\sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \left(\frac{AL_n(p)}{AL_n^*(p)} - \frac{AL_F(p)}{AL_F^*(p)}\right) dp.$$

Simple algebra gives the representation

$$\begin{aligned} \sqrt{n}(Z_n - Z_F) &= -\sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_n(p) - AL_F(p)}{AL_F^*(p)} dp \\ &\quad + \sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^{*2}(p)} (AL_n^*(p) - AL_F^*(p)) dp \\ &\quad - r_{n,1} + r_{n,2}, \end{aligned} \quad (8.1)$$

where the two remainder terms are:

$$r_{n,1} = \sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) (AL_n(p) - AL_F(p)) \left(\frac{1}{AL_n^*(p)} - \frac{1}{AL_F^*(p)}\right) dp$$

and

$$r_{n,2} = \sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^*(p)} (AL_n^*(p) - AL_F^*(p)) \left(\frac{1}{AL_n^*(p)} - \frac{1}{AL_F^*(p)}\right) dp.$$

We shall later show (Lemmas 9.1 and 9.2 below) that the remainder terms $r_{n,1}$ and $r_{n,2}$ are of the order $\mathbf{o}_{\mathbf{P}}(1)$. Hence, we proceed with an analysis of the first two terms on the right-hand side of equation (8.1), for which we use the (general) Vervaat process

$$V_n(p) = \int_0^p (F_n^{-1}(t) - F^{-1}(t)) dt + \int_0^{F^{-1}(p)} (F_n(x) - F(x)) dx \quad (8.2)$$

and its dual version

$$V_n^*(p) = \int_p^1 (F_n^{-1}(t) - F^{-1}(t))dt + \int_{F^{-1}(p)}^\infty (F_n(x) - F(x))dx. \quad (8.3)$$

For mathematical and historical details on the Vervaat process, see Zitikis (1998), Greselin et al. (2009), and references therein. Since $\int_0^1 (F_n^{-1}(t) - F^{-1}(t))dt = \bar{X} - \mu_F$ and $\int_0^\infty (F_n(x) - F(x))dx = -(\bar{X} - \mu_F)$, adding the right-hand sides of equations (8.2) and (8.3) gives the equation $V_n^*(p) = -V_n(p)$. Hence, whatever upper bound we have for $|V_n(p)|$, the same bound also holds for $|V_n^*(p)|$. In fact, the absolute value can be dropped from $|V_n(p)|$ since $V_n(p)$ is non-negative. Among other facts that we know about $V_n(p)$ is that it does not exceed $(p - F_n(F^{-1}(p)))(F_n^{-1}(p) - F^{-1}(p))$. Hence, with $e_n(p) = \sqrt{n}(F_n(F^{-1}(p)) - p)$, which is the uniform on $[0, 1]$ empirical process, we have that

$$\sqrt{n} V_n(p) \leq |e_n(p)| |F_n^{-1}(p) - F^{-1}(p)|. \quad (8.4)$$

Bound (8.4) implies the following asymptotic representation for the first term on the right-hand side of equation (8.1):

$$\begin{aligned} & -\sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_n(p) - AL_F(p)}{AL_F^*(p)} dp \\ &= \sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{1}{AL_F^*(p)} \left(\int_0^{F^{-1}(p)} (F_n(x) - F(x))dx \right) dp + O_{\mathbf{P}}(r_{n,3}), \end{aligned} \quad (8.5)$$

where

$$r_{n,3} = \int_0^1 \left(\frac{1}{p} - 1\right) \frac{1}{AL_F^*(p)} |e_n(p)| |F_n^{-1}(p) - F^{-1}(p)| dp.$$

We shall later show (Lemma 9.3 below) that $r_{n,3} = o_{\mathbf{P}}(1)$. Furthermore, we have the following asymptotic representation for the second term on the right-hand side of equation (8.1):

$$\begin{aligned} & \sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^{*2}(p)} (AL_n^*(p) - AL_F^*(p)) dp \\ &= -\sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^{*2}(p)} \left(\int_{F^{-1}(p)}^\infty (F_n(x) - F(x))dx \right) dp + O_{\mathbf{P}}(r_{n,4}), \end{aligned} \quad (8.6)$$

where

$$r_{n,4} = \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^{*2}(p)} |e_n(p)| |F_n^{-1}(p) - F^{-1}(p)| dp.$$

We shall later show (Lemma 9.4 below) that $r_{n,4} = o_{\mathbf{P}}(1)$. Hence, equations (8.1), (8.5) and (8.6) together with the statements $r_{n,1}, \dots, r_{n,4} = o_{\mathbf{P}}(1)$ imply that

$$\begin{aligned} \sqrt{n}(Z_n - Z_F) &= \sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{1}{AL_F^*(p)} \left(\int_0^{F^{-1}(p)} (F_n(x) - F(x)) dx \right) dp \\ &\quad - \sqrt{n} \int_0^1 \left(\frac{1}{p} - 1\right) \frac{AL_F(p)}{AL_F^{*2}(p)} \left(\int_{F^{-1}(p)}^\infty (F_n(x) - F(x)) dx \right) dp + o_{\mathbf{P}}(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n h(X_i) + o_{\mathbf{P}}(1), \end{aligned}$$

which completes the proof of Theorem 2.1.

9. NEGLIGIBILITY OF REMAINDER TERMS

The following four lemmas establish the earlier noted statements that the remainder terms $r_{n,1}, \dots, r_{n,4}$ are of the order $o_{\mathbf{P}}(1)$. In the proofs of the lemmas we shall use a parameter $\delta \in (0, 1/2]$, possibly different from line to line but never depending on n . Furthermore, we shall frequently use the fact that

$$\mathbf{E}[X^q] < \infty \implies \int_0^1 |F_n^{-1}(t) - F^{-1}(t)|^q dt = o_{\mathbf{P}}(1). \quad (9.1)$$

Another technical result that we shall frequently use is the fact that, for any $\varepsilon > 0$ as small as desired,

$$\sup_{x \in \mathbf{R}} \frac{\sqrt{n} |F_n(x) - F(x)|}{F(x)^{1/2-\varepsilon} (1 - F(x))^{1/2-\varepsilon}} = O_{\mathbf{P}}(1) \quad (9.2)$$

when $n \rightarrow \infty$.

Lemma 9.1. *Under the conditions of Theorem 2.1, we have that $r_{n,1} = o_{\mathbf{P}}(1)$.*

Proof. We split the remainder term $r_{n,1} = \sqrt{n} \int_0^1 \dots dp$ into the sum of $r_{n,1}^*(\delta) = \sqrt{n} \int_0^{1-\delta} \dots dp$ and $r_{n,1}^{**}(\delta) = \sqrt{n} \int_{1-\delta}^1 \dots dp$. The lemma follows if:

- (1) For every $\delta > 0$, the statement $r_{n,1}^*(\delta) = o_{\mathbf{P}}(1)$ holds when $n \rightarrow \infty$.
- (2) $r_{n,1}^{**}(\delta) = h(\delta)O_{\mathbf{P}}(1)$ for a deterministic $h(\delta) \downarrow 0$ when $\delta \downarrow 0$, where $O_{\mathbf{P}}(1)$ does not depend on δ .

To prove part (1), we first note that when $0 < p < 1-\delta$, then $AL_F^*(p) \geq \int_{1-\delta}^1 F^{-1}(t) dt$, which is positive, and $AL_n^*(p) \geq \int_{1-\delta}^1 F^{-1}(t) dt + o_{\mathbf{P}}(1)$ due to statement (9.1) with $q = 1$. Hence, we are left to show that, when $n \rightarrow \infty$,

$$\sqrt{n} \int_0^{1-\delta} \frac{1}{p} |AL_n(p) - AL_F(p)| |AL_n^*(p) - AL_F^*(p)| dp = o_{\mathbf{P}}(1). \quad (9.3)$$

Since $AL_n^*(p) - AL_F^*(p) = (\bar{X} - \mu_F) - (AL_n(p) - AL_F(p))$, statement (9.3) follows if

$$\sqrt{n} |\bar{X} - \mu_F| \int_0^{1-\delta} \frac{1}{p} |AL_n(p) - AL_F(p)| dp = o_{\mathbf{P}}(1) \quad (9.4)$$

and

$$\sqrt{n} \int_0^{1-\delta} \frac{1}{p} |AL_n(p) - AL_F(p)|^2 dp = o_{\mathbf{P}}(1). \quad (9.5)$$

We have $\sqrt{n} |\bar{X} - \mu_F| = O_{\mathbf{P}}(1)$ and $|AL_n(p) - AL_F(p)| \leq \sqrt{p} (\int_0^1 |F_n^{-1}(p) - F^{-1}(p)|^2 dp)^{1/2}$.

Since $\int_0^1 |F_n^{-1}(p) - F^{-1}(p)|^2 dp = o_{\mathbf{P}}(1)$ and $\int_0^{1-\delta} p^{-1} \sqrt{p} dp < \infty$, we have statement (9.4). To prove statement (9.5), we use bound (8.4) and reduce the proof to showing that

$$\frac{1}{\sqrt{n}} \int_0^{1-\delta} \frac{1}{p} \left| \int_0^{F^{-1}(p)} \sqrt{n} (F_n(x) - F(x)) dx \right|^2 dp = o_{\mathbf{P}}(1) \quad (9.6)$$

and

$$\frac{1}{\sqrt{n}} \int_0^{1-\delta} \frac{1}{p} |e_n(p)|^2 |F_n^{-1}(p) - F^{-1}(p)|^2 dp = o_{\mathbf{P}}(1). \quad (9.7)$$

To prove statement (9.6), we use statement (9.2) and observe that

$$\int_0^{1-\delta} \frac{1}{p} \left(\int_0^{F^{-1}(p)} F(x)^{1/2-\varepsilon} dx \right)^2 dp \leq c(F, \delta) \int_0^{1-\delta} \frac{1}{p} p^{1-2\varepsilon} dp < \infty. \quad (9.8)$$

To prove statement (9.7), we use the uniform on $[0, 1]$ version of statement (9.2) and Hölder's inequality, and in this way reduce the proof to showing that

$$\frac{1}{\sqrt{n}} \left(\int_0^{1-\delta} \frac{1}{p^{2\varepsilon a}} dp \right)^{1/a} \left(\int_0^{1-\delta} |F_n^{-1}(p) - F^{-1}(p)|^{2b} dp \right)^{1/b} = o_{\mathbf{P}}(1) \quad (9.9)$$

for some $a, b > 1$ such that $a^{-1} + b^{-1} = 1$. We choose a and b as follows. First, since $\mathbf{E}[X^{2+\alpha}] < \infty$, we set $b = (2 + \alpha)/2$. Next, we choose $\varepsilon > 0$ on the left-hand side of statement (9.9) so that $2\varepsilon a < 1$, which holds when $\varepsilon < \alpha/(4 + 2\alpha)$ in view of the equation $a^{-1} + b^{-1} = 1$. Hence, statement (9.9) holds and thus statement (9.7) follows.

This completes the proof of part (1).

To establish part (2), we first estimate $|r_{n,1}^{**}(\delta)|$ from above using the bounds $AL_F^*(p) \geq (1-p)F^{-1}(1/2)$ and $AL_n^*(p) \geq (1-p)F_n^{-1}(1/2)$, which hold since $\delta \leq 1/2$. Hence, we have reduced our task to showing that $\sqrt{n} \int_{1-\delta}^1 |AL_n(p) - AL_F(p)| dp = h(\delta) O_{\mathbf{P}}(1)$.

Using the Vervaat process, we reduce the latter statement to showing that the integrals

$$\int_{1-\delta}^1 \left(\int_0^{F^{-1}(p)} \sqrt{n} |F_n(x) - F(x)| dx \right) dp \quad (9.10)$$

and

$$\int_{1-\delta}^1 |e_n(p)| |F_n^{-1}(p) - F^{-1}(p)| dp \quad (9.11)$$

are of the order $h(\delta)O_{\mathbf{P}}(1)$ with possibly different $h(\delta) \downarrow 0$ in each case. In view of statement (9.2), we have the desired statement for integral (9.10) if the quantity

$$\int_{1-\delta}^1 \left(\int_0^{F^{-1}(p)} (1 - F(x))^{1/2-\varepsilon} dx \right) dp \quad (9.12)$$

converges to 0 when $\delta \downarrow 0$, in which case we set the quantity to be our $h(\delta)$. The inner integral of (9.12) does not exceed $\int_0^\infty (1 - F(x))^{1/2-\varepsilon} dx$, which is finite for all sufficiently small $\varepsilon > 0$ since $\mathbf{E}[X^{2+\alpha}] < \infty$ for some $\alpha > 0$. This completes the proof that quantity (9.10) is of the order $h(\delta)O_{\mathbf{P}}(1)$. To show that quantity (9.11) is of the same order, we use the uniform on $[0, 1]$ version of statement (9.2) and reduce the task to showing that $\int_{1-\delta}^1 |F_n^{-1}(p) - F^{-1}(p)| dp$ is of the desired order. By the Cauchy-Bunyakowski-Schwarz inequality, we have

$$\int_{1-\delta}^1 |F_n^{-1}(p) - F^{-1}(p)| dp \leq \sqrt{\delta} \left(\int_0^1 |F_n^{-1}(p) - F^{-1}(p)|^2 dp \right)^{1/2}.$$

Since $\mathbf{E}[X^2] < \infty$, we have $\int_0^1 |F_n^{-1}(p) - F^{-1}(p)|^2 dp = o_{\mathbf{P}}(1)$, and so setting $h(\delta) = \sqrt{\delta}$ establishes the desired asymptotic result for integral (9.11). This also completes the proof of part (2) and also of Lemma 9.1. \square

Lemma 9.2. *Under the conditions of Theorem 2.1, we have that $r_{n,2} = o_{\mathbf{P}}(1)$.*

Proof. Like in the proof of Lemma 9.1, we split the remainder term $r_{n,2} = \sqrt{n} \int_0^1 \dots dp$ into the sum of $r_{n,2}^*(\delta) = \sqrt{n} \int_0^{1-\delta} \dots dp$ and $r_{n,2}^{**}(\delta) = \sqrt{n} \int_{1-\delta}^1 \dots dp$. To prove the lemma, we need to show that:

- (1) For every $\delta > 0$, the statement $r_{n,2}^*(\delta) = o_{\mathbf{P}}(1)$ holds when $n \rightarrow \infty$.
- (2) $r_{n,2}^{**}(\delta) = h(\delta)O_{\mathbf{P}}(1)$ for a deterministic $h(\delta) \downarrow 0$ when $\delta \downarrow 0$, where $O_{\mathbf{P}}(1)$ does not depend on δ .

To prove part (1), we first estimate $|r_{n,2}^*(\delta)|$ from above using the bounds $p^{-1}AL_F(p) \leq F^{-1}(1 - \delta) < \infty$, $AL_F^*(p) \geq \int_{1-\delta}^1 F^{-1}(t) dt > 0$, and $AL_n^*(p) \geq \int_{1-\delta}^1 F^{-1}(t) dt + o_{\mathbf{P}}(1)$. This reduces our task to showing that, for every $\delta > 0$,

$$\sqrt{n} \int_0^{1-\delta} |AL_n^*(p) - AL_F^*(p)|^2 dp = o_{\mathbf{P}}(1). \quad (9.13)$$

Since $AL_n^*(p) - AL_F^*(p) = (\bar{X} - \mu_F) - (AL_n(p) - AL_F(p))$ and $\sqrt{n}(\bar{X} - \mu_F)^2 = o_{\mathbf{P}}(1)$, statement (9.13) follows from

$$\sqrt{n} \int_0^{1-\delta} |AL_n(p) - AL_F(p)|^2 dp = o_{\mathbf{P}}(1),$$

which is an elementary consequence of statement (9.5). This establishes part (1).

To prove part (2), we first estimate $|r_{n,2}^{**}(\delta)|$ from above using the bounds $AL_F^*(p) \geq (1-p)F^{-1}(1/2)$ and $AL_n^*(p) \geq (1-p)F_n^{-1}(1/2)$, and in this way reduce the task to showing that

$$\sqrt{n} \int_{1-\delta}^1 \frac{1}{1-p} |AL_n^*(p) - AL_F^*(p)| dp = h(\delta)O_{\mathbf{P}}(1). \quad (9.14)$$

Using the Vervaat process, the proof of statement (9.14) follows if

$$\int_{1-\delta}^1 \frac{1}{1-p} \left(\int_{F^{-1}(p)}^{\infty} \sqrt{n} |F_n(x) - F(x)| dx \right) dp = h(\delta)O_{\mathbf{P}}(1) \quad (9.15)$$

and

$$\int_{1-\delta}^1 \frac{1}{1-p} |e_n(p)| |F_n^{-1}(p) - F^{-1}(p)| dp = h(\delta)O_{\mathbf{P}}(1) \quad (9.16)$$

with possibly different $h(\delta) \downarrow 0$ in each case. Using statement (9.2), we have that statement (9.15) holds with $h(\delta)$ set as the integral

$$\int_{1-\delta}^1 \frac{1}{1-p} \left(\int_{F^{-1}(p)}^{\infty} (1-F(x))^{1/2-\varepsilon} dx \right) dp, \quad (9.17)$$

which converges to 0 when $\delta \downarrow 0$ as the following argument shows. First, we write the integrand as the product of $(1-F(x))^\varepsilon$ and $(1-F(x))^{1/2-2\varepsilon}$. Then we estimate the first factor by $(1-p)^\varepsilon$. The integral $\int_0^\infty (1-F(x))^{1/2-2\varepsilon} dx$ is finite for all sufficiently small $\varepsilon > 0$ since $\mathbf{E}[X^{2+\alpha}] < \infty$ for some $\alpha > 0$. Since $\int_{1-\delta}^1 (1-p)^{-1+\varepsilon} dp \downarrow 0$ when $\delta \downarrow 0$, integral (9.17) converges to 0 when $\delta \downarrow 0$. The proof of statement (9.15) is finished.

We are left to prove statement (9.16). Using the uniform on $[0, 1]$ version of statement (9.2), we reduce the task to showing that

$$\int_{1-\delta}^1 \frac{1}{(1-p)^{1/2+\varepsilon}} |F_n^{-1}(p) - F^{-1}(p)| dp = h(\delta)O_{\mathbf{P}}(1). \quad (9.18)$$

(In fact, we shall see below that $O_{\mathbf{P}}(1)$ can be replaced by $o_{\mathbf{P}}(1)$.) Using Hölder's inequality, we have that the right-hand side of equation (9.18) does not exceed

$$\left(\int_{1-\delta}^1 \frac{1}{(1-p)^{(1/2+\varepsilon)a} } dp \right)^{1/a} \left(\int_{1-\delta}^1 |F_n^{-1}(p) - F^{-1}(p)|^b dp \right)^{1/b} \quad (9.19)$$

for some $a, b > 1$ such that $a^{-1} + b^{-1} = 1$, which we choose as follows. Since $\mathbf{E}[X^{2+\alpha}] < \infty$, we set $b = 2 + \alpha$, and so the right-most integral of (9.19) is of the order $o_{\mathbf{P}}(1)$. Furthermore, $a = (2 + \alpha)/(1 + \alpha) < 2$, which can be made arbitrarily close to 2 by choosing sufficiently small $\alpha > 0$. Choosing $\varepsilon > 0$ so small that $(1/2 + \varepsilon)a < 1$, we have that the left-most integral in (9.19) converges to 0 when $\delta \downarrow 0$ and, as a consequence,

we set the integral to be our function $h(\delta)$. This establishes statement (9.16) and completes the proof of Lemma 9.2. \square

Lemma 9.3. *Under the conditions of Theorem 2.1, we have that $r_{n,3} = o_{\mathbf{P}}(1)$.*

Proof. We split the remainder term $r_{n,3} = \int_0^1 \dots dp$ into the sum of $r_{n,3}^* = \int_0^{1/2} \dots dp$ and $r_{n,3}^{**} = \int_{1/2}^1 \dots dp$. The lemma follows if the two summands are of the order $o_{\mathbf{P}}(1)$.

To prove $r_{n,3}^* = o_{\mathbf{P}}(1)$, we use the bound $AL_F^*(p) \geq \int_{1/2}^1 F^{-1}(p) dp$ and the uniform on $[0, 1]$ version of statement (9.2), and in this way reduce our task to showing that

$$\int_0^{1/2} \frac{1}{p^{1/2+\varepsilon}} |F_n^{-1}(p) - F^{-1}(p)| dp = o_{\mathbf{P}}(1).$$

This statement can be established following the proof of statement (9.18), with minor modifications.

To prove $r_{n,3}^{**} = o_{\mathbf{P}}(1)$, we use the bound $AL_F^*(p) \geq (1-p)F^{-1}(1/2)$, the fact that $\sup_t |e_n(t)| = O_{\mathbf{P}}(1)$, and statement (9.1) with $q = 1$. The desired result for $r_{n,3}^{**}$ follows, which finishes the proof of Lemma 9.3. \square

Lemma 9.4. *Under the conditions of Theorem 2.1, we have that $r_{n,4} = o_{\mathbf{P}}(1)$.*

Proof. We split $r_{n,4} = \int_0^1 \dots dp$ into the sum of $r_{n,4}^* = \int_0^{1/2} \dots dp$ and $r_{n,4}^{**} = \int_{1/2}^1 \dots dp$, and then show that the two summands are of the order $o_{\mathbf{P}}(1)$.

To prove $r_{n,4}^* = o_{\mathbf{P}}(1)$, we use the bounds $p^{-1}AL_F(p) \leq F^{-1}(1/2) < \infty$ and $AL_F^*(p) \geq \int_{1/2}^1 F^{-1}(p) dp > 0$ together with the uniform on $[0, 1]$ version of statement (9.2). This reduces our task to showing that $\int_0^{1/2} |F_n^{-1}(p) - F^{-1}(p)| dp = o_{\mathbf{P}}(1)$, which holds due to statement (9.1) with $q = 1$.

To prove $r_{n,4}^{**} = o_{\mathbf{P}}(1)$, we use the bound $AL_F^*(p) \geq (1-p)F^{-1}(1/2)$ and the uniform on $[0, 1]$ version of statement (9.2), and in this way reduce the proof to showing that

$$\int_{1/2}^1 \frac{1}{(1-p)^{1/2+\varepsilon}} |F_n^{-1}(p) - F^{-1}(p)| dp = o_{\mathbf{P}}(1).$$

This statement can be established following the proof of statement (9.18). The proof of Lemma 9.4 is finished. \square

REFERENCES

BANK OF ITALY (2006). *Household Income and Wealth in 2004. Supplements to the Statistical Bulletin-Sample Surveys*, XVI-n.7.

- DAVISON A.C., HINKLEY D.V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge.
- EFRON B. (1987). Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association*, 82, 171–200.
- GASTWIRTH, J.L. (1971). A general definition of the Lorenz curve. *Econometrica*, 39, 1037–1039.
- GINI C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. In: *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*. Anno Accademico 1913–1914, Tomo LXXII parte seconda. Premiate Officine Grafiche C. Ferrari, Venezia, 1201–1248.
- GRESELIN, F. AND PASQUAZZI, L. (2009). Asymptotic confidence intervals for a new inequality measure. *Communications in Statistics: Computation and Simulation*, 38(8), 17–42.
- GRESELIN, F., PURI, M.L. AND ZITIKIS, R. (2009). *L*-functions, processes, and statistics in measuring economic inequality and actuarial risks. *Statistics and Its Interface*, 2, 227–245.
- KARIAN, Z.A. AND DUDEWICZ, E.J. (2000). *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*. CRC Press, Boca Raton, FL.
- KLEIBER, C. AND KOTZ, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, New York.
- LEHMANN, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, 37, 1137–1153.
- MAASOUMI E. (1994). Empirical analysis of welfare and inequality. In: *Handbook of Applied Econometrics, Volume II: Microeconomics*. (Eds.: M.H. Pesaran and P. Schmidt). Blackwell, Oxford.
- PIETRA G. (1915). Delle relazioni fra indici di variabilità, note I e II. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, 74, 775–804.
- ZENGA, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni* 5, 3–27.
- ZITIKIS, R. (1998). The Vervaat process. In: *Asymptotic Methods in Probability and Statistics*, pp. 667–694. North-Holland, Amsterdam.