ATTI DELLA XLV RIUNIONE SCIENTIFICA (2010)

XLV Riunione Scientifica
(Padova, 16-18 giugno 2010)
ISBN 978-88-6129-566-7
Editore: CLEUP – Padova

Francesca Greselin, Leo Pasquazzi

Dagum confidence intervals for inequality measures

Dagum confidence intervals for inequality measures and an application to Swiss income data

Francesca Greselin, Leo Pasquazzi

Abstract In this work we compare parametric Dagum confidence intervals with non parametric confidence intervals for two inequality measures, Gini's (1914) traditional index and Zenga's (2007) new measure. We point out some problems in the computation of parametric Dagum confidence intervals and present a simulation study to assess what is gained with respect to non parametric confidence intervals when we exploit the asymptotic efficiency of ML estimators. Finally, we employ the results to analyze data coming from the *Income and Consumption Survey* of Switzerland.

Key words: Gini index, Zenga index, Lognormal model, Dagum model, confidence interval, measuring poverty and inequality.

1 Introduction

Since more than a century economists and statisticians have been concerned with the problem of modeling income and wealth distributions and measuring inequality. Despite the fact that data are nowadays available in very large samples, it was noted by many authors that point estimates of inequality measures are not quite reliable and that comparisons should be based on confidence intervals. In this work we will compare parametric Dagum confidence intervals with non parametric confidence intervals. The choice of the Dagum model is due to its well known ability to fit economic size distributions. We firstly present the results of a simulation study with the aim to determine coverage accuracy and length of parametric Dagum and non parametric confidence intervals for two inequality measures, Zenga's new (Zenga, 2007) and Gini's traditional index (Gini, 1914). Then

Francesca Greselin, Leo Pasquazzi
Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali, Universitá di Milano Bicocca, Milan, Italy, e-mail: francesca.greselin@unimib.it,leo.pasquazzi@unimib.it

The rest of the paper is organized as follows. The next section provides a brief description of technical details for the computation of the confidence intervals and presents the simulation results. In section 3 we introduce and discuss our inferential results on inequality in Switzerland Major Regions. Conclusions and final remarks end the paper in section 4.

2 Asymptotic confidence intervals

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from an unknown distribution F. Gini's index may be defined by

$$G(F) = \int_0^1 2(p - L(p;F))dp, \tag{1}$$

where

$$L(p;F) = \frac{\int_0^p F^{-1}(t)dt}{\int_0^1 F^{-1}(t)dt}, \quad 0 < p < 1 \tag{2}$$

is the Lorenz curve, while Zenga's new measure is given by

$$Z(F) = \int_0^1 \left(1 - \frac{1-p}{p} \cdot \frac{L(p;F)}{1 - L(p;F)}\right) dp. \tag{3}$$

As usual, we assume that the support of F is a subset of the non negative real line. Moreover, in order that the two inequality measures be well defined we need to assume that the first moment of F is finite.

If $\widehat{F_n}$ is the empirical CDF associated to the observed sample, we may estimate the two inequality measures simply by plugging in $\widehat{F_n}$ instead of F in (1), (2) and (3). Under mild restrictions on F (Hoeffding, 1948; Greselin et al., 2009) both inequality measures may be represented as

$$T(\widehat{F_n}) = T(F) + \frac{1}{n}\sum_{i=1}^{n} h_T(X_i;F) + o_p(n^{-1/2}) \tag{4}$$

where the $h_T(X_i;F)$ is the influence function evaluated at the point X_i, i.e. (as usual δ_X denotes the distribution with a unit mass at the point X)

$$h_T(X_i;F) = \lim_{\lambda \downarrow 0} \frac{T(F + \lambda(\delta_{X_i} - F)) - T(F)}{\lambda}.$$

It follows that both inequality measures have normal asymptotic distribution, i.e.

$$\sqrt{n}\left(T(\widehat{F_n}) - T(F)\right) \xrightarrow{d} N(0, \sigma_T^2),$$

where $\sigma_T^2 = Var_F(h_T(X_i;F))$. Let $S_{T;n}^2$ be a consistent estimator for σ_T^2; hence, we may compute the non parametric normal $(1-2\alpha)$ confidence interval given by

$$\left(T(\widehat{F}_n) - z_{1-\alpha}\frac{S_{T;n}}{\sqrt{n}}; T(\widehat{F}_n) + z_{1-\alpha}\frac{S_{T;n}}{\sqrt{n}} \right),$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-percentile of the standard normal distribution.

If F is known to belong to a parametric family $\mathscr{F}_\Theta$ indexed by a k-dimensional real parameter vector $\theta \in \Theta \subset \mathbb{R}^k$, then the two functionals in (1) and (3) are functions of θ and we will simply write $T(\theta)$ instead of $T(F_\theta)$. In this case we may estimate $T(\theta)$ by $T(\widehat{\theta}_n)$, where $\widehat{\theta}_n$ is the maximum likelihood estimate of the unknown value of θ. If T and $\mathscr{F}_\Theta$ satisfy suitable regularity conditions, then this estimator is asymptotically normal and efficient, i.e.

$$\sqrt{n}\left(T(\widehat{\theta}_n) - T(\theta) \right) \xrightarrow{d} N(0, \sigma_T^2(\theta)), \tag{5}$$

where $\sigma_T^2(\theta) = \frac{\partial T}{\partial \theta'} \mathbf{I}_\theta^{-1} \frac{\partial T}{\partial \theta}$. In the variance expression $\frac{\partial T}{\partial \theta}$ and $\mathbf{I}_\theta$ indicate the (column) vector of partial derivatives of T with respect to the components of the parameter vector and the information matrix at the unknown value of θ, respectively. If $\sigma_T^2(\theta)$ is continuous in θ, then $\sigma_T^2(\widehat{\theta}_n)$ is a consistent estimator of $\sigma_T^2(\theta)$.

Besides the normal confidence intervals just described, we will also consider different types of bootstrap confidence intervals, i.e. percentile, Bias Corrected Accelerated Bootstrap (Bca) and t-bootstrap confidence intervals (Davison and Hinkley, 1997).

For the non parametric versions of these confidence intervals, we proceed as in Greselin and Pasquazzi (2009) and estimate the bootstrap distributions by taking $R = 9999$ resamples from the original sample (i.e. from $\widehat{F}_n$). As variance estimator for σ_T^2 we use

$$S_{T;n}^2 = \frac{1}{n}\sum_{i=1}^n h_T(X_i; \widehat{F}_n)^2,$$

and, following Efron (1987), we estimate the acceleration constant for the Bca confidence intervals by

$$\widehat{a} = \frac{1}{6}\frac{\sum_{i=1}^n h_T(X_i; \widehat{F}_n)^3}{\left(\sum_{i=1}^n h_T(X_i; \widehat{F}_n)^2\right)^{3/2}}. \tag{6}$$

Heuristically, we may say that in the parametric versions of the confidence intervals nothing changes with respect to the non parametric setting, except that $F_{\widehat{\theta}_n}$ plays the role of $\widehat{F}_n$. Indeed, the expansion corresponding to (4) in the non parametric setting, may now be replaced by

$$T(\widehat{\theta}_n) = T(\theta) + \frac{1}{n}\sum_{i=1}^n \frac{\partial \ln f_\theta(X_i)}{\partial \theta'} \mathbf{I}_\theta^{-1} \frac{\partial T}{\partial \theta} + o_p(n^{-1/2}), \tag{7}$$

where $\frac{\partial \ln f_\theta(X_i)}{\partial \theta}$ is the score vector (a column vector) of the i-th sample component X_i at the unknown true value of θ. Thus, we may use

$$h_T(X_i;\widehat{\theta}_n) = \left.\frac{\partial \ln f_\theta(X_i)}{\partial \theta'}\right|_{\theta=\widehat{\theta}_n} \mathbf{I}^{-1}_{\widehat{\theta}_n} \left.\frac{\partial T}{\partial \theta}\right|_{\theta=\widehat{\theta}_n} \tag{8}$$

instead of $h_T(X_i;\widehat{F}_n)$ for estimating the variance $\sigma_T^2(\theta)$, which results in

$$\begin{aligned} V^2_{T;n} &= \frac{1}{n}\sum_{i=1}^{n} h_T(X_i;F_{\widehat{\theta}_n})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} \left.\frac{\partial T}{\partial \theta'}\right|_{\theta=\widehat{\theta}_n} \mathbf{I}^{-1}_{\widehat{\theta}_n} \left.\frac{\partial \ln f_\theta(X_i)}{\partial \theta}\right|_{\theta=\widehat{\theta}_n} \left.\frac{\partial \ln f_\theta(X_i)}{\partial \theta'}\right|_{\theta=\widehat{\theta}_n} \mathbf{I}^{-1}_{\widehat{\theta}_n} \left.\frac{\partial T}{\partial \theta}\right|_{\theta=\widehat{\theta}_n} . \end{aligned} \tag{9}$$

In the same way we just substitute $h_T(X_i;\widehat{F}_n)$ by $h_T(X_i;\widehat{\theta}_n)$ in (6) to get the estimate of the acceleration constant for the parametric Bca confidence interval.

We shall now briefly discuss technical details for the computation of parametric Dagum confidence intervals. First, we recall that F belongs to the Dagum family if its density function is given by

$$f(x) = \frac{apx^{ap-1}}{b^{ap}\left[1+\left(\frac{x}{b}\right)^a\right]^{p+1}}, \quad x > 0$$

for some $a,b,p > 0$ (Kleiber and Kotz, 2003). Notice that the first moment of a Dagum distribution is finite if and only if $a > 1$, and therefore the inequality measures we consider in this paper are only defined for the subfamily of Dagum distributions with $a > 1$.

In our simulation study we used the maximum likelihood estimates from the Italian equivalent income distribution as parameter values for the parent distribution. Thus we simulated $10,000$ samples from the Dagum distribution with $a = 3.6781$, $b = 19,262$ and $p = 0.6875$ in order to estimate coverage accuracy and size of parametric and non parametric confidence intervals for Gini's and Zenga's new index.

Given an i.i.d. sample $x_1,x_2,...,x_n$, the likelihood equations for the Dagum family are given by

$$\begin{aligned} \frac{n}{a} + p\sum_{i=1}^{n}\ln\left(\frac{x_i}{b}\right) - (p+1)\sum_{i=1}^{n}\frac{\ln\left(\frac{x_i}{b}\right)}{1+\left(\frac{b}{x_i}\right)^a} &= 0 \\ np - (p+1)\sum_{i=1}^{n}\frac{1}{1+\left(\frac{b}{x_i}\right)^a} &= 0 \\ \frac{n}{p} + a\sum_{i=1}^{n}\ln\left(\frac{x_i}{b}\right) - \sum_{i=1}^{n}\ln\left[1+\left(\frac{x_i}{b}\right)^a\right] &= 0 \end{aligned} \tag{10}$$

However, no closed form solution of this system is known. The ML estimation problem is easier to handle if we observe that the natural logarithm of a Dagum random variable follows a generalized logistic distribution with density function given by

$$f(y) = \frac{\alpha}{\sigma} \frac{e^{-\frac{x-\theta}{\sigma}}}{\left(1+e^{-\frac{x-\theta}{\sigma}}\right)^{\alpha+1}}, \quad -\infty < y < \infty,$$

where $-\infty < \theta < \infty$ and $\alpha, \sigma > 0$. Notice that θ and σ are the location and scale parameter, respectively, while α is a shape parameter that affects asymmetry. The parameters of the generalized logistic distribution are linked to those of the Dagum distribution by the relations

$$a = \frac{1}{\sigma}, \quad b = e^{\theta}, \quad p = \alpha.$$

Thus, the problem of solving the system in (10) is equivalent to the problem of finding a solution of the likelihood equations of the generalized logisitic distribution, which are given by

$$\begin{aligned} \frac{n}{\alpha} - \sum_{i=1}^{n} \ln\left(1+e^{-\frac{y_i-\theta}{\sigma}}\right) &= 0 \\ -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} \frac{y_i-\theta}{\sigma} - \frac{\alpha+1}{\sigma}\sum_{i=1}^{n} \frac{\frac{y_i-\theta}{\sigma}}{1+e^{\frac{y_i-\theta}{\sigma}}} &= 0 \\ \frac{n}{\sigma} - \frac{\alpha+1}{\sigma}\sum_{i=1}^{n} \frac{1}{1+e^{\frac{y_i-\theta}{\sigma}}} &= 0 \end{aligned} \tag{11}$$

We employed an iterative two-step procedure for solving this system. At step i we first find an update θ_i and σ_i of the location and scale parameters through a single Newton-Raphson step applied to the last two equations of the system (11). Then we use θ_i and σ_i in the first equation in order to get an update α_i of the shape parameter. If the likelihood function has a local maximum, this two-step procedure will converge to it provided that the starting values α_0, θ_0, σ_0 are not too far from the solution. The initial values for this algorithm are thus of crucial importance. We select them by least squares fitting the three quartiles of the generalized logistic distribution with shape parameter $\alpha = 1$ to the corresponding quartiles of the natural logarithm of the sample observations. Since the quantile function of the generalized logistic distribution, given by

$$y(t) = \theta + \sigma \ln\left(\frac{t^{\alpha}}{1-t^{\alpha}}\right), \qquad 0 < t < 1,$$

is linear in θ and σ, we find a closed form solution for the initial values. Indeed, putting $\alpha_0 = 1$ and denoting by Q_1, Q_2 and Q_3 the quartiles of the natural logarithm of the equivalent incomes x_i, we see that the initial values θ_0 and σ_0 are given by the least squares solution of the linear system

$$Q_1 = \theta - \sigma \ln 3$$
$$Q_2 = \theta$$
$$Q_3 = \theta + \sigma \ln 3,$$

which yields $\theta_0 = (Q_1 + Q_2 + Q_3)/3$ and $\sigma_0 = (-Q_1 \ln 3 + Q_3 \ln 3)/(2\ln^2 3)$.

We will allow for each sample a maximum number of $1,000$ iterations of the two-step procedure above. If the algorithm reduces the gradient of the likelihood function to a value smaller than 10^{-6} within $1,000$ iterations, we test the hessian matrix for negative definiteness at the solution. If this test is positive we conclude that the solution is a local maximum of the likelihood equation. Notice that beyond a bad choice of the initial values, there may be another simple reason why our procedure does not deliver a local maximum. Indeed, as Shao (2002) points out, there exist points in the sample space such that a solution of the likelihood equations in (11), and therefore also of the system in (10), does not exist. Nevertheless, with probability tending to 1 as the sample size increases, there exists a sequence of solutions of the likelihood equations of the generalized logistic distribution that is consistent and asymptotically normally distributed (Abberger and Heiler, 2000).

So how do we handle samples on which the algorithm does not deliver a local maximum? And what happens if the algorithm finds a local maximum, but the inequality measures are not defined at that local maximum (i.e. the ML estimate of the parameter a is not larger than 1)? Our answer to these questions depends on whether we are dealing with a bootstrap resample or not. In the former case we simply discard the sample and take another bootstrap resample until we reach a total of $9,999$ bootstrap resamples on which the algorithm converges to a local maximum at which the inequality measures are defined. Otherwise, if the sample we are dealing with is one of the *original* samples from the Dagum parent distribution of the simulation study, we use it for estimating the probability of the subset of the sample space on which the ML estimates of the inequality measures do not exist. For the Dagum parent distribution in our simulation study corresponding to the sample sizes $n = 100$, 200 and 400 these estimates are given by 0.0712, 0.0141 and 0.0010, respectively.

Let us now turn to the expressions of the inequality measures in the Dagum model. The Lorenz curve and Gini's index are respectively given by (Dagum, 1977)

$$L(t; a, b, p) = B\left(t^{1/p}; p + \frac{1}{a}, 1 - \frac{1}{a}\right), \quad 0 < t < 1 \tag{12}$$

and

$$G(p, a, b) = \frac{\Gamma(p)\Gamma(2p + 1/a)}{\Gamma(2p)\Gamma(p + 1/a)} - 1. \tag{13}$$

In (12) we used $B(t; a, b)$ to indicate the beta cdf, while $\Gamma(x)$ indicates the Gamma function in (13). Substituting the Lorenz curve (12) in the formula for Zenga's index (3), we get

$$Z(p, a, b) = \int_0^1 \frac{t - B\left(t^{1/p}; p + \frac{1}{a}, 1 - \frac{1}{a}\right)}{t\left[1 - B\left(t^{1/p}; p + \frac{1}{a}, 1 - \frac{1}{a}\right)\right]} dt. \tag{14}$$

As noticed at the beginning of this section, the Lorenz curve, and thus the two inequality measures, are defined if and only if $a > 1$.

In order to get the influence values $h_T(X_i, \theta)$, we need the gradient vectors of the two inequality functionals with respect to the parameters. The simplest way to solve this problem is to approximate the gradient by Newton's difference quotient. Finally, the components of the information matrix, also needed to get the influence values, can be found, for example, in Kleiber and Kotz (2003).

Table 1. *Simulation results: coverages and mean sizes.*

Dagum parent distribution										
$1-2\alpha$	0.9	0.95	0.975	0.99	mean size	0.9	0.95	0.975	0.99	mean size
sample size	**Normal confidence intervals**					**Percentile confidence intervals**				
	Gini - non parametric									
100	0.8469	0.9088	0.9430	0.9691	0.0977	0.8759	0.9346	0.9633	0.9819	0.0968
200	0.8666	0.9229	0.9551	0.9752	0.0727	0.8802	0.9357	0.9644	0.9841	0.0722
400	0.8835	0.9384	0.9653	0.9832	0.0533	0.8897	0.9442	0.9721	0.9876	0.0532
	Gini - parametric									
100	0.7839	0.8439	0.8810	0.9042	0.0993	0.8092	0.8631	0.8915	0.9106	0.0994
200	0.8620	0.9175	0.9485	0.9678	0.0723	0.8730	0.9254	0.9539	0.9698	0.0723
400	0.8932	0.9459	0.9718	0.9862	0.0520	0.8938	0.9464	0.9726	0.9883	0.0520
	Zenga - non parametric									
100	0.8486	0.9143	0.9506	0.9748	0.1218	0.8757	0.9336	0.9643	0.9833	0.1210
200	0.8644	0.9240	0.9596	0.9818	0.0890	0.8815	0.9359	0.9672	0.9870	0.0886
400	0.8781	0.9375	0.9676	0.9857	0.0647	0.8899	0.9456	0.9719	0.9883	0.0646
	Zenga - parametric									
100	0.8081	0.8645	0.8933	0.9140	0.1234	0.8263	0.8745	0.9029	0.9181	0.1234
200	0.8786	0.9276	0.9561	0.9727	0.0880	0.8841	0.9324	0.9587	0.9747	0.0881
400	0.8962	0.9499	0.9737	0.9891	0.0625	0.8983	0.9509	0.9742	0.9895	0.0625
	Bca confidence intervals					**t-bootstrap confidence intervals**				
	Gini - non parametric									
100	0.8279	0.8836	0.9213	0.9504	0.1016	0.8530	0.9135	0.9483	0.9725	0.1196
200	0.8514	0.9063	0.9401	0.9661	0.0758	0.8644	0.9206	0.9534	0.9778	0.0837
400	0.8757	0.9309	0.9581	0.9778	0.0553	0.8783	0.9339	0.9657	0.9833	0.0586
	Gini - parametric									
100	0.8182	0.8672	0.8939	0.9103	0.1045	0.8120	0.8678	0.8944	0.9141	0.1046
200	0.8806	0.9319	0.9584	0.9737	0.0742	0.8751	0.9290	0.9557	0.9714	0.0740
400	0.9004	0.9503	0.9735	0.9885	0.0526	0.8963	0.9497	0.9742	0.9889	0.0523
	Zenga - non parametric									
100	0.8363	0.8939	0.9275	0.9564	0.1213	0.8603	0.9215	0.9539	0.9770	0.1359
200	0.8495	0.9064	0.9425	0.9680	0.0895	0.8678	0.9242	0.9562	0.9816	0.0973
400	0.8744	0.9263	0.9567	0.9773	0.0654	0.8793	0.9352	0.9640	0.9843	0.0692
	Zenga - parametric									
100	0.8158	0.8677	0.8942	0.9116	0.1240	0.8251	0.8735	0.9014	0.9173	0.1252
200	0.8781	0.9304	0.9574	0.9744	0.0883	0.8831	0.9322	0.9589	0.9753	0.0886
400	0.8967	0.9505	0.9746	0.9874	0.0626	0.8994	0.9506	0.9750	0.9898	0.0626

Table 1 contains the main results of our simulations. The coverages reported in the table are the fraction of samples (on a total of $10,000$ samples) that give rise to a confidence interval that contains the true value of the inequality measure. Besides the coverages, the table reports also the mean sizes of the confidence intervals of each type.

We immediately notice the low coverages of the parametric confidence intervals for $n = 100$. This is due to the fact that our algorithm was not able to find a local maximum of the likelihood function in 712 of the $10,000$ samples of size $n = 100$. For $n = 200$ and $n = 400$ the problem of non existence of the ML estimates becomes less severe, and the parametric confidence intervals perform better. Indeed, for sample size $n = 400$ the parametric confidence intervals are both shorter and have larger coverages than the non parametric ones.

3 An Analysis of income data in Switzerland Major Regions

This section aims at presenting how our inferential result can be exploited to analyze data coming from the *Income and Consumption Survey* of Switzerland. Cross regional levels of income inequality will be measured by the Gini and Zenga indexes, their confidence intervals, both in the non parametric and in the parametric setting. The Major regions in Switzerland are Lake Geneva Region, Espace Mittelland, Northwestern Switzerland, Zurich, Eastern Switzerland, Central Switzerland and Ticino.

4 Conclusions

In this work we compared non parametric with parametric Dagum confidence intervals for Gini's and Zenga's inequality measures in a simulation study. We used the Dagum model both as parent distribution in the simulations and for the computation of the parametric confidence intervals. For small sample sizes the problem of non existence of the ML estimates spoils the performance of the parametric Dagum confidence intervals. For large samples, as the problem of non existence of local maxima of the likelihood function becomes negligible, the parametric confidence intervals are both shorter and have larger coverages than the non parametric ones.

Acknowledgements This research has been partially supported by the Milano Bicocca FAR 2009. The authors wish to thank the *Swiss Federal Statistical Office* for kindly supplying them microdata from the *Income and Consumption Survey*.

	non-parametric		parametric Dagum	
Lake Geneva Region (Sample size $n = 502$)				
AD	-		4.5411	
p-value	-		0.0470	(10016)
	Gini	Zenga	Gini	Zenga
point est.	0.270879	0.593947	0.262476	0.582933
normal	0.2383÷0.3052	0.5538÷0.6390	0.2425÷0.2826	0.5562÷0.6107
perc	0.2398÷0.3056	0.5498÷0.6339	0.2423÷0.2825	0.5545÷0.6091
Bca	0.2459÷0.3181	0.5585÷0.6465	0.2438÷0.2845	0.5564÷0.6115
t-boot	0.2436÷0.3355	0.5567÷0.6686	0.2444÷0.2848	0.5572÷0.6121
Espace Mittelland (Sample size $n = 751$)				
AD	-		176.6472	
p-value	-		0	(9999)
	Gini	Zenga	Gini	Zenga
point est.	0.25365	0.570771	0.257791	0.578956
normal	0.2394÷0.2685	0.5497÷0.5934	0.2423÷0.2736	0.5573÷0.6014
perc	0.2390÷0.2682	0.5479÷0.5916	0.2423÷0.2736	0.5564÷0.6005
Bca	0.2400÷0.2694	0.5502÷0.5940	0.2428÷0.2745	0.5578÷0.6022
t-boot	0.2399÷0.2695	0.5500÷0.5945	0.2432÷0.2746	0.5576÷0.6019
Northwestern Switzerland (Sample size $n = 400$)				
AD	-		2.7296	
p-value	-		0.1636	(10339)
	Gini	Zenga	Gini	Zenga
point est.	0.254257	0.568909	0.256258	0.571469
normal	0.2336÷0.2762	0.5396÷0.6012	0.2346÷0.2791	0.5418÷0.6034
perc	0.2325÷0.2752	0.5360÷0.5979	0.2340÷0.2790	0.5392÷0.6014
Bca	0.2351÷0.2786	0.5408÷0.6036	0.2360÷0.2816	0.5426÷0.6050
t-boot	0.2351÷0.2800	0.5403÷0.6044	0.2360÷0.2814	0.5423÷0.6046
Zurich (Sample size $n = 531$)				
AD	-		3.403	
p-value	-		0.0828	(10009)
	Gini	Zenga	Gini	Zenga
point est.	0.26549	0.587588	0.26608	0.588763
normal	0.2464÷0.2856	0.5615÷0.6159	0.2465÷0.2862	0.5627÷0.6163
perc	0.2457÷0.2848	0.5589÷0.6131	0.2464÷0.2863	0.5612÷0.6149
Bca	0.2479÷0.2872	0.5626÷0.6163	0.2478÷0.2877	0.5631÷0.6172
t-boot	0.2473÷0.2874	0.5619÷0.6177	0.2477÷0.2878	0.5632÷0.6169
Eastern Switzerland (Sample size $n = 417$)				
AD	-		20.6457	
p-value	-		0.0011	(9999)
	Gini	Zenga	Gini	Zenga
point est.	0.248861	0.567391	0.251134	0.57159
normal	0.2303÷0.2688	0.5396÷0.5980	0.2315÷0.2719	0.5433÷0.6023
perc	0.2293÷0.2679	0.5366÷0.5954	0.2310÷0.2709	0.5405÷0.5995
Bca	0.2316÷0.2709	0.5407÷0.6000	0.2328÷0.2734	0.5443÷0.6033
t-boot	0.2310÷0.2704	0.5399÷0.5997	0.2331÷0.2732	0.5440÷0.6032
Central Switzerland (Sample size $n = 245$)				
AD	-		2.2903	
p-value	-		0.1997	(10276)
	Gini	Zenga	Gini	Zenga
point est.	0.278121	0.603434	0.269792	0.592761
normal	0.2374÷0.3223	0.5542÷0.6602	0.2408÷0.3005	0.5546÷0.6344
perc	0.2368÷0.3214	0.5456÷0.6519	0.2397÷0.2997	0.5507÷0.6302
Bca	0.2450÷0.3358	0.5584÷0.6642	0.2436÷0.3043	0.5557÷0.6354
t-boot	0.2417÷0.3568	0.5552÷0.6890	0.2438÷0.3047	0.5564÷0.6364
Ticino (Sample size $n = 241$)				
AD	-		1.5055	
p-value	-		0.634	(10183)
	Gini	Zenga	Gini	Zenga
point est.	0.252847	0.567987	0.250516	0.565618
normal	0.2191÷0.2893	0.5223÷0.6205	0.2241÷0.2784	0.5285÷0.6060
perc	0.2199÷0.2894	0.5164÷0.6140	0.2237÷0.2783	0.5252÷0.6032
Bca	0.2272÷0.3048	0.5287÷0.6315	0.2268÷0.2825	0.5296÷0.6081
t-boot	0.2244÷0.3121	0.5249÷0.6457	0.2264÷0.2814	0.5290÷0.6073

Table 1 *Cross-regional levels of income inequality expressed by the Gini and the Zenga indexes, followed by their 95% confidence intervals, in non-parametric (columns 1-2) and parametric setting (columns 3-4), for Swiss Major regions.*

References

ABBERGER K. AND HEILER S.(2000). Simultaneous estimation of parameters for a generalized distribution and application to time series models. *All. Stat. Arch.*, 84, 41–49.

DAGUM C. (1977). A new model of personal distribution: specification and estimation. *Economie Appliquée*, 30, 413–437.

DAVISON A.C. AND 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.

GINI C.(1914). Sulla misura della concentrazione e della variabilit dei caratteri. *Atti del Reale Istituto Veneto di scienze, lettere ed arti 1914*, 73, part 2.

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., PASQUAZZI, L. AND ZITIKIS, R. (2009). Zenga's new index of economic inequality, its estimation, and an analysis of incomes in Italy. *MPRA Paper 17147* available at http://ideas.repec.org/p/pra/mprapa/17147.html

HOEFFDING W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathemtical Statistics*, 19 (3), 293–325.

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.

KLEIBER, C. AND KOTZ, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, New York.

SHAO J. AND TU D. (1995). *The Jackknife and Bootstrap*. Springer Series in Statistics.

ZENGA, M.(2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni*, 5, 3–27.