Asymptotic Confidence Intervals for a New Inequality Measure

Francesca Greselin, Leo Pasquazzi

Asymptotic Confidence Intervals for a New Inequality Measure

Journal:	*Communications in Statistics - Simulation and Computation*
Manuscript ID:	LSSP-2008-0070.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	06-Mar-2009
Complete List of Authors:	Greselin, Francesca; University Milan Bicocca, Quantitative methods for economics and business sciences Pasquazzi, Leo; Milano-Bicocca University, Quantitative methods for economics and business sciences
Keywords:	inequality curve, Zenga's index, asymptotic confidence intervals
Abstract:	Recently Zenga (2007) introduced a new inequality measure based on ratios between lower and upper group means. Both Zenga's new measure and Gini's traditional index may be interpreted in terms of areas beneath inequality curves. In this work we compare the performance of some types of asymptotic confidence intervals for the new measure and for Gini's index.
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
cis-review.tex	

ASYMPTOTIC CONFIDENCE INTERVALS FOR A NEW INEQUALITY MEASURE

1 Introduction

Recently Zenga (2007) introduced a new inequality measure based on ratios between lower and upper group means. **Both Zenga's new measure and Gini's index may be interpreted in terms of areas beneath inequality curves.** In this **work** the performance of asymptotic confidence intervals for Gini's measure and for the new measure is tested. Several types of confidence intervals are considered: the normal, the percentile, the BCa and the t-bootstrap. While the underlying asymptotic theory for Gini's measure is well established, formal proofs for Zenga's index are **currently under development**. Indeed, also in view of our simulation results, asymptotic properties similar to those of Gini's index can be expected to hold also for Zenga's new inequality measure.

The present work is organized as follows. Section 2 introduces the inequality functionals. Section 3 presents a brief review on non parametric confidence intervals. The simulation study is presented and discussed in Section 4. Some concluding remarks end the paper in Section 5.

2 The Inequality Functionals

Let $X_1, ..., X_n$ be independent random variables with common distribution function F. Denote by $\widehat{F}$ the empirical cumulative distribution function, and let $\theta(F)$ be the functional of interest. Using this notation, a sample statistic may be written as $\theta(\widehat{F})$.

Under suitable smoothness restrictions on θ (Serfling, 1980) the sample statistic may be represented as

$$\theta(\widehat{F}) = \theta(F) + \frac{1}{n}\sum_{i=1}^{n} h(F; X_i) + o_p(n^{-1/2}), \tag{1}$$

where

$$h(F;x) = \lim_{\lambda \to 0^+} \frac{\theta(F + \lambda(\delta_x - F)) - \theta(F)}{\lambda} \tag{2}$$

is the value of the influence function at x (δ_x as usual denotes the distribution function which assigns probability 1 to the point x). If $\sigma_\theta^2 = Var(h(F;X))$ is finite, then the Taylor expansion in (1) implies that

$$\left(\frac{\theta(\widehat{F}) - \theta(F)}{\sigma_\theta / \sqrt{n}} \right) \xrightarrow{d} N(0,1), \tag{3}$$

Section 3 reviews some asymptotic confidence intervals for smooth functionals like θ to which the above *non parametric delta method* applies.

Let us now focus on the two inequality measures considered in this work. Using the functional notation, Gini's measure is defined by

$$G(F) = 2 \int_0^1 (p - L_F(p))\, dp,$$

where

$$L_F(p) = \frac{1}{\mu_F} \int_0^p F^{-1}(t) dt, \quad 0 < p < 1$$

is the Lorenz curve associated with the distribution F and μ_F is the first moment of F. As usual, we restrict the domain of Gini's index to the class of distribution functions F with non negative support. Further, we impose $0 < \mu_F < \infty$ so that $L_F(p)$ is well defined. The asymptotic normality of Gini's inequality measure in case of i.i.d. sampling was first established by Hoeffding (1948), under second order moment restrictions on F.

For continuous distributions F with non negative support and finite first moment the new inequality measure of Zenga (2007) is defined by

$$Z(F) = \int_0^1 Z_F(p) dp \tag{4}$$

where

$$Z_F(p) = 1 - \frac{1-p}{p} \cdot \frac{L_F(p)}{1 - L_F(p)}, \quad 0 < p < 1$$

is Zenga's new inequality curve.

The ratio behind this definition is straightforward. Indeed, $\frac{L_F(p)}{p}$ **in** $Z_F(p)$ **is the mean income of the poorest** p **percent of the population, while** $\frac{1-L_F(p)}{1-p}$ **is the mean income of the remaining part of the population. Notice that the ratio between this two means takes on values in** $[0,1]$**. Small values correspond to high inequality between the two groups, while large values correspond to situations close to equity. Thus** $Z_F(p)$ **is a point inequality measure and** $Z(F)$ **is the mean of these point measures.**

In order to compare the two inequality measures it is convenient to rewrite also Gini's index as a (weighted) mean of point inequality measures, i.e.

$$G(F) = \int_0^1 G_F(p)\ 2p\ dp \tag{5}$$

where, according to Gini (1914),

$$G_F(p) = \frac{p - L_F(p)}{p}, \quad 0 < p < 1$$

measures the relative inequality at p**. As** $0 \leq L_F(p) \leq p$**, we see that also** $G_F(p)$ **takes on values in** $[0,1]$ **and that** $G_F(p)$ **is large (small) if the fraction of total income of the poorest** p **percent of the population is small (large). Comparing now (4) with (5) we observe that both** $G(F)$ **and** $Z(F)$ **are means of their respective inequality curves, but while Gini's index emphasizes the upper part of the income distribution by the weight function** $w(p) = 2p$**, Zenga's index has a constant weight function.**

Substituting $p = F(x)$ in all integrals involved in definition (4), $Z(F)$ may be expressed as

$$Z(F) = 1 - \int_0^\infty \frac{1 - F(x)}{F(x)} \cdot \frac{\int_0^x y dF(y)}{\int_x^\infty y dF(y)}\ dF(x). \tag{6}$$

The latter analytic expression for the functional proves to be more convenient for computational purposes in inferential problems. Indeed, the empirical distribution function $\widehat{F}$

associated to an i.i.d. sample is discrete, so that substituting F with $\widehat{F}$ in (6) yields a sum of n ratios. If instead we substitute F with $\widehat{F}$ in (4), we have to evaluate the area below a continuous curve, which is more time consuming. For this reason, we will use definition (6) in our simulation study. **For discrete distribution functions like $\widehat{F}$ this requires slight changes in the defining formula:**

$$Z(F) = 1 - \int_{[0,\infty)} g_F(x)\ dF(x), \tag{7}$$

where

$$g_F(x) = \begin{cases} \frac{1-F(x)}{F(x)} \cdot \frac{\int_{[0,x]} y dF(y)}{\int_{(x,\infty)} y dF(y)}, & \textbf{if } 0 < F(x) < 1; \\ 0, & \textbf{otherwise.} \end{cases}$$

Notice that $g_F(x)$ is well defined for each x and each distribution F with non negative support. With this definition we get

$$Z(\widehat{F}) = 1 - \frac{1}{n} \sum_{i=1}^{n-1} \left(\frac{n-i}{i} \cdot \frac{\sum_{j \le i} x_j}{\sum_{j > i} x_j} \right).$$

Neither (4) nor (7) coincide with the original definition that the author provided for *discrete* distributions. This will not cause any troubles here, since the scope is merely to compute confidence intervals for the functional (4) (as originally defined) in case of i.i.d. sampling from a *continuous* distribution F and **hence we need a *smooth extension* of the functional to some convex space that contains the parent distribution F (a continuous distribution with non negative support and finite first moment) and all empirical distribution functions $\widehat{F}$ that may occur when we sample from F.** By smooth extension we mean an extension such that the Taylor expansion in (1) and (2) holds. **In the appendix we shall compute $h(F;X)$ for Zenga's new measure and show that $\sigma_\theta^2 = Var(h(F;X))$ is finite if $\int x^{2+\epsilon} dF(x)$ is finite for some $\epsilon > 0$. Condition (1) will however not be checked in this work.** In what follows we shall confine ourselves to a simulation study for assessing the coverage probability and the length of some types of asymptotic confidence intervals for Zenga's new measure and for Gini's index.

3 Confidence Intervals

This section reviews some of the most commonly used types of confidence intervals and the theory behind them.

Firstly, notice that if (3) holds and we substitute σ_θ^2 with a consistent estimator[1] $\widehat{\sigma}_\theta^2$, we obtain the $(1-2\alpha)$ *normal* confidence interval

$$\left(\theta(\widehat{F}) - z_{1-\alpha}\frac{\widehat{\sigma}_\theta}{\sqrt{n}}; \theta(\widehat{F}) + z_{1-\alpha}\frac{\widehat{\sigma}_\theta}{\sqrt{n}}\right)$$

for $\theta(F)$. In what follows we will briefly discuss confidence intervals for θ obtained by non parametric bootstrap methods.

3.1 The Percentile Confidence Interval

The *percentile* method relies on the assumption that there exists a strictly increasing continuous transformation $g(\theta)$ such that

$$P\left\{g(\theta(\widehat{F})) - g(\theta(F)) \leq x\right\} = \Psi(x) \qquad \textit{for all } F, \tag{8}$$

where Ψ is a continuous and symmetric distribution. If Ψ is the standard normal distribution then g is just the normalizing and variance stabilizing transformation. Let z_α be the α quantile of Ψ. If g and Ψ were known, we could compute an exact $1-2\alpha$ confidence interval for $\theta(F)$ in the following way:

$$\left(g^{-1}(g(\theta(\widehat{F})) - z_{1-\alpha}); g^{-1}(g(\theta(\widehat{F})) - z_\alpha)\right). \tag{9}$$

We will now **show** that this confidence interval can also be computed *without* knowing neither g nor Ψ. This will be done by exploiting condition (8) and the bootstrap distribution of θ, i.e. the distribution of the functional θ if the underlying parent distribution is $\widehat{F}$. In what follows we will denote by P_* probabilities which refer to i.i.d. sampling from $\widehat{F}$ and F^* will be used for the empirical cumulative distribution function induced by n independent random

[1] Several consistent estimators have been proposed in literature, provided the functional satisfies quite general smoothness conditions. See Shao and Tu (1996) for an overview about sufficient conditions.

variables with common distribution $\widehat{F}$. Using this notation, the bootstrap distribution of θ is given by $P_*\{\theta(F^*) < x\}$. Now, if θ_p is the p-th quantile of the latter distribution, then

$$\begin{aligned} P_*\{\theta(F^*) \le \theta_p\} &= P_*\left\{g(\theta(F^*)) - g(\theta(\widehat{F})) \le g(\theta_p) - g(\theta(\widehat{F}))\right\} \\ [\text{by (8)}] &= \Psi\left(g(\theta_p) - g(\theta(\widehat{F}))\right) \qquad (10) \\ [\text{by continuity of } \Psi \textbf{ and } g] &= p \end{aligned}$$

so that the θ_p satisfies the condition $g(\theta_p) - g(\theta(\widehat{F})) = z_p$. Substituting this in (9) and using the symmetry of Ψ, we get the *bootstrap percentile confidence interval* whose bounds are given simply by θ_α and $\theta_{1-\alpha}$. Notice that these bounds can be computed without knowing neither g nor Ψ.

If assumption (8) holds exactly for the parent distribution F and for all $\widehat{F}$ that can occur with probability 1,[2] then the percentile method presented here gives exact confidence intervals. In practice, however, assumption (8) holds only asymptotically as the sample size tends to infinity. In this case the coverage accuracy of the percentile confidence interval depends on how good the approximation is at F and at the empirical distribution functions $\widehat{F}$.

3.2 The Boostrap Accelerated Bias Corrected Confidence Interval

The *bootstrap accelerated bias corrected confidence interval* relies on a more general assumption than (8), that takes into account also bias and skewness. In particular, it is assumed that there exist a strictly increasing continuous transformation g as well as constants a and b such that

$$P\left\{\frac{g(\theta(\widehat{F})) - g(\theta(F))}{1 + a g(\theta(F))} + b \le x\right\} = \Psi(x) \qquad \textit{for all } F, \qquad (11)$$

where Ψ is assumed to be continuous and strictly increasing. According to Efron (1987) the parameter a measures how fast the standard deviation of $g(\theta(\widehat{F}))$ is changing with respect to $g(\theta(F))$ and is therefore called the *acceleration constant*. If g, Ψ, a and b were known, the

[2] notice that if $F = \widehat{F}$ in (8), then $\widehat{F} = F^*$ and $P = P_*$ on the left hand side

lower and upper bounds of an exact $(1-2\alpha)$ confidence interval for θ could be computed in the following way:

$$g^{-1}\left(g(\theta(\widehat{F})) + \frac{(b-z_{1-\alpha})(1+ag(\theta(\widehat{F})))}{1-a(b-z_{1-\alpha})}\right) \tag{12}$$

and

$$g^{-1}\left(g(\theta(\widehat{F})) + \frac{(b-z_{\alpha})(1+ag(\theta(\widehat{F})))}{1-a(b-z_{\alpha})}\right). \tag{13}$$

Now, using assumption (11) we get

$$\begin{aligned} P_*\{\theta(F^*) \le \theta_p\} &= P_*\left\{\frac{g(\theta(F^*)) - g(\theta(\widehat{F}))}{1+ag(\theta(\widehat{F}))} + b \le \frac{g(\theta_p) - g(\theta(\widehat{F}))}{1+ag(\theta(\widehat{F}))} + b\right\} \\ &[\text{by (11)}] = \Psi\left(\frac{g(\theta_p) - g(\theta(\widehat{F}))}{1+ag(\theta(\widehat{F}))} + b\right) \\ &[\text{by continuity of } \Psi \text{ and } g] = p \end{aligned}$$

so that the p-th quantile θ_p of the bootstrap distribution satisfies the following condition:

$$\frac{g(\theta_p) - g(\theta(\widehat{F}))}{1+ag(\theta(\widehat{F}))} + b = z_p.$$

Solving for θ_p we get

$$\theta_p = g^{-1}\left(g(\theta(\widehat{F})) + (z_p - b)(1+ag(\theta(\widehat{F})))\right).$$

Then, equating θ_p with the lower confidence bound in (12), and solving for p we see that

$$\begin{aligned} g(\theta(\widehat{F})) + (z_p-b)(1+ag(\theta(\widehat{F}))) &= g(\theta(\widehat{F})) + \frac{(b-z_{1-\alpha})(1+ag(\theta(\widehat{F})))}{1-a(b-z_{1-\alpha})} \\ (z_p-b) &= \frac{(b-z_{1-\alpha})}{1-a(b-z_{1-\alpha})} \\ p &= \Psi\left(b + \frac{(b-z_{1-\alpha})}{1-a(b-z_{1-\alpha})}\right). \end{aligned}$$

Hence, the lower bound of the exact confidence interval in (12) may be obtained as a quantile of the bootstrap distribution of θ and can therefore be computed *without knowing the increasing transformation* g. The same proof may be applied to the upper bound. Thus the

BCa confidence interval is given by $(\theta_{\underline{p}}; \theta_{\overline{p}})$, where θ_p is the p-th quantile of the bootstrap distribution of θ and

$$\underline{p} = \Psi\left(b + \frac{(b - z_{1-\alpha})}{1 - a(b - z_{1-\alpha})}\right),$$
$$\overline{p} = \Psi\left(b + \frac{(b - z_{\alpha})}{1 - a(b - z_{\alpha})}\right).$$

In order to find b notice that (use again assumption (11))

$$P_*\left\{\theta(F^*) \leq \theta(\widehat{F})\right\} = P_*\left\{\frac{g(\theta(F^*)) - g(\theta(\widehat{F}))}{1 + ag(\theta(\widehat{F}))} + b \leq b\right\} = \Psi(b)$$

so that $b = \Psi^{-1}\left(P_*\left\{\theta(F^*) \leq \theta(\widehat{F})\right\}\right)$. The acceleration constant a, however, is not easy to determine. **As suggested by Efron (1987), in our simulation study we approximated the constant a by**

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

where h is the influence function defined in (2). Since $\sum_{i=1}^n h(\widehat{F}; X_i) = 0$ for smooth functionals, $\widehat{a}$ may be simply interpreted as one sixth of the skewness of the *empirical influence values* $h(\widehat{F}; X_i)$. The latter were obtained by numeric differentiation, i.e.

$$h(\widehat{F}; X_i) \approx \frac{\theta\left(\widehat{F} + \epsilon(\delta_x - \widehat{F})\right) - \theta\left(\widehat{F}\right)}{\epsilon},$$

with $\epsilon = 0.0001/n$.

Again, assumption (11) holds in practice only asymptotically as the sample size diverges and the coverage accuracy of the BCa confidence interval depends on how good the approximation is at the parent distribution F and at the empirical distribution functions $\widehat{F}$.

3.3 The t-bootstrap Confidence Interval

The idea behind the *t-bootstrap confidence interval* is easy to understand. Assume there exists a pivotal quantity, or at least an asymptotically pivotal quantity, like

$$\frac{\theta(\widehat{F}) - \theta(F)}{\sigma_\theta(\widehat{F})/\sqrt{n}}, \tag{14}$$

where $\sigma_\theta^2(\widehat{F})$ is a consistent estimator of σ_θ^2 in (3). Given the empirical distribution $\widehat{F}$, the distribution of the random variable in (14) may be estimated by

$$P_* \left\{ \frac{\theta(F^*) - \theta(\widehat{F})}{\sigma_\theta(F^*)/\sqrt{n}} \leq t \right\}. \tag{15}$$

Using this estimate we may construct the $(1-2\alpha)$ t-bootstrap confidence interval

$$\left(\theta(\widehat{F}) - t^*_{1-\alpha} \frac{\sigma_\theta(\widehat{F})}{\sqrt{n}}; \theta(\widehat{F}) - t^*_{\alpha} \frac{\sigma_\theta(\widehat{F})}{\sqrt{n}} \right),$$

where t^*_p is the p-th quantile of the bootstrap distribution in (15). The drawback of the t-bootstrap confidence interval is that it requires a consistent variance estimator $\sigma_\theta^2(\widehat{F})$ to be known. In the simulation study described in the following section we used

$$\widehat{\sigma}_\theta^2 = \frac{1}{n} \sum_{i=1}^{n} h(\widehat{F}; X_i)^2,$$

as variance estimator in the normal and t-bootstrap confidence intervals. Notice that $\widehat{\sigma}_\theta^2$ is the variance of the empirical influence values $h(\widehat{F}; X_i)$. Intuitively, $\widehat{\sigma}_\theta^2$ will be consistent for σ_θ^2 if $h(\widehat{F}; x)$ is close to $h(F; x)$ uniformly in x whenever $\widehat{F}$ is close to F.

Usually the analytic expressions of the bootstrap distributions of θ or of its studentized version are unknown and must be approximated by Monte Carlo methods. This introduces a simulation error into the confidence intervals. Most authors compute from 999 to 9999 replicates of θ or its studentized version to estimate the quantiles of the corresponding bootstrap distribution. The ratio behind computing, say, 999 instead of 1000 replicates is that in this way the quantiles of common interest (i.e the percentiles) are order statistics of the bootstrap replicates. The computational effort, in particular for BCa and t-bootstrap confidence intervals, is justified by a higher order of accuracy with respect to normal confidence intervals.

4 Simulation Results

The design of the simulation study begins with the choice of the underlying distributions. As income analysis is the natural framework in which inequality measures are needed, we will focus our attention on two widely used income models: the Dagum and the Lognormal models. Their probability density functions are respectively given by

$$f(x) = \lambda\beta\theta x^{-(\theta+1)}\left(1+\lambda x^{-\theta}\right)^{-(\beta+1)}, \qquad (\lambda,\beta,\theta>0,\ x>0);$$

$$f(x) = \frac{1}{\sqrt{2\pi}\delta}\frac{1}{x}e^{-\frac{1}{2}\left(\frac{\ln x-\gamma}{\delta}\right)^2}, \qquad (-\infty<\gamma<\infty,\delta>0,\ x>0).$$

In the simulations we set the parameters equal to their maximum likelihood estimates obtained on real data. In the Dagum case, the Italian expenditure distribution, as given by the Banca d'Italia Survey on Household Income and Wealth in 2002 (8001 households) yields $\widehat{\beta}=1.055$, $\widehat{\theta}=3.095$ and $\widehat{\lambda}=44030$ (Bianchi, 2003). On this distribution the value of Gini's and Zenga's inequality measures are respectively given by $\widehat{G}=0.3193$ and $\widehat{Z}=0.6505$. In the Lognormal case, Latorre (1989) obtained $\widehat{\gamma}=2.8171$ and $\widehat{\delta}=0.6262$, using the Italian income distribution of 1983 (4107 households; source: Banca d'Italia Survey on Italian Household Income and Wealth, 1983). Hence, we get $\widehat{G}=0.3420$ and $\widehat{Z}=0.6774$. Both estimated models proved to fit well the corresponding empirical distribution, as assessed by the Kolmogorov-Smirnov test and the χ^2 test.

Having chosen the parent distributions, we drew 10000 samples of size ranging from 100 to 400 from both of them and for each sample we checked whether the confidence intervals in section 3 contained the true value of the inequality measure. The coverage probabilities reported in Table 1 are the relative frequencies (on a total of 10000) of the confidence intervals that contain the true value of the inequality measure, while Table 2 displays the average size of the 10000 confidence intervals with 95% nominal confidence level. The bootstrap distributions were obtained by taking 9999 resamples from each sample and by computing the functionals or their standardized versions (for the t-bootstrap confidence intervals) on each resample.

The results in Table 1 show that the confidence intervals suffer from undercoverage. Comparing the different types of confidence intervals it is seen, as expected, that the t-bootstrap performs best in terms of coverage probability, at the cost of providing larger intervals (see Table 1 and 2). As the sample size reaches 400, the coverage probability of the t-bootstrap confidence intervals is quite close to the nominal confidence level.

If large confidence levels are required, the number of bootstrap resamples has to be increased beyond 9999 in order to get meaningful BCa confidence intervals. Indeed, in the Dagum case with samples of size 400, we observed that 825 (out of 10000) BCa confidence intervals for Gini's index with 99% nominal confidence level were based on the 100th percentile of the bootstrap distribution (for Zenga's index this figure is 659). This may be explained by the heaviness of the tail of the Dagum model. Indeed, with the Lognormal parent distribution, only 34 of the 99%-Bca confidence intervals for Gini's index and 29 of those for Zenga's index required the 100th percentile of the bootstrap distribution for their computation.

Consider now the results in Table 2. Combined with the above observed undercoverage, the rather large average sizes of the confidence intervals suggest that point estimates for the two inequality measures are not very accurate even in samples of 400 observations. In practical situations, in order to distinguish between substantial differences in the degree of inequality and differences that are merely due to sample variability, we may therefore need larger samples.

Finally, comparing the performance of the confidence intervals for Gini's index with that of Zenga's new measure we do not notice large differences in the coverage accuracy or in the size. It seems that both indices may be estimated with about the same precision in finite samples from economic size distributions.

5 Concluding Remarks

In this work we tested the performance of several types of asymptotic confidence intervals for Zenga's new inequality measure and for Gini's traditional measure in sampling from economic size distributions. The results show that the confidence intervals suffer from some undercoverage. Nevertheless, it appears that the coverage accuracy of the t-bootstrap confidence intervals is quite close to the nominal confidence level in samples of size 400. **The rather large average size of the confidence intervals, however, suggests that point estimates for the two inequality measures are not very accurate in samples of size up to 400 and that larger samples may be needed in order to get a reliable inequality ranking. Besides this, it appears that the confidence intervals for the Zenga and Gini indices have similar coverage accuracy and size.**

Appendix

We will first derive an analytic expression for the influence function $h(F;x)$ of Zenga's index as defined in equation (7), i.e. we shall find

$$\lim_{\lambda\to 0^+}\frac{Z\left(F_\lambda\right)-Z\left(F\right)}{\lambda}, \tag{16}$$

where F is a continuous distribution with non negative support and finite first moment μ, X is a fixed real number such that $0<F(X)<1$ and $F_\lambda = F+\lambda(\delta_X - F)$. We shall use the following notation:

$$\mu_\lambda = \int_{[0,\infty)} y dF_\lambda(y), \quad Q_\lambda(x) = \int_{[0,x]} y dF_\lambda(y)$$

and

$$g_\lambda(x) = \begin{cases} \frac{1-F_\lambda(x)}{F_\lambda(x)}\frac{Q_\lambda(x)}{\mu_\lambda - Q_\lambda(x)}, & \text{if } 0<F_\lambda(x)<1; \\ 0, & \text{otherwise;} \end{cases}$$

so that

$$Z(F_\lambda) = 1 - \int_{[0,\infty)} g_\lambda(x) dF_\lambda(x).$$

If $\lambda = 0$ we shall simply write μ, $Q(x)$ and $g(x)$.

For each $\lambda \in [0,1]$, we have

$$Z\left(F_\lambda\right) - Z\left(F\right) = -\int_{[0,\infty)} \left[g_\lambda(x) - g(x)\right] dF_\lambda(x) - \int_{[0,\infty)} g(x) d\left(F_\lambda(x) - F(x)\right)$$

Using the definition of F_λ, the first integral on the right hand side becomes

$$(1-\lambda)\int_{[0,\infty)} \left[g_\lambda(x) - g(x)\right] dF(x) + \lambda\left[g_\lambda(X) - g(X)\right],$$

while the second intergral can be written as

$$\lambda\left[g(X) - \int_{[0,\infty)} g(x) dF(x)\right]$$

It follows that

$$\begin{aligned}\frac{Z\left(F_\lambda\right) - Z\left(F\right)}{\lambda} = &-\int_{[0,\infty)} \frac{g_\lambda(x) - g(x)}{\lambda} dF(x) + \\ &+\int_{[0,\infty)} \left[g_\lambda(x) - g(x)\right] dF(x) + \\ &- \left[g_\lambda(X) - g(X)\right] + \\ &- \left[g(X) - \int_{[0,\infty)} g(x) dF(x)\right].\end{aligned} \tag{17}$$

Since $g_\lambda(x) \to g(x)$ for all x as λ approaches 0 from the right, the expression in the third line of equation (17) tends to 0 when we take the limit in (16). Moreover $|g_\lambda(x)| \leq 1$ for all x and for all $\lambda \in [0,1]$ and therefore also the integral in the second line of equation (17) will tend to 0 as λ approaches 0 from the right. Consider now the integral in the first line of (17). Suppose that $0 < \lambda < 1$. Then $0 < F(x) < 1$ is equivalent to $0 < F_\lambda(x) < 1$. If these conditions fail, then

$$\frac{g_\lambda(x) - g(x)}{\lambda} = 0.$$

On the other hand, if at x the condition $0 < F(x) < 1$ holds, we get (after some simple algebraic manipulations)

$$\begin{aligned}\frac{g_\lambda(x)-g(x)}{\lambda} =& \frac{1-F_\lambda(x)}{(\mu_\lambda - Q_\lambda(x))\,F_\lambda(x)} \cdot \frac{Q_\lambda(x)-Q(x)}{\lambda} + \\ &+ \frac{Q(x)}{(\mu_\lambda - Q_\lambda(x))\,F_\lambda(x)} \cdot \frac{(1-F_\lambda(x))-(1-F(x))}{\lambda} + \\ &- \frac{Q(x)\,(1-F(x))}{(\mu_\lambda - Q_\lambda(x))\,F_\lambda(x)} \cdot \frac{1}{F(x)} \cdot \frac{F_\lambda(x)-F(x)}{\lambda} + \\ &- \frac{Q(x)\,(1-F(x))}{(\mu_\lambda - Q_\lambda(x))\,F(x)} \cdot \frac{1}{\mu - Q(x)} \cdot \frac{(\mu_\lambda - Q_\lambda(x))-(\mu - Q(x))}{\lambda}.\end{aligned}$$

Simplifying λ in the rightmost fractions of each line, we obtain

$$\begin{aligned}\frac{g_\lambda(x)-g(x)}{\lambda} =& \frac{1-F_\lambda(x)}{(\mu_\lambda - Q_\lambda(x))\,F_\lambda(x)} \left(X\delta_X(x) - Q(x)\right) + \\ &+ \frac{Q(x)}{(\mu_\lambda - Q_\lambda(x))\,F_\lambda(x)} \left[(1-\delta_X(x)) - (1-F(x))\right] + \\ &- \frac{Q(x)\,(1-F(x))}{(\mu_\lambda - Q_\lambda(x))\,F_\lambda(x)F(x)} \left(\delta_X(x) - F(x)\right) + \\ &- \frac{Q(x)\,(1-F(x))}{(\mu_\lambda - Q_\lambda(x))\,F(x)\,(\mu - Q(x))} \left[X\,(1-\delta_X(x)) - (\mu - Q(x))\right].\end{aligned} \tag{18}$$

The first term on the right hand side is bounded by

$$\left|\frac{X\delta_X(x)-Q(x)}{Q_\lambda(x)}\right| \le \begin{cases} \frac{1}{1-\lambda} & \text{if } x < X \\ \frac{1}{1-\lambda}\left(\frac{X}{Q(X)}+1\right) & \text{if } x \ge X. \end{cases}$$

Therefore, if $\lambda < \lambda_0 \in [0,1)$, there there exists a constant K (depending on F, X and λ_0) that bounds the first term on the right hand side of the equation (18) for all x. In a similar way we can also find constants that bound the other three terms on the right hand side of equation (18) for all λ smaller than some $\lambda_0 \in [0,1)$ and for all x. It follows that when we let λ approach zero from the right in the first integral on the right hand side of (17), we can apply Lebesgue's dominated convergence theorem and take the limit under the integral sign.

Finally we get the value of the influence function at the point X

$$\begin{aligned} h(F,X) &= \lim_{\lambda\to 0^+} \frac{Z\left(F_\lambda\right) - Z\left(F\right)}{\lambda} \\ &= -\int_{[0,\infty)} \frac{1-F(x)}{\left(\mu - Q(x)\right)F(x)} \left(X\delta_X(x) - Q(x)\right) dF(x) + \\ &\quad -\int_{[0,\infty)} \frac{Q(x)}{\left(\mu - Q(x)\right)F(x)} \left[\left(1-\delta_X(x)\right) - \left(1-F(x)\right)\right] dF(x) + \\ &\quad +\int_{[0,\infty)} \frac{Q(x)\left(1-F(x)\right)}{\left(\mu - Q(x)\right)F(x)^2} \left(\delta_X(x) - F(x)\right) dF(x) + \\ &\quad +\int_{[0,\infty)} \frac{Q(x)\left(1-F(x)\right)}{\left(\mu - Q(x)\right)^2 F(x)} \left[X\left(1-\delta_X(x)\right) - \left(\mu - Q(x)\right)\right] dF(x) \\ &\quad -\left[g(X) - \int_{[0,\infty)} g(x) dF(x)\right]. \end{aligned} \tag{19}$$

Notice that, since each of the integrals in (19) is finite, we could split up the integral $\int \lim_{\lambda\to 0^+} \frac{g_\lambda(x)-g(x)}{\lambda} dF(x)$.

Let now X be redefined as a random variable with distribution given by F and consider $E\{h(F;X)\}$. Applying Fubini's theorem to the first four terms on the right hand side of (19) we see that the expectation $E\{h(F;X)\}$ vanishes. We shall now find a restriction on the moments of F that ensures

$$Var\{h(F;X)\} = E\{h(F;X)^2\} < \infty.$$

Recalling that $|g_\lambda(x)| \leq 1$ and ignoring constant terms in (19), we see that the variance of $h(F;x)$ is finite if and only if the second moments of the following random variables

$$X \cdot \int_{[X,\infty)} \frac{1-F(x)}{\left(\mu - Q(x)\right)F(x)} dF(x) \tag{20}$$

$$\int_{[0,X)} \frac{Q(x)}{\left(\mu - Q(x)\right)F(x)} dF(x) \tag{21}$$

$$\int_{[X,\infty)} \frac{Q(x)\left(1-F(x)\right)}{\left(\mu - Q(x)\right)F(x)^2} dF(x) \tag{22}$$

$$X \cdot \int_{[0,X)} \frac{Q(x)\left(1-F(x)\right)}{\left(\mu - Q(x)\right)^2 F(x)} dF(x) \tag{23}$$

are finite. Consider first the random variables in (20), (21) and (22). Notice that they are bounded from above respectively by

$$\begin{aligned} X \cdot \int_{[X,\infty)} \frac{1}{XF(x)} dF(x) &= -\ln F(X), \\ \int_{[0,X)} \frac{1}{1-F(X)} dF(x) &= -\ln\left(1-F(X)\right), \\ \int_{[X,\infty)} \frac{1}{F(x)} dF(x) &= -\ln F(X). \end{aligned}$$

Since we assumed F to be continuous, we have $0 < F(X) < 1$ with probability 1, and therefore the random variables $\ln F(X)$ and $\ln(1-F(X))$ are well defined with probability 1. Since

$$P\left\{-\ln^k F(X) > t\right\} = P\left\{F(X) < e^{-t^{1/k}}\right\} = e^{-t^{1/k}}, \quad k = 1, 2, ...,$$

and

$$P\left\{-\ln^k (1-F(X)) > t\right\} = P\left\{F(X) > 1 - e^{-t^{1/k}}\right\} = e^{-t^{1/k}}, \quad k = 1, 2, ..., \tag{24}$$

all moments of the random variables $-\ln F(X)$ and $-\ln\left(1-F(X)\right)$ are finite. It follows that no additional restriction on the parent distribution F is needed in order that the second moments of the random variables in (20), (21) and (22) be finite. Now, our last effort is devoted to the random variable in (23)

$$\begin{aligned} X \cdot \int_{[0,X)} \frac{Q(x)\left(1-F(x)\right)}{\left(\mu - Q(x)\right)^2 F(x)} dF(x) &\leq X \cdot \int_{[0,X)} \frac{1}{\mu\left(1-F(x)\right)} dF(x) \\ &= -\frac{X}{\mu} \ln\left(1-F(X)\right). \end{aligned}$$

Let $\epsilon > 0$. If $0 < F(x) < 1$ then either

$$x^{\epsilon} \geq -\ln(1-F(x))$$

or

$$x < -\ln^{1/\epsilon}(1-F(x)).$$

Therefore,

$$\begin{aligned} -\frac{X}{\mu} \ln\left(1-F(X)\right) &\leq \frac{1}{\mu} \max\left\{X^{1+\epsilon}; -\ln^{1+1/\epsilon}(1-F(X))\right\} \\ &\leq \frac{1}{\mu}\left(X^{1+\epsilon} - \ln^{1+1/\epsilon}(1-F(X))\right). \end{aligned}$$

By (24) and the Schwartz inequality, the random variable $\frac{1}{\mu}\left(X^{1+\epsilon} - \ln^{1+1/\epsilon}(1-F(X))\right)$ has finite second moment provided that $\int x^{2+2\epsilon}dF(x) < \infty$. The same condition ensures that the random variable in (23) and hence $h(F;X)$ have finite variance.

BIBLIOGRAPHY

Bianchi E. (2003). La distribuzione di Dagum: applicazione ai consumi in Italia. it Tesi di laurea della Facolt di Economia dell'Universit Milano Bicocca, a.a. 2003-2004.

Davison A.C., Hinkley D.V. (1997). *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics.

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. 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.

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

Latorre G. (1989). Asymptotic distributions of Indices of Concentration: empirical verification and applications. *Income and Wealth Distribution, Inequality and Poverty, Camilo Dagum, Michele Zenga (Eds)*. Springer-Verlag, 149–170.

Shao,J., Tu D. (1995). *The Jackknife and the Bootstrap*. Springer-Verlag, New York.

Zenga M. (2007). Inequality Curve and Inequality Index based on the Ratios between Lower and Upper Arithmetic Means. *Statistica & Applicazioni*, **V**, 1, 3–27.

Table 1: Coverage accuracy of the asymptotic confidence intervals

Dagum parent distribution: $\lambda = 44030$, $\beta = 1.055$, $\theta = 3.095$; $G = 0.3193$, $I = 0.6505$								
	Gini's index				***Zenga's index***			
	***normal* confidence intervals**							
n	***0.9000***	***0.9500***	***0.9750***	***0.9900***	***0.9000***	***0.9500***	***0.9750***	***0.9900***
100	0.8228	0.8842	0.9196	0.9468	0.8407	0.9032	0.9418	0.9662
200	0.8418	0.9005	0.9331	0.9592	0.8522	0.9102	0.9476	0.9715
400	0.8643	0.9194	0.9482	0.9689	0.8641	0.9236	0.9560	0.9758
	***percentile* confidence intervals**							
100	0.7986	0.8571	0.8933	0.9269	0.8157	0.8743	0.9115	0.9410
200	0.8271	0.8880	0.9183	0.9489	0.8313	0.8915	0.9235	0.9554
400	0.8539	0.9108	0.9421	0.9637	0.8569	0.9092	0.9417	0.9670
	***BCa* confidence intervals**							
100	0.8334	0.8959	0.9347	0.9593	0.8472	0.9075	0.9449	0.9694
200	0.8487	0.9020	0.9416	0.9670	0.8523	0.9098	0.9463	0.9728
400	0.8602	0.9181	0.9496	0.9733	0.8633	0.9212	0.9529	0.9762
	***t-boostrap* confidence intervals**							
100	0.8642	0.9226	0.9524	0.9734	0.8647	0.9276	0.9576	0.9778
200	0.8699	0.9232	0.9575	0.9767	0.8696	0.9268	0.9613	0.9807
400	0.8780	0.9338	0.9608	0.9810	0.8769	0.9344	0.9634	0.9829
Lognormal parent distribution: $\gamma = 2.8171$, $\delta = 0.6262$; $G = 0.3420$, $Z = 0.6774$								
	Gini's index				***Zenga's index***			
	***normal* confidence intervals**							
n	***0.9000***	***0.9500***	***0.9750***	***0.9900***	***0.9000***	***0.9500***	***0.9750***	***0.9900***
100	0.8617	0.9188	0.9485	0.9709	0.8731	0.9282	0.9593	0.9791
200	0.8821	0.9354	0.9640	0.9812	0.8875	0.9429	0.9679	0.9846
400	0.8877	0.9431	0.9687	0.9861	0.8906	0.9457	0.9726	0.9875
	***percentile* confidence intervals**							
100	0.8404	0.8971	0.9310	0.9586	0.8536	0.9101	0.9392	0.9669
200	0.8687	0.9243	0.9553	0.9756	0.8721	0.9260	0.9567	0.9772
400	0.8822	0.9345	0.9647	0.9831	0.8817	0.9369	0.9660	0.9827
	***BCa* confidence intervals**							
100	0.8671	0.9234	0.9536	0.9738	0.8729	0.9270	0.9589	0.9797
200	0.8819	0.9369	0.9646	0.9821	0.8847	0.9412	0.9668	0.9841
400	0.8847	0.9423	0.9700	0.9876	0.8892	0.9451	0.9711	0.9872
	***t-boostrap* confidence intervals**							
100	0.8863	0.9391	0.9664	0.9843	0.8833	0.9391	0.9664	0.9850
200	0.8960	0.9488	0.9715	0.9882	0.8960	0.9479	0.9736	0.9884
400	0.8919	0.9483	0.9747	0.9904	0.8973	0.9508	0.9757	0.9897

Table 2: Average size of the 95% asymptotic confidence intervals

	Dagum parent distribution ***with*** $\lambda = 44030$, $\beta = 1.055$, $\theta = 3.095$		***Lognormal parent distribution*** ***with*** $\gamma = 2.8171$, $\delta = 0.6262$	
	Gini	***Zenga***	***Gini***	***Zenga***
		***normal* confidence intervals**		
100	0.1114	0.1272	0.0938	0.1036
200	0.0850	0.0945	0.0686	0.0745
400	0.0641	0.0702	0.0494	0.0533
		***percentile* confidence intervals**		
100	0.1095	0.1258	0.0933	0.1033
200	0.0840	0.0938	0.0685	0.0743
400	0.0636	0.0698	0.0494	0.0532
		***BCa* confidence intervals**		
100	0.1162	0.1262	0.0961	0.1025
200	0.0896	0.0952	0.0702	0.0743
400	0.0676	0.0713	0.0503	0.0533
		***t-boostrap* confidence intervals**		
100	0.1512	0.1491	0.1082	0.1114
200	0.1064	0.1081	0.0747	0.0784
400	0.0763	0.0786	0.0519	0.0550

Lognorma distribution (gamma=2,8171; delta=0,6262; Gini=0,3420; Zenga=0,6774)

ONLY FOR THE REFEREE. NOT TO BE INCLUDED IN THE PAPER !!!

The bootstrap distributions were obtained by R=9999 bootstrap
replicates of the sample statistics or their standardized versions.
The estimates of the coverage accuracies and average sizes
for sample size n=100 refer to N=10000 simulated samples, whereas
(due to compuation time) the estimates of coverage accuracies and
average sizes for samples of size n=400 refer to only N=1000 simulated samples.
This explains why sometimes the estimate of the coverage
accuracy is larger for samples of size n=100, than for samples of size 400.

Gini's index (G = 0,3420)

Non parametric Normal confidence intervals

n\1-alpha	0.9000	0.9500	0.9750	0.9900	av.size
100	0.8597	0.9178	0.9514	0.9738	0.0940
400	0.8830	0.9380	0.9690	0.9830	0.0490

Parametric Normal confidence intervals

	0.9000	0.9500	0.9750	0.9900	av.size
	0.8923	0.9447	0.9700	0.9863	0.0882
	0.9010	0.9410	0.9720	0.9850	0.0443

Non parametric Percentile confidence intervals

n\1-alpha	0.9000	0.9500	0.9750	0.9900	av.size
100	0.8426	0.8945	0.9311	0.9602	0.0936
400	0.8670	0.9270	0.9640	0.9770	0.0490

Parametric Percentile confidence intervals

	0.9000	0.9500	0.9750	0.9900	av.size
	0.8862	0.9356	0.9649	0.9811	0.0882
	0.8940	0.9410	0.9700	0.9850	0.0443

Non parametric BCa confidence intervals

n\1-alpha	0.9000	0.9500	0.9750	0.9900	av.size
100	0.8663	0.9223	0.9558	0.9772	0.0962
400	0.8810	0.9390	0.9610	0.9820	0.0497

Parametric BCa confidence intervals

	0.9000	0.9500	0.9750	0.9900	av.size
	0.8972	0.9498	0.9746	0.9909	0.0900
	0.9040	0.9530	0.9720	0.9850	0.0445

Non parametric t-bootstrap confidence intervals

n\1-alpha	0.9000	0.9500	0.9750	0.9900	av.size
100	0.8863	0.9403	0.9679	0.9858	0.1083
400	0.8900	0.9450	0.9730	0.9870	0.0513

Parametric t-bootstrap confidence intervals

	0.9000	0.9500	0.9750	0.9900	av.size
	0.8975	0.9437	0.9690	0.9829	0.0950
	0.9000	0.9520	0.9720	0.9900	0.0451

Zenga's index (Z = 0,6774)

Non parami Parametric t-bootstrap confidence intervals

n\1-alpha	0.9000	0.9500	0.9750	0.9900	av.size
100	0.8975	0.9437	0.9690	0.9829	0.0950
400	0.9000	0.9520	0.9720	0.9900	0.0451

Parametric Normal confidence intervals

	0.9000	0.9500	0.9750	0.9900	av.size
	0.8960	0.9490	0.9728	0.9902	0.1031
	0.9020	0.9470	0.9740	0.9880	0.0513

Non parametric Percentile confidence intervals

n\1-alpha	0.9000	0.9500	0.9750	0.9900	av.size
100	0.8545	0.9081	0.9410	0.9665	0.1034
400	0.8590	0.9290	0.9610	0.9790	0.0528

Parametric Percentile confidence intervals

	0.9000	0.9500	0.9750	0.9900	av.size
	0.8862	0.9356	0.9649	0.9811	0.1031
	0.8940	0.9410	0.9700	0.9850	0.0513

Non parametric BCa confidence intervals

n\1-alpha	0.9000	0.9500	0.9750	0.9900	av.size
100	0.8711	0.9285	0.9594	0.9809	0.1026
400	0.8820	0.9360	0.9710	0.9810	0.0529

Parametric BCa confidence intervals

	0.9000	0.9500	0.9750	0.9900	av.size
	0.8972	0.9498	0.9746	0.9909	0.1026
	0.9040	0.9530	0.9720	0.9850	0.0513

Non parametric t-bootstrap confidence intervals

n\1-alpha	0.9000	0.9500	0.9750	0.9900	av.size
100	0.8843	0.9385	0.9683	0.9865	0.1114
400	0.8890	0.9410	0.9740	0.9870	0.0545

Parametric t-bootstrap confidence intervals

	0.9000	0.9500	0.9750	0.9900	av.size
	0.9077	0.9565	0.9802	0.9928	0.1059
	0.9050	0.9560	0.9750	0.9890	0.0517