Asymptotic Confidence Intervals for a New Inequality Measure

F. Greselin L. Pasquazzi

Dipartimento di Metodi Quantitativi per le scienze Economiche e Aziendali
University of Milano-Bicocca

Workshop on Income Distribution
Second Meeting - Milan, April 23, 2009

Introduction

In this work we compare the performance of asymptotic confidence intervals for Zenga's new inequality measure to that of Gini's traditional measure.

Normal, percentile, BCa and t-bootstrap confidence intervals have been considered.

Gini's measure is defined by

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

where

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

is the Lorenz curve associated with the distribution F and μ_F is the first moment of F, $0 < \mu_F < \infty$.
Hoeffding (1948) proved the asymptotic normality of G(F), under second order moment restrictions on F.

For continuous distributions F,
with non negative support and finite first moment,
the new Zenga inequality measure is given by

$$I(F) = \int_0^1 I_F(p)dp \tag{1}$$

where

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

is Zenga's new inequality curve.

Gini's index can be expressed also as

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

where

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

measures the relative inequality at p (Gini,1914).

The sample estimators are the sample Gini concentration ratio

$$\widehat{G} = \widehat{\Delta}/(2\overline{X})$$

for Gini's index and

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

for the Zenga inequality functional.

Normal Confidence Intervals

The normal confidence interval is based on the quantiles z_α of the standard normal distribution:

$$(\widehat{\theta} - z_{1-\alpha}\widehat{\sigma}; \widehat{\theta} + z_{1-\alpha}\widehat{\sigma}),$$

where $\widehat{\theta}$ is the inequality functional computed on the ECDF, and $\widehat{\sigma}$ is the variance of the empirical influence values.

Percentile and BCa Confidence Intervals

The percentile and BCa confidence intervals are based on the quantiles θ_α of the bootstrap distributions P^* of the statistical functionals. In our simulation study we obtained the bootstrap distribution by taking 9999 resamples.

- The percentile confidence interval is $(\theta_\alpha, \theta_{1-\alpha})$
- The BCa confidence interval is given by $(\theta_{\underline{\alpha}}, \theta_{\overline{\alpha}})$, where

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

$b = \Phi^{-1}(P^*(\widehat{\theta}))$, and a is $1/6$ times the standardized third moment of the empirical influence values.

t-bootstrap Confidence Intervals

The t-bootstrap confidence interval is based on the quantiles t_α of the bootstrap distribution of the *standardized* statistical functional,

$$(\widehat{\theta} - t_\alpha \widehat{\sigma}; \widehat{\theta} + t_{1-\alpha} \widehat{\sigma}),$$

where standard deviations are estimated by the variance of the empirical influence values.

Simulation Study

Choice of the underlying distributions

The Dagum parent distribution has density function

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

with $\widehat{\beta} = 1.055$, $\widehat{\theta} = 3.095$ and $\widehat{\lambda} = 44030$
(MLE from Italian Expenses distribution,
2002 Bank of Italy Survey on Household Income and Wealth
(8001 households)).
This yields $\widehat{G} = 0.3193$ and $\widehat{I} = 0.6505$.

Choice of the underlying distributions

The Lognormal parent distribution has density function

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

with $\widehat{\gamma} = 2.8171$ and $\widehat{\delta} = 0.6262$, from Latorre (1989) using the Italian Income distribution,
(1983 Bank of Italy Survey on Household Income and Wealth (4107 households)).
In this case $\widehat{G} = 0.3420$ and $\widehat{I} = 0.6774$.

From each distribution, we draw 10000 samples of size 100, 200 and 400.

The relative frequencies of the confidence intervals containing the true value of the inequality measures are the coverage probabilities reported in our results.

Simulation Results

Dagum ($\beta = 1.055$; $\theta = 3.095$; $\lambda = 44030$) $\Rightarrow G = 0.3193$ and $I = 0.6505$
95% confidence intervals

Coverage probabilities

	normal		percentile		Bca		*t*-bootstrap	
n	Gini	Zenga	Gini	Zenga	Gini	Zenga	Gini	Zenga
100	0.8842	0.9032	0.8571	0.8743	0.8959	0.9075	0.9226	0.9276
200	0.9005	0.9102	0.8880	0.8915	0.9020	0.9098	0.9232	0.9268
400	0.9194	0.9236	0.9108	0.9092	0.9181	0.9212	0.9338	0.9344

Average sizes

	normal		percentile		Bca		*t*-bootstrap	
n	Gini	Zenga	Gini	Zenga	Gini	Zenga	Gini	Zenga
100	0.1114	0.1272	0.1095	0.1258	0.1162	0.1262	0.1512	0.1491
200	0.0850	0.0945	0.0840	0.0938	0.0896	0.0952	0.1064	0.1081
400	0.0641	0.0702	0.0636	0.0698	0.0676	0.0713	0.0763	0.0786

Simulation Results

Lognormal ($\gamma = 2.8171, \delta = 0.6262$) $\Rightarrow G = 0.3420, I = 0.6774$

95% confidence intervals

Coverage probabilities

	normal		percentile		Bca		t-bootstrap	
n	Gini	Zenga	Gini	Zenga	Gini	Zenga	Gini	Zenga
100	0.9188	0.9282	0.8971	0.9101	0.9234	0.9270	0.9391	0.9391
200	0.9354	0.9429	0.9243	0.9260	0.9369	0.9412	0.9488	0.9479
400	0.9431	0.9457	0.9345	0.9369	0.9423	0.9451	0.9483	0.9508

Average sizes

	normal		percentile		Bca		t-bootstrap	
n	Gini	Zenga	Gini	Zenga	Gini	Zenga	Gini	Zenga
100	0.0938	0.1036	0.0933	0.1033	0.0961	0.1025	0.1082	0.1114
200	0.0686	0.0745	0.0685	0.0743	0.0702	0.0743	0.0747	0.0784
400	0.0494	0.0533	0.0494	0.0532	0.0503	0.0533	0.0519	0.0550

Simulation Results

- Samples of size 400 are large enough for accurate confidence intervals for both inequality measures.
- The t-bootstrap confidence interval performs best in terms of coverage probability at the cost of providing larger confidence intervals.
- There is no substantial difference in the coverage probabilities and in the sizes of confidence intervals for the two inequality measures.

References

Banca d'Italia, *Indagine sui bilanci delle famiglie italiane*, 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 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.

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.