

Estimating Gini and Zengas inequality indices on the ECHP dataset

F. Greselin, L. Pasquazzi

January 19, 2011

Outline

- ▶ Asymptotic theory for Zenga's index
 - ▶ asymptotic expansion of the index
 - ▶ confidence intervals in cross-sectional and longitudinal settings
- ▶ Simulation results
- ▶ Application: Confidence intervals for Zenga's index for income distributions in 15 EU countries.

Asymptotic theory for Zenga's index

Let X_1, X_2, \dots, X_n be i.i.d. random observations from F . In order that the definition of Zenga's index

$$Z_F = \int_0^1 Z_F(p) dp, \quad \text{where} \quad Z_F(p) = 1 - \frac{L_F(p)}{p} \cdot \frac{1-p}{1-L_F(p)}$$

makes sense, we need to assume that F is defined on the non-negative real axis and that its first moment is finite.

Asymptotic theory for Zenga's index

Let \tilde{Z}_n be the plug-in estimator for Zenga's index. If there exists $\varepsilon > 0$ such that $E(X_1^{2+\varepsilon}) < \infty$, then we get the asymptotic representation

$$\tilde{Z}_n = Z_F + \frac{1}{n} \sum_{i=1}^n h_F(X_i) + o_{\mathbf{P}} \left(\frac{1}{\sqrt{n}} \right),$$

where

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

with the weight function

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

Proof of asymptotic expansion

The proof relies crucially on some properties of the Vervaat process (Zitikis, 1998; Davydov and Zitikis, 2003, 2004):

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

When F is continuous we have

$$\sqrt{n} V_n(t) = o_{\mathbf{P}}(1),$$

which allows us to substitute

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

with

$$\int_0^{F^{-1}(p)} (F_n(x) - F(x))dx + o_{\mathbf{P}}\left(\frac{1}{\sqrt{n}}\right).$$

Asymptotic theory for Zenga's index

We may exploit the asymptotic expansion for computing confidence intervals

- ▶ in the one sample case (needed for confidence intervals in a cross-sectional setting)
- ▶ in case of two independent samples (needed to compare the value of Zenga's index in two populations using two independent samples of size m and n respectively)
- ▶ in case of paired Samples (needed to compare the value of Zenga's index in two different points in time in a longitudinal setting)

Simulation Study

Estimates of Coverage accuracies and size of confidence intervals based on 10000 simulated samples

sample size	Nominal confidence level				Size of 95% confidence intervals		
	0.9000	0.9500	0.9750	0.9900	min	average	max
Normal							
200	0.7881	0.8527	0.8926	0.9266	0.0674	0.1500	0.7300
400	0.8047	0.8693	0.9078	0.9396	0.0563	0.1167	0.7465
800	0.8246	0.8882	0.9237	0.9503	0.0462	0.0900	0.6535
Percentile							
200	0.7629	0.8190	0.8567	0.8892	0.0667	0.1462	0.4782
400	0.7934	0.8487	0.8864	0.9179	0.0561	0.1143	0.4721
800	0.8168	0.8751	0.9119	0.9393	0.0468	0.0884	0.4117
Bias Corrected Accelerated							
200	0.8054	0.8670	0.9047	0.9374	0.0661	0.1497	0.4652
400	0.8204	0.8860	0.9212	0.9523	0.0558	0.1186	0.4629
800	0.8338	0.8983	0.9323	0.9634	0.0467	0.0927	0.4085
t-bootstrap							
200	0.8485	0.9049	0.9400	0.9675	0.0680	0.2099	2.5148
400	0.8534	0.9120	0.9463	0.9709	0.0573	0.1559	2.1009
800	0.8572	0.9169	0.9504	0.9754	0.0474	0.1162	2.2051

EHCP Survey

- ▶ The EHCP survey is an instrument developed by Eurostat to gather information on households and individual incomes in a common framework for national systems.
- ▶ Longitudinal Panel design allows to follow up a common set of private households over several consecutive years.
- ▶ The 2001 wave sample contains data coming from 58,783 households spread over 15 European countries.
- ▶ OECD equivalence scale to deal with different household sizes (weight 1 to household head, 0.5 to the other adult members of the household, and 0.3 to the members under 14 years of age)

99% confidence intervals for Gini's index

Country	<i>n</i>	Gini	Normal		<i>t</i> -bootstrap		BCa	
Sweden	5085	0.2466	0.2336	0.2595	0.2360	0.2636	0.2364	0.2652
Denmark	2279	0.2488	0.2348	0.2628	0.2359	0.2647	0.2366	0.2650
Germany	5474	0.2569	0.2448	0.2689	0.2473	0.2732	0.2473	0.2728
Netherlands	4824	0.2615	0.2495	0.2735	0.2505	0.2762	0.2509	0.2759
Austria	2535	0.2628	0.2509	0.2746	0.2513	0.2756	0.2514	0.2756
Luxembourg	2428	0.2649	0.2526	0.2772	0.2537	0.2784	0.2537	0.2787
France	5268	0.2797	0.2708	0.2886	0.2711	0.2891	0.2714	0.2894
Finland	3106	0.2897	0.2718	0.3076	0.2746	0.3129	0.2758	0.3129
Belgium	2322	0.3005	0.2642	0.3367	0.2715	0.3789	0.2740	0.3552
Italy	5525	0.3047	0.2942	0.3151	0.2950	0.3162	0.2953	0.3164
UK	4749	0.3228	0.3064	0.3391	0.3092	0.3459	0.3102	0.3460
Ireland	1757	0.3228	0.3038	0.3418	0.3054	0.3465	0.3061	0.3460
Spain	4948	0.3233	0.3111	0.3355	0.3125	0.3377	0.3128	0.3376
Greece	3895	0.3447	0.3316	0.3577	0.3327	0.3594	0.3330	0.3592
Portugal	4588	0.3681	0.3556	0.3806	0.3562	0.3816	0.3564	0.3818

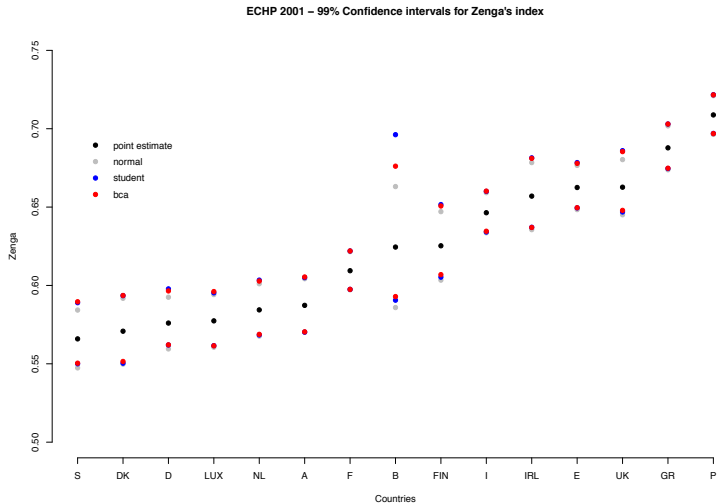
(2001 wave, ECHP income data).

99% confidence intervals for Zenga's index

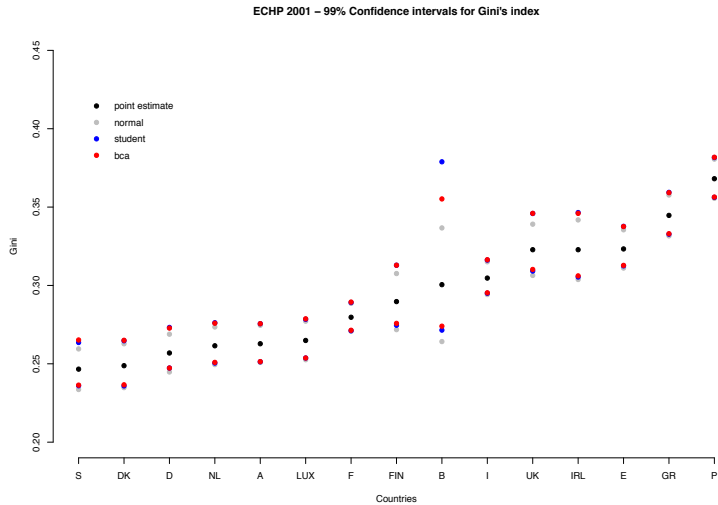
Country	<i>n</i>	Zenga	Normal		<i>t</i> -bootstrap		BCa	
Sweden	5085	0.5659	0.5474	0.5843	0.5500	0.5890	0.5504	0.5896
Denmark	2279	0.5708	0.5498	0.5918	0.5504	0.5935	0.5515	0.5936
Germany	5474	0.5760	0.5595	0.5925	0.5620	0.5978	0.5621	0.5965
Luxembourg	2428	0.5774	0.5606	0.5942	0.5616	0.5953	0.5615	0.5961
Netherlands	4824	0.5844	0.5677	0.6011	0.5684	0.6034	0.5688	0.6028
Austria	2535	0.5873	0.5702	0.6043	0.5702	0.6050	0.5704	0.6054
France	5268	0.6094	0.5972	0.6216	0.5975	0.6220	0.5975	0.6220
Belgium	2322	0.6245	0.5859	0.6631	0.5906	0.6962	0.5929	0.6761
Finland	3106	0.6253	0.6034	0.6471	0.6053	0.6516	0.6069	0.6507
Italy	5525	0.6464	0.6336	0.6592	0.6341	0.6600	0.6346	0.6602
Ireland	1757	0.6570	0.6356	0.6784	0.6371	0.6814	0.6370	0.6811
Spain	4948	0.6625	0.6484	0.6766	0.6495	0.6784	0.6496	0.6779
UK	4749	0.6627	0.6451	0.6803	0.6469	0.6860	0.6479	0.6854
Greece	3895	0.6878	0.6738	0.7018	0.6746	0.7030	0.6748	0.7030
Portugal	4588	0.7088	0.6964	0.7211	0.6970	0.7217	0.6969	0.7216

(2001 wave, ECHP income data).

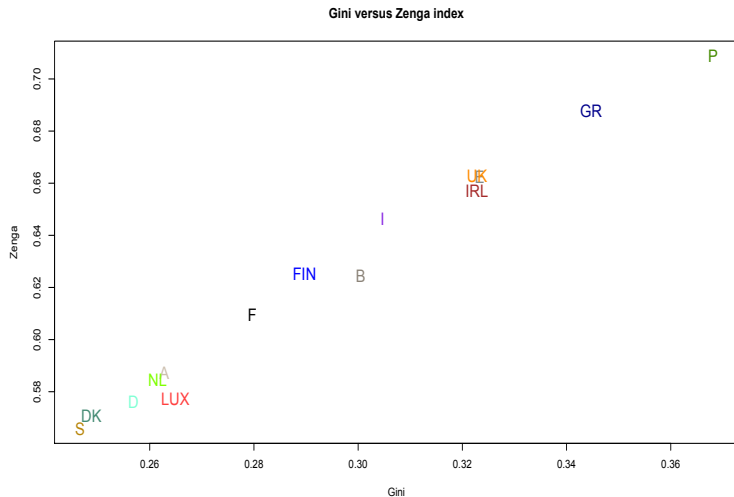
99% confidence intervals for Zenga's index



99% confidence intervals for Gini's index



Gini versus Zenga Index



References

- ▶ BICKEL, P.J. AND FREEDMAN, D.A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9, 1196–1217.
- ▶ DAVYDOV, Y. AND ZITIKIS, R. (2003). Generalized Lorenz curves and convexifications of stochastic processes. *Journal of Applied Probability*, 40, 906–925.
- ▶ DAVYDOV, Y. AND ZITIKIS, R. (2004). Convex rearrangements of random elements. In: *Asymptotic Methods in Stochastics*, Festschrift for Miklós Csörgö (Eds. L. Horváth and B. Szyszkowicz), *Fields Institute Communications*, Vol. 44, 141–171. American Mathematical Society, Providence, RI.
- ▶ ECHP (1996a). The European Community Household Panel: Volume 1 – Survey methodology and Implementation: Waves 1-3 – Theme 3, Series E, Eurostat, OPOCE, Luxembourg.
- ▶ ECHP (1996b). The European Community Household Panel: Volume 1 – Survey questionnaires: Waves 1-3 – Theme 3, Series E, Eurostat, OPOCE, Luxembourg.
- ▶ 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. (2010). Zenga's new index of economic inequality, its estimation, and an analysis of incomes in Italy. *Journal of Probability and Statistics* (Special Issue on "Actuarial and Financial Risks: Models, Statistical Inference, and Case Studies"), 2010, Article ID 718905, 26 pages.
- ▶ GRESELIN, F., PURI, M.L. AND ZITIKIS, R. (2009b). L-functions, processes, and statistics in measuring economic inequality and actuarial risks. *Statistics and Its Interface*, 2, 227–245.
- ▶ Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical 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.
- ▶ 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.