

Konfidenzintervalle für einen neuen Ungleichverteilungskoeffizienten

F. Greselin L. Pasquazzi

Abteilung für Quantitative Methoden in Wirtschaft - und Finanzwissenschaften
Universität Mailand - Bicocca

Schweizer Statistiktage 2009

In dieser Arbeit vergleichen wir asymptotische Kle für Zengas neuen Ungleichverteilungskoeffizienten

$$Z(F) = 1 - \int_0^1 \frac{1-p}{p} \frac{L_F(p)}{1-L_F(p)} dp, \quad \text{wobei } L_F(p) = \frac{\int_0^p F^{-1}(t) dt}{\int_0^1 F^{-1}(t) dt},$$

mit Klen für den Gini Koeffizienten

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

Wir untersuchen normale und einige Arten von bootstrap Klen.

Theorie

Beide Koeffizienten haben bei einer Zufallsauswahl X_1, X_2, \dots, X_n aus F eine Taylorreihenentwicklung der Art

$$T(\widehat{F}_n) = T(F) + \frac{1}{n} \sum_{i=1}^n h_F(X_i) + o_p(n^{-1/2}),$$

wobei \widehat{F}_n die Verteilungsfunktion der Stichprobe und

$$h_T(x; F) = \lim_{\lambda \downarrow 0} \frac{T(F_\lambda) - T(F)}{\lambda}, \quad F_\lambda = F + \lambda(\delta_x - F)$$

die Influenzfunktion bezeichnet. Es folgt, dass

$$\sqrt{n} \left(T(\widehat{F}_n) - T(F) \right) \xrightarrow{L} \mathcal{N}(0, \sigma_{T,F}^2), \quad \sigma_{T,F}^2 = \text{Var}_F(h_T(X; F)).$$

Computersimulation

- Computersimulation \longrightarrow $\left\{ \begin{array}{l} \text{Überdeckungshäufigkeit} \\ \text{Breite} \end{array} \right.$
- Stichprobenumfang $n = 100, 200, 400$
- 10000 Zufallsstichproben aus einer Dagumverteilung mit Dichtefunktion

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

$a, b, p > 0 \longrightarrow$ ML-Schätzungswerte der italienischen Konsumausgabenverteilung im Jahre 2002

$$\implies G(F) = 0.3193 \text{ und } Z(F) = 0.6505.$$

Normale und t-bootstrap Kle

$$(T(\hat{F}_n) - t_{1-\alpha} \hat{\sigma}_T^2 / \sqrt{n}; T(\hat{F}_n) + t_{1-\alpha} \hat{\sigma}_T^2 / \sqrt{n}),$$

- $\hat{\sigma}_T^2 = \frac{1}{n} \sum_{i=1}^n h_T(X_i; \hat{F}_n)^2;$
- $h_T(X_i; \hat{F}_n) = \frac{T(\hat{F}_n + \varepsilon(\delta_{X_i} - \hat{F}_n)) - T(\hat{F}_n)}{\varepsilon}, \varepsilon = 0.0001;$
- $t_\alpha = \begin{cases} \Phi^{-1}(\alpha) \text{ im normalen KI,} \\ \text{das } \alpha \text{ Quantil der bootstrap Vtlg von} \\ \sqrt{n} (T(\hat{F}_n) - T(F)) / \hat{\sigma}_T. \end{cases}$

Percentil und BCa Percentil Kle

Endpunkte: $(T_{p_\alpha}^*; T_{p_{1-\alpha}}^*)$, wobei

- T_p^* das p Quantil der bootstrap Vtlg von $T(\hat{F}_n)$ bezeichnet;

- $p_\alpha =$

$$\left\{ \begin{array}{l} \alpha \text{ im Percentil KI,} \\ \Phi\left(b + \frac{b+z_\alpha}{1-a(b+z_\alpha)}\right), \text{ im BCa Percentil KI;} \\ a = \frac{\sum_{i=1}^n h_T(X_i; \hat{F}_n)^3}{6 \left(\sum_{i=1}^n h_T(X_i; \hat{F}_n)^2\right)^{3/2}} \\ b = \Phi^{-1}\left(F_n^*(T(\hat{F}_n))\right), \quad F_n^* \text{ bstrp Vtlg von } T(\hat{F}_n) \end{array} \right.$$

Bootstrap Vtlgen \rightarrow 9999 bootstrap Stichproben.

Resultate

Dagum ($\beta = 1.055$; $\theta = 3.095$; $\lambda = 44030$) $\Rightarrow G = 0.3193$ und $Z = 0.6505$
 95% Konfidenzintervalle

Überdeckungshäufigkeit

n	normal		percentile		Bca		t-bootstrap	
	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

Mittlere Breite

n	normal		percentile		Bca		t-bootstrap	
	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

Resultate

- Kle für Gini und Zengas Index haben ähnliche Überdeckungshäufigkeit und Breite
- „wahres“ Kfdzniveau $<$ nominales Kfdzniveau
- Bei Stichprobenumfang $n = 400$ erreicht das t-bootstrap KI das nominale Kfdzniveau
- verhältnismäßig breite Kle \implies unpräzise Punktschätzer

Ital. Einkommensvertlg. - Ginis Koeffizient

Pos.	Region	n	Mittelwert	99% inf	Gini	99% sup
1	Valle d'Aosta	44	18131	0,1211	0,1629	0,2259
2	Piemonte	731	18958	0,2053	0,2218	0,2421
3	Trentino	173	19816	0,1966	0,2306	0,2781
4	Sardegna	334	15426	0,2272	0,2492	0,2766
5	Abruzzo	204	18835	0,2120	0,2577	0,3676
6	Umbria	256	19173	0,1780	0,2648	0,6092
7	Liguria	339	20851	0,2376	0,2663	0,3015
8	Calabria	190	12596	0,2324	0,2677	0,3106
9	Friuli	255	21610	0,2343	0,2681	0,3185
10	Toscana	619	23039	0,2413	0,2716	0,3326
11	Emilia Romagna	699	23490	0,2560	0,2868	0,3481
12	Veneto	596	20184	0,2367	0,2947	0,4918
13	Puglia	415	13471	0,2621	0,2965	0,3573
14	Sicilia	526	13299	0,2753	0,2995	0,3351
15	Molise	86	17064	0,2494	0,3018	0,3992
16	Lombardia	873	24215	0,2755	0,3038	0,3474
17	Basilicata	128	14199	0,2354	0,3088	0,4644
18	Lazio	343	20290	0,2819	0,3179	0,3836
19	Campania	595	11766	0,2921	0,3330	0,4224
20	Marche	356	21856	0,2067	0,3592	1,0642

Ital. Einkommensvertlg. - Zengas Index

Pos.	Region	n	Mittelwert	99% inf	Gini	99% sup
1	Valle d'Aosta	44	18131	0,3388	0,4242	0,5264
2	Piemonte	731	18958	0,4938	0,5208	0,5522
3	Trentino	173	19816	0,4824	0,5380	0,6054
4	Sardegna	334	15426	0,5238	0,5584	0,5996
5	Umbria	256	19173	0,4555	0,5715	0,8943
6	Abruzzo	204	18835	0,5008	0,5724	0,7106
7	Liguria	339	20851	0,5464	0,5860	0,6319
8	Friuli	255	21610	0,5394	0,5878	0,6510
9	Toscana	619	23039	0,5496	0,5913	0,6648
10	Calabria	190	12596	0,5427	0,5952	0,6526
11	Emilia Romagna	699	23490	0,5743	0,6144	0,6810
12	Veneto	596	20184	0,5502	0,6207	0,7970
13	Molise	86	17064	0,5557	0,6227	0,7284
14	Puglia	415	13471	0,5817	0,6271	0,6939
15	Sicilia	526	13299	0,5981	0,6299	0,6722
16	Lombardia	873	24215	0,5983	0,6324	0,6776
17	Basilicata	128	14199	0,5522	0,6396	0,7846
18	Lazio	343	20290	0,6091	0,6564	0,7281
19	Campania	595	11766	0,6280	0,6745	0,7524
20	Marche	356	21856	0,5322	0,6761	1,1035

Ital. Einkommensvertlg. - Zengas Index

Pos.	Region	n	Mittelwert	99% inf	Gini	99% sup
1	Valle d'Aosta	44	18131	0,3388	0,4242	0,5264
2	Piemonte	731	18958	0,4938	0,5208	0,5522
3	Trentino	173	19816	0,4824	0,5380	0,6054
4	Sardegna	334	15426	0,5238	0,5584	0,5996
5	Umbria	256	19173	0,4555	0,5715	0,8943
6	Abruzzo	204	18835	0,5008	0,5724	0,7106
7	Liguria	339	20851	0,5464	0,5860	0,6319
8	Friuli	255	21610	0,5394	0,5878	0,6510
9	Toscana	619	23039	0,5496	0,5913	0,6648
10	Calabria	190	12596	0,5427	0,5952	0,6526
11	Emilia Romagna	699	23490	0,5743	0,6144	0,6810
12	Veneto	596	20184	0,5502	0,6207	0,7970
13	Molise	86	17064	0,5557	0,6227	0,7284
14	Puglia	415	13471	0,5817	0,6271	0,6939
15	Sicilia	526	13299	0,5981	0,6299	0,6722
16	Lombardia	873	24215	0,5983	0,6324	0,6776
17	Basilicata	128	14199	0,5522	0,6396	0,7846
18	Lazio	343	20290	0,6091	0,6564	0,7281
19	Campania	595	11766	0,6280	0,6745	0,7524
20	Marche	356	21856	0,5322	0,6761	1,1035

Ital. Einkommensvertlg. - Zengas Index

Pos.	Region	n	Mittelwert	99% inf	Gini	99% sup
1	Valle d'Aosta	44	18131	0,3388	0,4242	0,5264
2	Piemonte	731	18958	0,4938	0,5208	0,5522
3	Trentino	173	19816	0,4824	0,5380	0,6054
4	Sardegna	334	15426	0,5238	0,5584	0,5996
5	Umbria	256	19173	0,4555	0,5715	0,8943
6	Abruzzo	204	18835	0,5008	0,5724	0,7106
7	Liguria	339	20851	0,5464	0,5860	0,6319
8	Friuli	255	21610	0,5394	0,5878	0,6510
9	Toscana	619	23039	0,5496	0,5913	0,6648
10	Calabria	190	12596	0,5427	0,5952	0,6526
11	Emilia Romagna	699	23490	0,5743	0,6144	0,6810
12	Veneto	596	20184	0,5502	0,6207	0,7970
13	Molise	86	17064	0,5557	0,6227	0,7284
14	Puglia	415	13471	0,5817	0,6271	0,6939
15	Sicilia	526	13299	0,5981	0,6299	0,6722
16	Lombardia	873	24215	0,5983	0,6324	0,6776
17	Basilicata	128	14199	0,5522	0,6396	0,7846
18	Lazio	343	20290	0,6091	0,6564	0,7281
19	Campania	595	11766	0,6280	0,6745	0,7524
20	Marche	356	21856	0,5322	0,6761	1,1035

Bibliografie



Banca d'Italia,

Indagine sui bilanci delle famiglie italiane, 2004.



M. Zenga.

Inequality Curve and Inequality Index based on the Ratios
between Lower and Upper Arithmetic Means,
Statistica & Applicazioni, Vol V, 1, pp.3-27,