Rivista Italiana di Economia Demografia e Statistica *Volume LXXII n. 2 Aprile-Giugno 2018*

INFERENCE FOR INEQUALITY MEASURES: A REVIEW

Margherita Gerolimetto, Stefano Magrini

1. Introduction

After the seminal paper by Schluter and Trede (2002), who introduced concepts and tools for extreme value theory to the econometrics, part of the literature on inequality measures has concentrated on the consequences on inference due to heavy upper tail of the income distribution. More in details, the focus is on inferential problems created by population income distributions whose right tail decays slowly like a power function, *i.e.*

$$P(Y > y) \sim \beta y^{-\alpha} \quad \text{as} \quad y \to \infty \qquad (1)$$

α is the stability index and provides information on moments finiteness, if $\alpha < 1$, the mean is infinite mean and also the variance, if $\alpha < 2$, the variance is infinite.

Examples of this type of distributions, often called "heavy tailed", are the Singh-Maddala, Pareto, Generalized Beta distributions. For these heavy tailed distributions, standard methods of inference, both asymptotic and bootstrap, are unreliable. The underlying intuition is that as tail heaviness increases, the population moments increase and eventually cease to exist, whilst the (finite) sample moments tend to underestimate them.

In the attempt to overcome severe inference problems, a number of contributes have presented in the literature. Broadly speaking, works on this topic can be divided into two groups. The first group includes methods focusing on point measures of inequality and aiming at handling the limits of conventional inference (in particular, bootstrap inference) in the presence of heavy tails. The second group includes methods based on partially ordering income distributions using stochastic dominance and related criteria, one such is Donald *et al.* (2012).

In the present review, we will concentrate in particular on the first group of methods about which a heat debated has developed in the last 10 years. On the one hand, there are approaches oriented to improve the finite sample performance in case of very heavy tails, based on bootstrap and permutation (Davidson and Flachaire, 2007; Dufour *et al*. 2017). On the other hand, there are approaches based on deriving asymptotic expansions for the distributions, to transform the statistics opportunely (Schluter, 2012; Schluter and van Garderen, 2009).

This paper is structured as follow. In the second section the contribute by Davidson and Flachaire (2007) is discussed. The third section is devoted to Schluter (2012). The fourth section presents the very recent paper by Dufour *et al*. (2017). The fifth section concludes.

2. Davidson and Flachaire (2007)

Davidson and Flachaire (2007), DF07 hereafter, begin by studying the finite-sample performance of asymptotic and bootstrap inference for inequality measures, in particular they concentrate on the Generalized Entropy class. Their preliminary simulations show that, in spite of large samples of usually *iid* observations, neither asymptotic nor standard bootstrap inference perform well, in particular severe overrejection of the null hypothesis is documented. As anticipated in the first section, the reason of this is that asymptotic and bootstrap inference is very sensitive to the exact nature of the upper tail of the income distribution, especially in cases of infinite variance.

Focussing on the Theil index[1], DF07 propose two alternative bootstrap to deal with these cases of infinite variance: i) a revised version of the *m* out of *n* bootstrap (moon bootstrap) and ii) a semiparametric bootstrap.

The *m* out of *n* bootstrap (Politis and Romano, 1994; Bickel *et al.* 1997) is based on bootstrap samples of dimension $m<n$, where *n* is the original sample size, and it is known in literature as a good option when the standard bootstrap fail or it is difficult to prove its consistency.

The performance of the moon bootstrap is shown via simulations, for the Theil index, *n*=50, *m*=2,...,50, from a Singh-Maddala distribution which can successfully mimic observed income distributions in various countries (Brachman *et al.*, 1996) and whose cumulative distribution function is

$$F(y) = 1 - \frac{1}{(1+ay^b)^c} \quad (2)$$

[1] For a random variable y with cumulative distribution function F, the Theil index can be written as $T(F) = \left(\frac{v_F}{\mu_F}\right) - \log(\mu_F)$, where $\mu_F = E_F(y)$ and $v_F = E_F(y \log y)$.

where a=100, b=2.8, c=1.7. The tail parameter α=bc=4.76 is a choice that closely mimics the net income distribution of German households. DF07 results show that for m=22 the percentage of rejection is very close to the nominal level. However, the moon bootstrap performance is very sensitive to the choice of m. In particular, for m=n the moon bootstrap coincides with the standard one and approaches serious overrejection; by reducing m the number of rejections decreases, but when m is too small, the moon bootstrap approaches underrejection. Given these results, DF07 introduce a different version of the moon bootstrap. To do this, they firstly quantify the bias that leads to overrejection, then they approximate it via the moon bootstrap (with some choice of m), finally, they obtain an adjusted p-value (denoted revised moon bootstrap p-value).

As an alternative to the revised moon bootstrap, DF07 propose a semiparametric bootstrap that combines a parametric bootstrap for the right tail with a standard bootstrap for the main body of the distribution. In a first step, it is estimated the index of stability of the right tail of the distribution (resorting to the k greatest order statistics of a sample of size n, for some integer $k \leq n$)

$$\hat{\alpha} = H_{k,n}^{-1} \text{ and } H_{k,n} = k^{-1} \sum_{i=0}^{k-1} \log y_{n-1} - \log y_{n-k+1} \tag{3}$$

where y_j is the j-th order statistic of the sample, the choice of k is a matter of trade-off between bias and variance. Common choice, usually based on graphical methods, and adopted here, is the square root of the sample size (Coles, 2001).

In a second step, the bootstrap samples are drawn from a distribution defined as a function of a probability p_{tail} that constitutes the tail of the distribution. Each observation of the bootstrap sample is, with probability p_{tail}, a drawing from the distribution with cumulative density function

$$F(y) = 1 - (y/y_0)^{\hat{\alpha}}$$

where $y > y_0$ and y_0 is the order statistic of rank $\bar{n} = n(1 - p_{tail})$ of the sample and, with probability $1 - p_{tail}$, a drawing from the empirical distribution of the sample of the smallest $n(1 - p_{tail})$.

Table 1 – *ERP (*$\boldsymbol{\alpha}$*=4.76, h=0.4, m is the closest integer to* $\sqrt{\boldsymbol{n}}$*)*

n	50	100	500	1000	2000	3000
Asymptotic	0.140	0.115	0.071	0.062	0.052	0.051
Standard bootstrap	0.058	0.049	0.032	0.022	0.021	0.020
Moon bootstrap (revised)	-0.043	-0.0018	0.021	0.020	0.020	0.020
Semiparametric bootstrap	0.026	0.022	0.005	0.001	-0.002	0.003

Simulations results by DF07 from Singh-Maddala distribution are summed up in tables 1 and 2, where $p_{tail} = hk/n$, *h=0.3, 0.4, 0.6, 0.8*. The performance is in terms of Error in Rejection Probability, ERP, (left hand tail) for the Theil index, hence the closer the figures are to zero, the better. The revised moon bootstrap yields a slight improvement over the standard bootstrap. For the the semiparametric bootstrap, instead, the performance improves dramatically over the standard bootstrap, with insignificant ERPs for sample sizes greater than around 1000 (table 1). Still ERP are unacceptably high in case of very heavy tails (table 2).

Table 2 – *ERP (in case of heavier tail, h=0.4, n=100, m is the closest integer to* $\sqrt{\boldsymbol{n}}$*)*

	Asy	Stand boot	Moon boot (rev)	Semiparam boot
α=2.1	0.41	0.24	0.15	0.16
α=1.9	0.48	0.28	0.20	0.18

All in all, the simulation results by DF07 show that by adopting their bootstrap proposals the inferential problem is mitigated, especially with the semiparametric bootstrap, but still performance deteriorates as the tail of the income distribution becomes heavier.

3. Schluter (2012)

Moving from the results by DF07 and by Schluter and van Gardener (2009), Schluter (2012), hereafter SCH12, proposes a normalizing transformation of inequality measures, in particular the Generalized Entropy class. The work is based on Edgeworth expansions to adjust asymptotic Gaussian approximations in order to deal with the inference problem due to the heavy upper tail of the income distribution discussed above.

SCH12 begins by observing a systematic relationship between $\hat{I}$ and $\widehat{var}(\hat{I})$ as potentially responsible of the severity of the inference problem. In particular, by plotting pairs of $\hat{I}$ and $\widehat{var}(\hat{I})$ and observing the corresponding coverage error, the author recognize that the wrong confidence limits are associated to particularly low realizations of both $\hat{I}$ and $\widehat{var}(\hat{I})$. Exploiting this relationship suggests the application of a variance stabilizing transform:

$$H(I) = \int_0^I \frac{du}{[\sigma(u)^2]^{1/2}}$$
(4)

where $\sigma^2(I)$ denotes the variance as a function of I. In conjunction with a bootstrap, the transform in (4) reduces the inference problem significantly. Finally, SCH12 develops asymptotic expansions for studentized (based on estimated variance)

$$S_n = \sqrt{n}\left(\frac{\hat{I}-I}{\hat{\sigma}}\right) \quad (5)$$

where $\hat{\sigma}$ is the asymptotic standard deviation, derived with the delta method, and standardized inequality measures (based on the theoretical variance)

$$S_n = \sqrt{n}\left(\frac{\hat{I}-I}{\sigma}\right) \quad (6)$$

A finite sample experiment documents (whose detailed results are not reported here) the positive effects of the stabilizing variance transform for various levels of the heaviness of the tail for a Singh-Maddala distribution. Compared to the poor quality of the Gaussian approximation (discussed in the previous section), the performance of the studentized bootstrap, coupled with the stabilizing variance transform, improves. In spite of this improvement in the performance, however for $\alpha = 2$ there is still a substantial difference between the nominal and the actual coverage behaviour.

4. Dufour, Flachaire and Khalaf (2018)

Dufour, Flachaire and Khalaf (2018), hereafter DFK18, propose Monte Carlo permutation and bootstrap methods for the problem of testing the equality of inequality measures between two samples. Their results cover the Generalized Entropy class. In addition to the previously discussed problems of heavy upper tail, DFK18 emphasize that inequality measures inference can also be confounded because those indices, as functionals of the cumulative distribution function, can be equal even in case the underlying distributions differ.

Consider two *iid* samples $X = \{X_1, X_2, \ldots, X_n\}$ and $Y = \{Y_1, Y_2, \ldots, Y_m\}$, from cumulative distribution functions F_X and F_Y, and the hypothesis testing

$$H_0: \theta(F_X) = \theta(F_Y) \quad (7)$$

where $\theta(.)$ is some functional on some subset F of distributions. Inequality indices are special cases of $\theta(.)$. A natural statistic is:

$$T = \theta(\hat{F}_X) - \theta(\hat{F}_Y)$$
(8)

where $\hat{F}_X$ and $\hat{F}_Y$ are the empirical cumulative distributions and the studentized version is

$$S = \frac{\theta(\hat{F}_X) - \theta(\hat{F}_Y)}{\sqrt{\hat{V}(\theta(\hat{F}_X)) + \hat{V}(\theta(\hat{F}_Y))}} \quad (9)$$

and, for both T and S, DFK18 consider three p-values (in addition to the asymptotic p-value, based on the Gaussian limiting distribution): (i) MC permutation p-value (ii) bootstrap p-value (iii) bootstrap p-value that imposes the null hypothesis.

The permutation p-value (i) is obtained from the distribution derived by permuting in all possible ways the $N = n + m$ observations of the combined sample

$$Z = \{X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m\}$$

under the assumption of *iid* samples. The permutations, in total $(m + n)!$, are all equally probably which in turn determines the permutational distribution of T or S.

B permutations are drawn at random (Dwass, 1957) from the set of all permutations and along with the actual data this yields $B + 1$ random permutations of Z., $\hat{F}_{X^*}$ and $\hat{F}_{Y^*}$ are the corresponding cumulative distribution function and the value of the test statistic is:

$$T_* = \theta(\hat{F}_X) - \theta(\hat{F}_Y)$$
(10)

The following is the permutation p-value function, where $j = 1, \dots, B$ refer to the series of permutation statistics and **1**(.) is the indicator function:

$$p_* = 2\min\left(\frac{\sum_{j=1}^{B} \mathbf{1}(T_{*j} \le x) + 1}{B+1}; \frac{\sum_{j=1}^{B} \mathbf{1}(T_{*j} \ge x) + 1}{B+1}\right) \quad (11)$$

similar arguments hold for S_{*j}, the studentized version of T_{*j}.

The bootstrap p-value (ii) is obtained moving from the bootstrap samples, (X_b, Y_b) and $\hat{F}_{X_b}$ and $\hat{F}_{Y_b}$, their corresponding empirical cumulative distribution functions; the bootstrap statistic is

$$S_b = \frac{\left(\theta\left(\hat{F}_{X_b}\right) - \theta\left(\hat{F}_{Y_b}\right)\right) - \left(\theta(\hat{F}_X) - \theta(\hat{F}_Y)\right)}{\sqrt{\hat{V}\left(\theta\left(\hat{F}_{X_b}\right)\right) + \hat{V}\left(\theta\left(\hat{F}_{Y_b}\right)\right)}} \tag{12}$$

hence for a two-tailed test, the bootstrap p-value, based on $j = 1, \dots, B$ bootstrap statistics:

$$p_b = 2 \min\left(\frac{1}{B}\sum_{j=1}^{B} \mathbf{1}\left(S_{bj} \leq S_0\right); \frac{1}{B}\sum_{j=1}^{B} \mathbf{1}\left(S_{bj} > S_0\right)\right) \tag{13}$$

It is interesting to observe that the permutation approach does not differ radically from the bootstrap approach. For example, a sample obtained by permuting elements of the combined sample Z is equivalent to resampling without replacement N observations from Z. Thus, resampling with replacement from Z represents an alternative bootstrap sample that respects the null hypothesis from which the bootstrap p-value under the null (iii) can be derived. This bootstrap sample is denoted by $(X_\circ, Y_\circ)$, $\hat{F}_{X_\circ}$ and $\hat{F}_{Y_\circ}$ are the corresponding empirical cumulative distribution functions; the bootstrap statistic is:

$$S_\circ = \frac{\left(\theta(\hat{F}_{X_\circ}) - \theta(\hat{F}_{Y_\circ})\right)}{\sqrt{\hat{V}\left(\theta(\hat{F}_{X_\circ})\right) + \hat{V}\left(\theta(\hat{F}_{Y_\circ})\right)}} \tag{14}$$

and for a two-tailed test the bootstrap p-value under the null, based on $j = 1, \dots, B$ bootstrap statistics, is

$$p_\circ = 2 \min\left(\frac{1}{B}\sum_{j=1}^{B} \mathbf{1}(S_\circ \leq S_0); \frac{1}{B}\sum_{j=1}^{B} \mathbf{1}(S_\circ > S_0)\right) \tag{15}$$

DFK18 provide asymptotic conditions for the validity[2] of the proposed methods (Romano, 1990; Chung and Romano, 2013) and they do it for two scenarios. In one case, they consider testing equality of the inequality measures when the population have the same distributions; this is equivalent to test $H_0: F_X = F_Y$ versus the alternative $\theta(F_X) \neq \theta(F_Y)$. Under this circumstance both level and size of the tests can be controlled, irrespective whether the distribution F_X or F_Y is continuous or discrete, without any restriction on the form of the functional θ. Moreover, permutation tests are exact both for the T and S statistic.

[2] Validity is intended in the sense that under the null hypothesis the rejection frequency tends to the nominal level as the sample size increases.

In the second case, when the populations do not have the same distributions things are more complicated since permutations test are no longer exact. However, such tests can be asymptotically valid if some restrictions are satisfied. In particular, for $\theta(.)$ linear functional, if

$$V_{as}\big(\theta(\hat{F}_X)\big) = V_{as}\big(\theta(\hat{F}_Y)\big)$$

implying that $n^{1/2}\left(\left(\theta\left(\hat{F}_X\right) - \theta(F_X)\right)\right.$ and $m^{1/2}\left(\left(\theta\left(\hat{F}_Y\right) - \theta(F_Y)\right)\right.$ have the same asymptotic variance or if $\frac{m}{m+n} \to \frac{1}{2}$ as $n \to \infty$ implying sample sizes asymptotically equal, asymptotic validity of the permutation test is guaranteed.

Note that the (arithmetic) mean is a linear functional, but the quantile is not. So comparing means from samples of similar size is then asymptotically valid when the sample sizes are similar even if the underlying distributions are not identical, while comparing quantiles with a permutation test is no longer valid, in general, if the underlying distribution are not identical.

The Generalized Entropy class considered by the authors is not a linear functional, unless the mean in the group is the same. However, given the scale invariance property of the Generalized Entropy class, it is possible to base a permutation test on the rescaled samples

$$\left\{\frac{X_1}{\mu(F_X)}, \dots, \frac{X_n}{\mu(F_X)}\right\} \text{ and } \left\{\frac{Y_1}{\mu(F_Y)}, \dots, \frac{Y_m}{\mu(F_Y)}\right\}$$

Comparing these indices from the rescaled samples makes no difference, while it validates (asymptotically) the use of permutation test. In practice, the following combined samples will be used where the sample means $\bar{X}$ and $\bar{Y}$ are adopted:

$$Z_s = \left\{\frac{X_1}{\bar{X}}, \dots, \frac{X_n}{\bar{X}}, \frac{Y_1}{\bar{Y}}, \dots, \frac{Y_m}{\bar{Y}}\right\}$$

and the ame holds for Gini index.

DFK18 carry out a simulations study focusing on extreme cases of heavy-tailed distributions in small samples to stress-test the methods employed in testing for Theil and Gini index. Competitors p-values are: asymptotic, standard bootstrap test S_b, permutation T rescaled (T_* based on Z_s), permutation S rescaled (S_* based on Z_s), permutation S standard (T_* based on Z), bootstrap S rescaled (bootstrap $S_\circ$ based on Z_s), bootstrap S standard (bootstrap $S_\circ$ based on Z).

Table 3 – *Theil index – Empirical size (same distribution, (r) stands for rescaled sample)*

α	asy	Boot S_b	Perm T_* (r)	Perm S_* (r)	Perm S_*	boot $S_\circ$ (r)	boot $S_\circ$
2.9	0.083	0.082	0.130	0.065	0.051	0.041	0.038
3.2	0.075	0.081	0.115	0.061	0.051	0.042	0.038
3.5	0.071	0.082	0.092	0.060	0.050	0.043	0.041
4	0.068	0.080	0.075	0.055	0.050	0.043	0.041
5	0.064	0.079	0.063	0.051	0.050	0.045	0.041
6	0.061	0.078	0.060	0.050	0.050	0.048	0.047

The size part of the experiment is based on data generated from several Singh-Maddala distributions for which the Theil (and Gini) inequality index is the same and the tail index varies in $[2.9,6.26]$, the value 2.9 corresponds to the most severe case of heavy tail. The sample size is small, $m = n = 50$. As a benchmark it is used the Singh-Maddala distribution with tail index equal to 4.76, the one used in DF07.

Results for the case of identical distributions are presented in tables 3 (Theil index) and 4 (Gini index).

Table 4 – *Gini index – Empirical size (same distribution, (r) stands for rescaled sample)*

α	asy	boot S_b	perm T_* (r)	perm S_* (r)	perm S_*	boot $S_\circ$ (r)	boot $S_\circ$
2.6	0.125	0.072	0.095	0.065	0.052	0.055	0.047
2.9	0.117	0.071	0.087	0.064	0.051	0.054	0.048
3.2	0.106	0.071	0.074	0.060	0.050	0.0500	0.048
4	0.090	0.070	0.068	0.057	0.050	0.050	0.049
5	0.081	0.069	0.062	0.053	0.050	0.050	0.049
6	0.077	0.068	0.057	0.050	0.050	0.050	0.049

As we can see from tables 3 and 4, in case of identical distributions results show that asymptotic and standard bootstrap do not perform well for the heaviest tails cases, whilst permutation tests and bootstrap based on S statistic perform well. Test based on original samples provide exact inference, unlike test based on rescaled samples (this is because rescaling is done via the sample mean).

Table 5 – *Theil index – Empirical size (different distribution, (r) stands for rescaled sample)*

α	asy	boot S_b	perm T_* (r)	perm S_* (r)	perm S_*	boot $S_\circ$ (r)	boot $S_\circ$
2.9	0.145	0.171	0.173	0.125	0.124	0.115	0.116
3.2	0.120	0.135	0.137	0.104	0.096	0.090	0.088
3.5	0.083	0.105	0.090	0.071	0.068	0.062	0.062
4	0.071	0.091	0.073	0.055	0.053	0.050	0.050
4.76	0.062	0.087	0.065	0.053	0.052	0.047	0.048
5.5	0.061	0.086	0.061	0.051	0.051	0.048	0.048
6	0.062	0.086	0.061	0.051	0.051	0.049	0.050

Table 6 – *Gini index – Empirical size (different distribution, (r) stands for rescaled sample)*

α	asy	boot S_b	perm T_* (r)	perm S_* (r)	perm S_*	boot $S_\circ$ (r)	boot $S_\circ$
2.6	0.138	0.090	0.105	0.085	0.083	0.082	0.081
2.9	0.082	0.087	0.092	0.077	0.078	0.070	0.070
3.5	0.082	0.072	0.071	0.091	0.059	0.057	0.056
4	0.078	0.070	0.068	0.058	0.055	0.053	0.053
4.76	0.077	0.069	0.065	0.058	0.054	0.050	0.050
5	0.072	0.068	0.065	0.054	0.053	0.050	0.050

In case of different distributions (empirical size in tables 5 and 6), the distribution of X is fixed (α=4.76) and the distribution of Y changes and its tail becomes heavier (the lower the value of α). The results show an overall performance that worsens the more F_Y is heavy tailed than F_X. Permutation tests and bootstrap based on S statistic perform similarly and better than other methods.DFK18 also consider the effects of the increase of the sample size (results not reported here). When $F_X = F_Y$, the rejection frequencies decrease slowly for asymptotic and standard bootstrap test. Instead, permutation and bootstrap under the null based on studentized statistic perform very well in all cases. When $F_X \neq F_Y$ the rejection frequencies decrease slowly for all methods, but permutation and bootstrap under the null based on studentized statistic outperform the other methods.

DFK18 also study the effects of the unequal sample sizes. In case $F_X = F_Y$, the more unequal are the sample sizes, the overrejections grow quickly for asymptotic and bootstrap test. Instead, permutation based on studentized statistic perform very well in all cases. In case $F_X \neq F_Y$ for all methods overrejections grow quickly as the sample sizes are more unequal, but permutation tests this happens more slowly.

As for the power side of the Monte Carlo experiment, DFK18 test the equality of an inequality measure between two samples, when the sample come from two distributions with different value of the inequality measure (F_X is fixed and F_Y varies). Power comparison of the considered permutation and bootstrap methods are valid since rejection probabilities under the null hypothesis $\theta(F_X) - \theta(F_Y) = 0$ are close to the nominal level. The permutation approach (rescaled and standard) is more powerful than the bootstrap under the null (rescaled and standard), the difference between the two approaches being resampling without replacement rather than with replacement. Studentized permutation test based on rescaled outperforms all other methods, especially when F_Y is heavier tailed than F_X.

All in all, results show that Monte Carlo methods outperform competitors both in terms of empirical size and power. Substantial reduction in size distortion is achieved more generally and studentized rescaled permutation tests outperforms the competing methods in terms of power.

5. Conclusions

In this work we present a review of some contributes of the econometric literature on comparing inequality measures. The main issue behind this bulk of recent literature is the heavy right tail of the income distribution, a condition under which standard methods of inference, both asymptotic and bootstrap are unreliable.

The papers we review are attempts to deal with this inference severe problem focusing on bootstrap and permutation, as well as on asymptotic expansion. Theoretical and simulations performance has been provided by all authors and, as far as we can tell by results, it is, in particular, the most recent proposal based on permutation methods by DFK18 the one that seems to better handle the issue.

References

BICKEL P., GOTZE F., VAN ZWET W.R., 1997. Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica* Vol 7, pp. 1–32.

BRACHMAN K., STICH A., TREDE M., 1996. Evaluating parametric income distribution models. *Allgemeines Statistisches Archiv*, Vol. 80, pp. 285–298.

COLES, S., 2001. *An Introduction to Statistical Modeling of Extreme Values.* Springer, London.

CHUNG E., ROMANO J. P., 2013. Exact and asymptotically robust permutation tests. *Annals of Statistics*, Vol. 41, pp. 484–507.

DAVIDSON R., FLACHAIRE E., 2007. Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics,* Vol. 141, pp. 141–66.

DONALD S.G., HSU Y.-C., BARRET G..F., 2012. Incorporating covariates in the measurement of welfare and inequality: methods and applications, *The Econometrics Journal*, Vol. 5, pp. C1-C30.

DUFOUR J.-M., FLACHAIRE E., KHALAF L., 2018. Permutation tests for comparing inequality measures. *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2017.1371027

DWASS M., 1957. Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics*, Vol. 28, pp. 181–187.

POLITIS D.N., ROMANO J.P., 1994. The stationary bootstrap. *Journal of American Statistical Association*, Vol. 89, pp. 1303–1313.

ROMANO J. P., 1990. On the behavior of randomized tests without a group invariance assumption. *Journal of the American Statistical Association*, Vol. 85, pp. 686–692.

SCHLUTER C., VAN GARDEREN K. J., 2009. Edgeworth Expansions and Normalizing Transforms for Inequality Measures. *Journal of Econometrics*, Vol. 150, pp. 16–29.

SCHLUTER C., TREDE M., 2002. Tails of Lorentz curves. *Journal of Econometrics*, Vol. 109, pp. 151–166.

SCHLUTER C., 2012. On the problem of inference for inequality measures for heavy-tailed distributions. *The Econometrics Journal*, Vol. 15, pp. 125–153.

SUMMARY

Inference for inequality measures: a review

In this paper we present a review of the most recent contributes of the econometric literature on comparing inequality measures, focusing in particular on Theil index and Gini index.

We will start by discussing the main issue behind this bulk of literature, which is the heavy tail of the income distribution. Specifically, the severity of the inference problem responds to the exact nature of the right tail of the distribution. Attention in the literature has been given to determining the limits of conventional inference in the presence of heavy tails and, in particular, of bootstrap inference. Then we review a number of methods based on alternative parametric bootstrap and, more recently on permutations that heated in this debate in the last 10 years.

MARGHERITA GEROLIMETTO, Università Ca' Foscari Venezia, Dipartimento di Economia. E-mail: margherita.gerolimetto@unive.it

STEFANO MAGRINI, Università Ca' Foscari Venezia, Dipartimento di Economia. E-mail: stefano.magrini@unive.it