Robust Inference on Income Inequality: t-Statistic Based Approach

Rustam Ibragimov[a,b,e], Paul Kattuman[c], Anton Skrobotov[d]

[a] Imperial College Business School

[b] New Economic School

[c] Judge Business School, the University of Cambridge

[d] HSE University

[e] Saint Petersburg State University

Abstract

Empirical analyses of income and wealth inequality often face the difficulty that the observations are heterogeneous, heavy-tailed or correlated in some unknown fashion. This paper focuses on applications of the recently developed computationally simple approach t-statistic based robust inference approach in the analysis of inequality. Two regions can be compared in terms of inequality as follows: the data in the samples relating to the two regions are partitioned into small numbers of groups, and the chosen inequality index/measure is estimated for each group. Inference is then based on standard t-tests with the resulting group estimators. The t-statistic based approach results in valid inference, as long as the group estimators of the inequality index are asymptotically independent, unbiased, and Gaussian, possibly with different variances. These conditions are typically satisfied in empirical applications. The presented method complements and compare favorably with other approaches to inference on inequality. We apply this approach to examine income inequality across Russian regions. Our analysis reveals that income distribution in Russia is notably heavy-tailed, with most regions exhibiting higher levels of inequality compared to Moscow. Robust comparisons of this type offer a good foundation for evaluating and shaping regional policies aimed at addressing income disparities.

Keywords: Income inequality, inequality indices, robust inference, heavy-tailedness, Russian economy.

1 Introduction

Many studies have found that income and wealth distributions (denoted by $X > 0$) are heavy-tailed and follow power laws.

[0]CONTACT: Rustam Ibragimov; irustam@imperial.ac.uk; Imperial College Business School, South Kensington Campus, London SW7 2AZ, United Kingdom

$$P(X > x) \sim Cx^{-\zeta}, \quad C > 0, \tag{1}$$

for large $x > 0$, with the tail index $\zeta > 0$ (see, among others, the discussion and reviews in Piketty and Saez, 2003, Atkinson, 2008, Gabaix, 2009, Gabaix et al., 2016, Toda and Wang, 2021, and references therein).[1]

The tail index parameters (ζ) of power law distributions (1) characterize the heaviness (the rate of decay) of their tails and thus govern the likelihood of observing outliers and extreme values of the random variables (r.v.s). Smaller values of ζ correspond to more pronounced heavy-tailedness in the distribution, and vice versa. Indeed, The values of ζ in power law income or wealth distributions (1) have been used as a measure of upper tail inequality (that is, inequality among the rich), with smaller values of the tail index corresponding to greater inequality in the upper tail.[2] Importantly, the tail index parameter governs the existence of moments of the r.v. $X > 0$ with power law distribution (1), with the moment EX^p of order $p > 0$ of X being finite if and only if $\zeta > p$. In particular, the second moment EX^2 of the r.v. X is finite and its variance $\mathrm{Var}(X)$ is defined if and only if $\zeta > 2$; and the first moment - the mean EX - of X is finite if and only if $\zeta > 1$. Empirical estimates suggest $\zeta \in (1.5, 3)$ for income distributions, and $\zeta \approx 1.5$ for wealth distributions. The implication is that the population variance is infinite for wealth and may be infinite for income.

Applicability of commonly used approaches to inference on inequality based on asymptotic normality of estimators of inequality measures becomes problematic under heavy-tailedness, heterogeneity, and correlation in the data. For example, estimators of measures of inequality such as the Gini coefficient converge to non-Gaussian limits given by stable r.v.'s under sufficiently pronounced heavy-tailedness, with infinite second moments and variances (see Fontanari et al., 2018, and the discussion in Appendix B).[3] Even when normal convergence holds for an estimator of an inequality index, asymptotic methods based on it often have poor finite sample properties in the presence of extreme values, i.e., when the data are heavy-tailed.[4] Similar problems also plague bootstrap methods (see

[1]As is well-known, heavy-tailedness and power law distributions are also exhibited by many other key variables in economics and finance, including financial returns, foreign exchange rates, insurance risks and losses from natural disasters, to name a few (see, among others, the reviews in Embrechts et al., 1997, Gabaix, 2009, Ibragimov et al., 2015, McNeil et al., 2015, and references therein).

[2]This may be motivated by the fact that in the case of Pareto distributions with $\zeta > 1$ for income or wealth, where (1) holds exactly for all values x greater than a certain threshold x_m, the Gini coefficient of inequality over the whole income/wealth distribution is equal to $1/(2\zeta - 1)$ and is thus decreasing in ζ (see also the discussion in Atkinson, 2008, Gabaix et al., 2016, that focuses on the analysis and estimation of the top income inequality measure $\eta = 1/\zeta$, Blanchet et al., 2018, and Ibragimov and Ibragimov, 2018).

[3]As is well-known, the finiteness of variances for the r.v.s is crucial for applicability of standard statistical and econometric methods, including regression and least squares. Similarly, the problem of potentially infinite fourth moments of economic and financial variables and time series needs to be taken into account in applications of autocorrelation-based methods and inference procedures (see, among others, the discussion in Granger and Orr, 1972, Embrechts et al., 1997, Cont, 2001, Ch. 1 in Ibragimov et al., 2015, and references therein.)

[4]More generally, poor finite sample properties are often observed for asymptotic methods based on normal convergence of estimators and consistent estimation of their limiting variances under heterogeneity and dependence in observations; for example, for inference approaches based on heteroskedasticity and autocorrelation consistent - HAC - and clustered standard errors, especially with data with pronounced autocorrelation, dependence and heterogeneity.

Cowell and Flachaire, 2007, Davidson and Flachaire, 2007). Bootstrap methods are also known to fail in heavy-tailed infinite variance settings (see the discussion in Section 5 in Davidson and Flachaire, 2007, and references therein).

The difficulties in inference related to inequality are discussed in detail in Dufour et al. (2019, 2020), where the paucity of reliable methods are highlighted. Difficulties apply both to the problem of one-sample inference based on a single estimate of an inequality index, as well as the two-sample problem of inference on the difference between or equality of inequality indices in two populations. The latter problem is much more challenging than the former (see also Ibragimov and Müller, 2016, for discussion and results on robust inference on the difference between or equality of two parameters under heterogeneity and dependence).

Dufour et al. (2019) propose permutation tests for the hypothesis of equality of an inequality index across two populations using independent samples. These tests outperform other asymptotic and bootstrap methods currently available (see also Canay et al., 2017, for permutation tests of equality of two general parameters of interest under heterogeneity and clustered dependence), but do not provide a way of drawing inferences on any specified difference (including zero) in inequality between the populations, nor of constructing a confidence interval for the difference in the inequality index. In this regard Dufour et al. (2020) propose Fieller-type methods for inference on the generalized entropy (GE) class of inequality indices for any, including possibly non-zero, difference in inequality between two populations. This approach can be used with independent samples of i.i.d. observations with possibly unequal sizes and equal-sized samples of i.i.d. observations with arbitrary dependence between the samples, but is computationally demanding.

This paper focuses on applications of the recently developed $t-$statistic approach (see Ibragimov and Müller, 2010, 2016, and also Ch. 3 in Ibragimov et al., 2015) for robust inference on income and wealth inequality measured using a class of inequality indices, when the data are heterogeneous, heavy-tailed or correlated in some unknown fashion. A robust large sample test on equality or a non-zero difference in inequality (between two regions, for example) can be conducted quite easily as follows: the data in the two samples are partitioned into fixed numbers $q_I, q_Y \geq 2$ (e.g., $q_I = q_Y = 2, 4,$ or 8) of groups, the considered inequality index is estimated for each group, and inference is based on a standard two-sample $t-$test with the resulting q_I, q_Y group estimators (see the next section). As follows from the results in Ibragimov and Müller (2010, 2016), the above $t-$statistic based robust inference approach results in valid inference under general conditions that the group estimators of the parameter of interest, the inequality index, are asymptotically independent and Gaussian, with possibly different variances.[5][6]

(See, among others, Andrews, 1991, den Haan and Levin, 1997, the discussion in Phillips, 2005, Ibragimov and Müller, 2010, 2016, Canay et al., 2017, Esarey and Menger, 2019, and references therein).

[5]Asymptotic validity of $t-$statistic robust inference approach in Ibragimov and Müller (2010, 2016) continues to hold in the case where group estimators of the parameter of interest weakly converge, at an arbitrary rate, to scale mixtures of normals, e.g., to symmetric stable distributions as in heavy-tailed infinite variance settings.

[6]As discussed in Ibragimov and Müller (2010) and Ch. 3 in Ibragimov et al. (2015) (see also Section 2), asymptotic Gaussianity of group estimators of the parameters of interest typically follows from the same reasoning and holds under the same conditions as the asymptotic Gaussianity of their full-sample estimators. Asymptotic normality holds for estimators of Theil and Gini indices in the case of power law income distributions with the tail index $\zeta > 2$ and

The $t-$statistic approach to robust inference on inequality complements and compares favorably with other inference methods available in the literature, including computationally expensive bootstrap procedures and permutation-based inference methods. Importantly, the approach can be used for constructing confidence intervals for any possibly non-zero difference in inequality between populations as measured by the chosen index, notwithstanding the problems of heterogeneity, heavy-tailedness, and possible dependence in the data.

In comparison to other methods in the literature including those in Dufour et al. (2019, 2020), the approach to inference on inequality that is proposed in this paper has a wider range of applicability. It can be used when observations on income or wealth levels in each of the samples considered are ***dependent*** among themselves - for example, due to spatial or clustered dependence (see Conley, 1999, and Bhattacharya, 2007, for a review of settings and methods of inference under spatial and clustered dependence, including complex stratified and clustered household surveys), common shocks (see Andrews, 2005, and Hwang, 2021, for a review of and inference using data with common shock dependence), or, in the case of time series or panel data on income or wealth levels, due to autocorrelation and dependence in observations over time. Further, for inference on the difference in inequality between two populations using two samples of possibly dependent observations, the $t-$statistic inference approach may be used under ***arbitrary*** dependence ***between*** the samples as well as with ***unequal*** sample sizes.

The paper is organized as follows. Section 2 describes the $t-$statistic approach to robust inference on inequality and discusses the conditions for its validity. Section 3 provides numerical results on the finite sample performance of this approach and its comparison with other inference methods in the literature. In particular, Section 3.2 provides the results on finite sample performance of $t-$statistic and other approaches in the case of testing equality of inequality between two populations and inference on the difference in inequality between them, with the main focus on the important and empirically relevant case of samples with different (income or wealth) distributions, and the case of samples with identical distributions discussed in Appendix C. Section 4 presents empirical applications of the robust $t-$statistic approach in the analysis of income inequality in Russia and comparisons of inequality across Russian regions. Section 5 offers some concluding remarks and suggestions for future research. Appendix A provides tables on the numerical and empirical results in the paper. Appendix B provides a review of the definitions and asymptotic properties of Gini and Theil inequality indices dealt with in the paper and a discussion of the asymptotics for their sample analogs. Appendix C discusses finite-sample performance of the $t-$statistic and other inference approaches in the illustrative case of samples with identical (income or wealth) distributions. For illustrative purposes, online Appendix D discusses $t-$statistic inference approach and its implementation for the one-sample case.

---

finite second moments (see the discussion in Appendix B).

2 Methodology: Robust *t*-statistic approach to inference on inequality measures

Our focus is on inference on inequality as measured using a set of commonly employed inequality indices, applying the computationally economic *t*-statistic based approach to robust inference recently developed in Ibragimov and Müller (2010, 2016). Ibragimov and Müller (2010) provides an approach to robust inference on an arbitrary single parameter of interest; Ibragimov and Müller (2016) provides an approach to robust testing of equality as well as to inference on the difference between two populations, in terms of an arbitrary parameter of interest.[7]

We focus on the important empirically relevant and difficult (see the discussion in the introduction) problem of two-sample inference on income or wealth inequality. [8]

In the two-sample problem, we consider inference on the difference $d = \mathcal{L}^I - \mathcal{L}^Y$ between the values of the inequality index $\mathcal{L}$ in two populations using large samples $I_1, I_2, ..., I_{N_I}$ and $Y_1, Y_2, ..., Y_{N_Y}$ on income or wealth levels in the populations. With two independent samples, inference is based on the two-sample $t-$statistic in group estimators of the inequality index $\mathcal{L}$ for the two samples.

2.1 Two independent samples: Inference using the two-sample $t-$statistic in group estimators

Throughout the paper, we denote by T_k the r.v. that has a Student-t distribution with $k \geq 1$ degrees of freedom. Further, for $q \geq 2$ and $0 < \alpha < 1$, we use $cv_{q,\alpha}$ to denote the $(1 - \alpha/2)-$quantile of the Student-t distribution with $q - 1$ degrees of freedom: $P(|T_{q-1}| > t_\alpha)= \alpha$.

Following the $t-$statistic approach to robust inference on two parameters in Ibragimov and Müller (2016), each of two independent samples ($I_1, I_2, ..., I_{N_I}$ and $Y_1, Y_2, ..., Y_{N_Y}$) is partitioned into fixed numbers $q_I, q_Y \geq 2$ (e.g., $q_I, q_Y = 2, 4, 8$) groups, respectively, and the income inequality index $\mathcal{L}$ is estimated using the data for each of the groups in the two samples. This thus results in $q_I + q_Y$ group empirical income inequality measures $\widehat{\mathcal{L}}_1^I$, ..., $\widehat{\mathcal{L}}_{q_I}^I$, and $\widehat{\mathcal{L}}_1^Y$, ..., $\widehat{\mathcal{L}}_{q_Y}^Y$. The robust test of the null hypothesis $H_0 : \mathcal{L}^I - \mathcal{L}^Y = d_0$ (when $d_0 = 0$, the test is for the hypothesis $H_0 : \mathcal{L}^I = \mathcal{L}^Y$) against the two-sided alternative $H_a : \mathcal{L}^I - \mathcal{L}^Y \neq d_0$ (resp., with $d_0 = 0$, against the two-sample alternative $H_a : \mathcal{L}_1 \neq \mathcal{L}_2$) is based on the usual two-sample $t-$statistic $\tilde{t}_{\mathcal{L}}$ in the $q_I + q_Y$ group inequality estimates $\widehat{\mathcal{L}}_j^I$, $\widehat{\mathcal{L}}_k^Y$, $j = 1, ..., q_I$, $k = 1, ..., q_Y$:

$$\tilde{t}_{\mathcal{L}} = \frac{\overline{\widehat{\mathcal{L}}^I} - \overline{\widehat{\mathcal{L}}^Y} - d_0}{\sqrt{s^2_{\widehat{\mathcal{L}}^I}/q_I + s^2_{\widehat{\mathcal{L}}^Y}/q_Y}} \tag{2}$$

[7]We refer to, among others, Section 13.F in Cowell and Flachaire (2007), Davidson and Flachaire (2007), Section 13. F in Marshall et al. (2011), Ibragimov and Ibragimov (2018), Dufour et al. (2019) and Dufour et al. (2020), and Appendix B for definitions of the most widely used inequality measures, including Gini, Generalized Entropy, and Theil indices, and their values for different income distributions, including empirically relevant heavy-tailed Pareto, double Pareto, and Singh-Maddala distributions that follow power laws (1) (see the next section).

[8]Implementation of the $t-$statistic approach in the illustrative one-sample case is discussed in online Appendix D and in Ibragimov et al., 2013).

with

$$\overline{\widehat{\mathcal{L}}^I} = \frac{1}{q_I}\sum_{j=1}^{q_I}\widehat{\mathcal{L}}_j^I, \overline{\widehat{\mathcal{L}}^Y} = \frac{1}{q_Y}\sum_{k=1}^{q_Y}\widehat{\mathcal{L}}_k^Y,$$

$$s^2_{\widehat{\mathcal{L}}^I} = \frac{1}{q_I-1}\sum_{j=1}^{q_I}\left(\widehat{\mathcal{L}}_j^I - \overline{\widehat{\mathcal{L}}^I}\right)^2, s^2_{\widehat{\mathcal{L}}^Y} = \frac{1}{q_Y-1}\sum_{k=1}^{q_Y}\left(\widehat{\mathcal{L}}_k^Y - \overline{\widehat{\mathcal{L}}^Y}\right)^2.$$

For the numbers of groups $q_I, q_Y \leq 14$ the above null hypothesis $H_0 : \mathcal{L}^I - \mathcal{L}^Y = d_0$ is rejected in favor of the alternative $H_a : \mathcal{L}^I - \mathcal{L}^Y \neq d_0$ at level $\alpha \in \{0.001, 0.002, ..., 0.099, 0.10\}$ (which includes the usual significance levels $\alpha = 0.01, 0.05$ and 0.1 if the absolute value $|\tilde{t}_{\mathcal{L}}|$ of the two-sample $t-$statistic in group inequality estimators $\widehat{\mathcal{L}}_j^I$, $\widehat{\mathcal{L}}_k^Y$, $j = 1, ..., q_I$, $k = 1, ..., q_Y$, exceeds the $(1-\alpha/2)-$quantile of the standard Student-t distribution with $q-1$ degrees of freedom, where $q = \min(q_I, q_Y) : |\tilde{t}_{\mathcal{L}}| > cv_{q,\alpha} = cv_{\min(q_I,q_Y),\alpha}$.[9] One-sided tests are conducted in a similar way (see also Appendix D for illustrations in the one-sample case).

For $\alpha = 0.01, 0.05, 0.1$, with the number of groups $q_I, q_Y \leq 14$, denoting $\min(q_I, q_Y) = q$, a confidence interval for the difference $d_0 = \mathcal{L}^I - \mathcal{L}^Y$ between the values of the inequality index $\mathcal{L}$ in two populations with asymptotic coverage of at least $1-\alpha$ may be constructed readily; for example, the 95% confidence interval for $\mathcal{L}$ is given by $\widehat{\mathcal{L}}^I - \widehat{\mathcal{L}}^Y \pm cv_{q,0.05}\sqrt{s^2_{\widehat{\mathcal{L}}^I}/q_I + s^2_{\widehat{\mathcal{L}}^Y}/q_Y}$, where $cv_{q,0.05}$ is the 0.975-quantile of the Student-t distribution with $\min(q_I, q_Y)-1$ degrees of freedom: $P(\left|T_{\min(q_I,q_Y)-1}\right| > cv_{q,0.05})$=0.05.

As follows from Ibragimov and Müller (2016), the two-sample $t-$statistic approach is asymptotically valid under the assumption that the group empirical income inequality measures $\widehat{\mathcal{L}}_j^I$, $j = 1, ..., q_I$, $\widehat{\mathcal{L}}_k^Y$, $k = 1, ..., q_Y$, are asymptotically independent, unbiased and Gaussian, even if of different variances.

More generally, asymptotic validity of t-statistic based inference also holds under convergence of the group estimators to conditionally normal r.v.s which are possibly unconditionally dependent through their second moments or have a common shock-type dependence (see Andrews, 2005, for inference methods under common shock dependence structures, and Hwang, 2021, for applications of $t-$statistic robust inference approach in such settings).[10] This implies that the approach can be applied to inference on $\mathcal{L}$ in the presence of extremes and outliers in observations generated by heavy-tailedness with infinite variances, as well as dependence structures that include models with multiplicative common shocks (see Ibragimov, 2007, 2009). The inference approach does not require the estimation of limiting variances of estimators of interest, in contrast to inference methods based on consistent (e.g., HAC or clustered) standard errors.[11]

[9] As follows from the analysis in Ibragimov and Müller (2016), these tests may also be used for all $q_I, q_Y \leq 50$ if $\alpha \in \{0.001, 0.002, ..., 0.083\}$, e.g., for the usual critical values $\alpha = 0.01, 0.05$.

[10] Justification of asymptotic validity of the robust $t-$statistic inference approach in Ibragimov and Müller (2010) is based on a small sample result in Bakirov and Székely (2006) that implies validity of the standard $t-$test on the mean under independent heterogeneous normal observations. Justification of asymptotic validity of the approach in inference on equality of two parameters in Ibragimov and Müller (2016) is based on the analogues of the above small sample result for two-sample $t-$tests and Behrens-Fisher problem obtained therein.

[11] The numerical analysis in Ibragimov and Müller (2010, 2016) and Section 3 in Ibragimov et al. (2015) indicates favorable finite sample performance of the approach in inference on models with time series, panel, clustered and

The conditions for asymptotic validity of $t-$statistic robust inference approach are discussed in further detail in Section 2.3 below and in Appendix D.

2.2 Possibly dependent samples: Inference using one-sample $t-$statistic in differences of group estimators

Let us now consider the problem of inference on the difference $d = \mathcal{L}^I - \mathcal{L}^Y$ between the two populations using income or wealth samples $I_1, I_2, ..., I_{N_I}$ and $Y_1, Y_2, ..., Y_{N_Y}$ of ***possibly unequal sizes*** N_I, N_Y, that may exhibit an ***arbitrary dependence between them***. Suppose that the samples are divided into an equal number of groups $q_I = q_Y = q \geq 2$ (e.g., $q_I, q_Y = 2, 4$ or 8), and estimates of the inequality index $\mathcal{L}$ are calculated using the data for each of the $2q$ groups. This produces group level income inequality estimates $\widehat{\mathcal{L}}_1^I$, ..., $\widehat{\mathcal{L}}_q^I$, and $\widehat{\mathcal{L}}_1^Y$, ..., $\widehat{\mathcal{L}}_q^Y$.

As follows from Ibragimov and Müller (2010) (see also Appendix D on the one-sample inference), the robust test of the null hypothesis $H_0 : \mathcal{L}^I - \mathcal{L}^Y = d_0$ (with $d_0 = 0$, the test is of $H_0 : \mathcal{L}^I = \mathcal{L}^Y$) against the two-sided alternative $H_a : \mathcal{L}^I - \mathcal{L}^Y \neq d_0$ may be based on the one-sample $t-$statistic $\tilde{\tilde{t}}_{\mathcal{L}}$ in the q differences $\widehat{\mathcal{L}}_j^I - \widehat{\mathcal{L}}_j^Y$, $j = 1, ..., q$, of the group empirical inequality estimates:

$$\tilde{\tilde{t}}_{\mathcal{L}} = \sqrt{q}\frac{\overline{\widehat{\mathcal{L}}^I} - \overline{\widehat{\mathcal{L}}^Y} - d_0}{s_{\widehat{\mathcal{L}}^{I-Y}}} \tag{3}$$

with

$$\overline{\widehat{\mathcal{L}}^I} = \frac{1}{q}\sum_{j=1}^{q} \widehat{\mathcal{L}}_j^I, \overline{\widehat{\mathcal{L}}^Y} = \frac{1}{q}\sum_{j=1}^{q} \widehat{\mathcal{L}}_j^Y,$$

$$s^2_{\widehat{\mathcal{L}}^{I-Y}} = \frac{1}{q-1}\sum_{j=1}^{q} \Big((\widehat{\mathcal{L}}_j^I - \widehat{\mathcal{L}}_j^Y) - (\overline{\widehat{\mathcal{L}}^I} - \overline{\widehat{\mathcal{L}}^Y}) \Big)^2.$$

As in the case of $t-$statistic based inference on one parameter in Ibragimov and Müller (2010) and Appendix D, for any $\alpha \leq 0.083$ (any $\alpha \leq 0.1$ for $2 \leq q \leq 14$), the null hypothesis $H_0 : \mathcal{L}^I - \mathcal{L}^Y = d_0$ is rejected in favor of the two-sided alternative $H_a : \mathcal{L}^I - \mathcal{L}^Y \neq d_0$ at level α if the absolute value $|\tilde{\tilde{t}}_{\mathcal{L}}|$ of the $t-$statistic in the differences $\widehat{\mathcal{L}}_j^I - \widehat{\mathcal{L}}_j^Y$, $j = 1, ..., q$ of group inequality estimates exceeds the $(1 - \alpha/2)-$quantile of the standard Student-t distribution with $q - 1$ degrees of freedom: $|\tilde{\tilde{t}}_{\mathcal{L}}| > cv_{q,\alpha}$. Further, as in the case of $t-$statistic inference on a single inequality measure in Appendix D, the $p-$values of the above tests can be calculated in the case of an arbitrary number $q = q_I = q_Y$ of groups thus enabling conducting robust tests of an arbitrary level.

For all $\alpha \leq 0.083$ (and all $\alpha \leq 0.1$ for $2 \leq q \leq 14$), a confidence interval for the difference $d_0 = \mathcal{L}^I - \mathcal{L}^Y$ between the values of the inequality index $\mathcal{L}$ in two populations with asymptotic coverage of at least $1 - \alpha$ may be constructed as $\widehat{\mathcal{L}}_j^I - \widehat{\mathcal{L}}_j^Y \pm cv_{q,\alpha} s_{\widehat{\mathcal{L}}^{I-Y}}$.

As discussed in Ibragimov and Müller (2010), Section 2.1 and Appendix D, the $t-$statistic based approach to robust inference based on (3) is asymptotically valid under the assumption that the the

spatially correlated data. See also Esarey and Menger (2019) for software (STATA and R) implementations and detailed numerical analysis of finite sample performances of different inference procedures including $t-$statistic and related approaches, under small number of clusters of dependent data.

bivariate vectors of group empirical inequality estimates $(\widehat{\mathcal{L}}_j^I,\ \widehat{\mathcal{L}}_j^Y)$, $j = 1, ..., q$, are asymptotically independent across j, with components that are asymptotically unbiased and Gaussian, with possibly different variances (across j). At the same time, for each $j = 1, ..., q$, the components of the vector of group estimators $\widehat{\mathcal{L}}_j^I$ and $\widehat{\mathcal{L}}_j^Y$ can exhibit an arbitrary dependence among themselves. Indeed, asymptotic independence of the vectors $(\widehat{\mathcal{L}}_j^I,\ \widehat{\mathcal{L}}_j^Y)$, $j = 1, ..., q$, of group estimators implies asymptotic independence of the differences $\widehat{\mathcal{L}}_j^I - \widehat{\mathcal{L}}_j^Y$, $j = 1, ..., q$, used in calculation of $t-$statistic (3), that is required for asymptotic validity of the inference approach based on the $t-$statistic, according to the results in Ibragimov and Müller (2010).

2.3 Asymptotic Gaussianity and asymptotic independence of group estimators and the choice of groups

As in the one-sample problem in Appendix D, asymptotic Gaussianity (or convergence to stable distributions under heavy-tailedness with infinite second moments) of group estimators of the inequality index $\mathcal{L}$ in the two samples - the group empirical inequality estimators $\hat{\mathcal{L}}_j^I$, $\hat{\mathcal{L}}_k^Y$ - typically holds (under the same conditions) as long as it holds for the full-sample estimators $\hat{\mathcal{L}}^I$, $\hat{\mathcal{L}}^Y$ of $\mathcal{L}$ calculated using all the observations in the two samples. In particular, asymptotic normality holds for estimators of Theil and Gini indices, in the case of power law income distributions (1) with tail indices $\zeta > 2$ and finite second moments (see Appendix B).

As discussed in Appendix D, with reference to inference on the difference between inequality in two regions in a country using i.i.d. data from household income surveys, groups may be formed by partitioning each of the two region-specific random samples. Namely, given the random samples $I_1, I_2, ..., I_{N_I}$, $Y_1, Y_2, ..., Y_{N_Y}$ of (i.i.d.) income levels, the q_I, q_Y groups may be formed as $\{I_k, (i-1)N_I/q_I < k \leq iN_I/q_I\}$, $\{Y_l, (j-1)N_Y/q_Y < l \leq jN_Y/q_Y\}$, $i = 1, ..., q_I$, $j = 1, ..., q_Y$. This is illustrated in the empirical application in Section 4. Asymptotic unbiasedness and independence of group inequality estimators hold due to i.i.d.ness of data in the samples.[12]

In applications of the two-sample t-statistic based approach, the appropriate choice of the numbers q_I and q_Y of groups is important. The numerical results in Section 3.2 indicate that, for different heavy-tailed distributions considered, the choice of the number of groups $q_I = q_Y = q = 4, 8, 12, 16$ leads to attractive finite sample performance of robust two-sample tests based on $t-$statistics $\tilde{t}_{\mathcal{L}}$ in (2) and $\tilde{\tilde{t}}_{\mathcal{L}}$ in (3).[13]

[12] Various methods for forming groups can be employed. Additionally, it would be valuable to explore random splits or all possible splits and use inference procedures based on metrics such as the median or average of the t-statistics calculated from the corresponding group estimates of parameters of interest. We thank an anonymous referee for suggesting the use of random splits as the basis for groups. The use of random splitting does not affect the validity of the asymptotic properties of the group estimators. Following a suggestion by one of the authors, Dagayev and Stoyan (2020) recently used random sample splits in their empirical application. They constructed t-statistics for group estimates of the parameters under analysis and based the inference on quantiles, including the median, from the empirical distribution of the t-statistics values for the splits considered.

[13] These conclusions align with the numerical results presented in Ibragimov and Müller (2010), which indicate that for many dependence and heterogeneity settings considered in the literature and typically encountered in practice (e.g.,

More generally, and naturally, the finite sample performance of the $t-$statistic robust inference approach with different numbers q_I, q_Y of groups and the choice of the optimal values of q_I, q_Y on its basis depend on distributional properties of the populations considered, including the degrees of their heavy-tailedness, and the sizes N_I, N_Y of the samples used in inference. The simplest way to choose the numbers of groups in the case of distributions that are not very different from each other is to have q_I/q_Y (approximately) equal to N_I/N_Y so that the sizes of all the groups considered are (approximately) the same. If the population distributions have similar tail indices, then in the case of inference on Gini measures, q_I and q_Y may be taken to be equal. In general, the size of the groups in the sample from a more heavy-tailed distribution should be larger than the size of the groups from a less heavy-tailed distribution. Thus in the case of equally sized samples, one should take the number of groups in the more heavy-tailed sample to be less than the number of groups in the less heavy-tailed sample.

3 Finite sample performance

In this section, we present numerical results on finite sample properties of the $t-$statistic approach, as well as of the bootstrap and permutation approaches to inference on inequality indices. Results are provided for inference on the Theil index and the Gini coefficient, as in Cowell and Flachaire (2007) and Dufour et al. (2019). We begin by addressing the finite sample approximations to sampling distributions of full-sample estimators in Section 3.1. In Section 3.2, we present results for the two-sample problem of inference on the difference between two inequality indices.

As in Dufour et al. (2019), our numerical analyses are based on simulations from the Singh-Maddala (S-M) family of distributions, which is known to provide a good fit to income distributions in various countries (see the review and the discussion in Section 6.1.6 in Kleiber and Kotz, 2003, Cowell and Flachaire, 2007, Davidson and Flachaire, 2007, Dufour et al., 2019, and references therein). We adopt the same parameter values used in these studies and largely follow the notation in Cowell and Flachaire (2007), Davidson and Flachaire (2007) and Dufour et al. (2019).

The cdf of an S-M distribution with the scale parameter $b > 0$ and the shape parameters $a, c > 0$ is given by

$$F(x) = 1 - \left[1 + \left(\frac{x}{b}\right)^a\right]^{-c}, x > 0, \tag{4}$$

As in Dufour et al. (2019), the S-M distribution with parameters $a, b, c > 0$ is denoted by $SM(a, b, c)$ in what follows. It is easy to see that the cdf $F(x)$ of $SM(a, b, c)$ satisfies $F(x) \sim c\left(\frac{x}{b}\right)^a$ as

time series, panel, clustered, and spatially correlated data), selecting the number of groups $q = 8$ or $q = 16$ leads to robust one-sample t-statistic tests with attractive finite sample performance. It is important to emphasize that the asymptotic efficiency results for t-statistic based robust inference in Ibragimov and Müller (2010) imply that it is not feasible to use data-dependent methods to determine the optimal number of groups q when the only assumption imposed on the data-generating process is asymptotic normality and asymptotic independence of group estimators of the parameter of interest. That being so, it is both interesting and important for future research to explore if data-driven optimal values can be derived for the number of groups in robust t-statistic inference, conditional on additional assumptions such as the degree of heavy-tailedness in the populations.

$x \to 0$, and $1 - F(x) \sim \left(\frac{x}{b}\right)^{-ac}$ as $x \to \infty$. [14] Therefore the S-M distribution has the (double) power law or (double) Pareto behavior in both tails (see Toda, 2012, for the analysis of double Pareto and related distributions for income). For large $x > 0$ the r.v.s (income or wealth levels) $X > 0$ with the S-M distribution $SM(a, b, c)$ follows power law (1) with the tail index $\zeta = ac$.

Following Cowell and Flachaire (2007), Davidson and Flachaire (2007) and Dufour et al. (2019), we use parameter values $a_0 = 2.8$, $b_0 = 100^{-1/2.8}$, $c_0 = 1.7$ for the S-M distribution, with the corresponding tail index $\zeta = a_0 c_0 = 4.76$, as a benchmark.[15] For these values, the Theil index equals 0.140, and the Gini coefficient equals 0.289 (see Dufour et al., 2019). Cowell and Flachaire (2007) and Davidson and Flachaire (2007) demonstrate the poor finite-sample performance of asymptotic and bootstrap inference approaches using these parameter values.

Further, as in Dufour et al. (2019), we also consider several other S-M distributions $SM(a, b_0, c)$ with the above scale parameter $b_0 = 100^{-1/2.8}$ for which the Theil inequality index and the Gini index are the same as in the case of $SM(a_0, b_0, c_0)$.[16] [17]

We note that the tail indices $\zeta = 2.78, 2.59, 2.9$ for the S-M distributions lie in the interval $(1.5, 3)$ as is typically the case for real-world income distributions, as discussed above. We also consider heavier-tailed distributions $SM(a, b_0, c)$ with $(a, c) = (2, 1.1)$, $(2, 0.7)$ and $b_0 = 100^{-1/2.8}$. The corresponding tail indices ζ equal $2.2, 1.4$.

3.1 Finite sample approximations to sampling distributions of full-sample estimators

In this section, $\mathcal{L}_0 = \mathcal{L}(F)$ denotes the true value of the inequality index $\mathcal{L}$ (e.g., the Theil index or Gini coefficient) for the population; $\hat{\mathcal{L}} = \hat{\mathcal{L}}(I_1, ..., I_N)$ denotes the full-sample estimator of $\mathcal{L}$ calculated using a sample of observations $I_1, ..., I_N$ from the population; and $\hat{\mathcal{L}}_j$, $j = 1, ..., q$, denote the group estimators of $\mathcal{L}$.

Asymptotic approaches to inference on inequality index $\mathcal{L}$ are based on normal approximations to sampling distributions of full-sample estimators $\hat{\mathcal{L}}$ of the index – more precisely, on standard normal approximations to sampling distributions of (full-sample) $t-$statistics $S_{\hat{\mathcal{L}}} = (\hat{\mathcal{L}} - \mathcal{L}_0)/s.e._{\hat{\mathcal{L}}}$, where $s.e._{\mathcal{L}}$ denotes the usual consistent standard error of $\hat{\mathcal{L}}$ (see the formulae for inequality index estimators and their standard errors in Cowell and Flachaire, 2007, Davidson and Flachaire, 2007, and Dufour et al., 2019). As discussed in Section 2, the validity of $t-$statistic based robust inference

[14] As usual, we write $f(x) \sim g(x)$ as $x \to x_0$ or $x \to \infty$ for two positive functions $f(x)$ and $g(x)$ if $f(x)/g(x) \to 1$ as $x \to x_0$ or $x \to \infty$.

[15] This distribution is reported to provide a good fit to income distribution of German households, up to a scale factor.

[16] In simulations involving the Theil index, we allow the parameters (a, c) of the distribution $SM(a, b_0, c)$ to equal $(2.5, 2.502)$, $(2.6, 2.1497)$, $(2.7, 1.894)$, $(2.8, 1.7)$, $(3.0, 1.422)$, $(3.2, 1.232)$, $(3.4, 1.092)$, $(3.8, 0.898)$, $(4.8, 0.637)$ and $(5.8, 0.4996)$. The corresponding tail indices ζ are $\zeta = 6.26, 5.59, 5.11, 4.76, 4.27, 3.94, 3.71, 3.41, 3.06, 2.9$.

[17] In simulations involving the Gini index, we allow, again as in Dufour et al. (2019), the parameters (a, c) of the distribution $SM(a, b_0, c)$ to equal (2.5,2.640), (2.6,2.218), (2.7,1.921), (2.8,1.7), (3.0,1.392), (3.2,1.187), (3.4,1.039), (3.8,0.838), (4.8,0.578) and (5.8,0.447). The corresponding tail indices ζ are ζ =6.6, 5.77, 5.19, 4.76, 4.18, 3.80, 3.53, 3.19, 2.78, 2.59.

approach requires weak convergence of group estimators $\hat{\mathcal{L}}_j$, $j = 1, ..., q$, of the inequality index $\mathcal{L}$ to possibly heterogeneous Gaussian distributions (without any Studentization/normalization of the group estimators by their standard errors, in contrast to the $t-$statistics ($S_{\hat{\mathcal{L}}}$) calculated using the full-sample estimators). Asymptotic normality of group estimators $\hat{\mathcal{L}}_j$ holds under the same conditions as in the case of the full-sample estimators $\hat{\mathcal{L}}$ (see the discussion in the introduction, Section 2 and Appendix D).

We begin with an assessment of finite-sample distributions of (full-sample) inequality estimators $\hat{\mathcal{L}}$ and (full-sample) $t-$statistics $S_{\hat{\mathcal{L}}}$ calculated using them. We focus on the closeness of the above finite-sample distributions to the Gaussian, and on comparisons of finite-sample distributions of the (full-sample) $t-$statistics $S_{\hat{\mathcal{L}}}$ with those of the centered inequality index estimators normalized by their true standard deviations, i.e., of the statistics $Z_{\hat{\mathcal{L}}} = (\hat{\mathcal{L}} - \mathcal{L}_0)/\sigma_{\hat{\mathcal{L}}}$, where $\sigma^2_{\hat{\mathcal{L}}} = Var(\hat{\mathcal{L}})$. The true values of the standard deviations $\sigma_{\hat{\mathcal{L}}}$ for the populations and sample sizes considered are obtained using direct simulations.

Figures 1-3 provide kernel estimates of densities of the finite-sample distributions of the statistics $Z_{\hat{\mathcal{L}}}$ and $S_{\hat{\mathcal{L}}}$ for different population distributions and sample sizes.[18] Figures 1 and 2 provide kernel densities of the statistics $Z_{\hat{\mathcal{L}}}$ (sample sizes $N = 50, 100, 1000$) and $S_{\hat{\mathcal{L}}}$ (sample size $N = 100$)[19] for, respectively, the estimates of Theil and Gini inequality indices for samples drawn from the S-M distributions $SM(a_0, b_0, c_0)$ with the parameters $a_0 = 2.8$, $b_0 = 100^{-1/2.8}$, $c_0 = 1.7$ and the corresponding tail index $\zeta = 4.76$, mentioned earlier.

In the case of the Theil index in Figure 1, we observe some non-Gaussianity in the distribution of the statistics $Z_{\hat{\mathcal{L}}}$ and $S_{\hat{\mathcal{L}}}$ for small and moderate-sized samples. In addition, the density of the $t-$statistic $S_{\hat{\mathcal{L}}}$ for Theil index is considerably (left) skewed in comparison to the densities of the statistic $Z_{\hat{\mathcal{L}}}$. In the case of the Gini coefficient in Figure 2, the distribution of the statistic $Z_{\hat{\mathcal{L}}}$ is very close to the standard normal even in small samples. In contrast, the distribution of the $t-$statistic S is again skewed towards the left.

For S-M distributions with heavier tails, as with parameters $(a, c) = (5.8, 0.447)$ and correspondingly, tail index $\zeta = 2.59$ in Figure 3, the finite sample distributions of the statistics $Z_{\hat{\mathcal{L}}}$ and $S_{\hat{\mathcal{L}}}$ for the Gini coefficient become more skewed (the same is observed for the Theil index; the results are omitted for brevity and available on request). Skewness is especially pronounced in the case of small samples and the $t-$statistic S.

Overall, according to Figures 1-3, the normal approximation appears to perform better for finite-sample distributions of the statistic $Z_{\hat{\mathcal{L}}}$ as compared to those of the full-sample $t-$statistic $S_{\hat{\mathcal{L}}}$ used in asymptotic tests and inference. Note that the group estimators used in the $t-$statistic based robust inference approach are just scaled versions of the statistics $Z_{\hat{\mathcal{L}}}$ calculated using observations in the groups. Therefore, the above comparisons are to be expected to translate into better finite-sample performance of the $t-$statistic approach as compared to the asymptotic approaches, provided that the number of observations in each of the groups in the $t-$statistic approach is sufficiently large, e.g., greater than 100 (usually, it is the case in empirical applications with the number of groups $q = 4, 8$). For better size control, the number of groups, q, should be fewer if the total sample size N is not very

[18]The number of replications in all simulation experiments is equal to 100,000.

[19]Qualitatively similar results for other sample sizes N are omitted for brevity and available on request.

large.

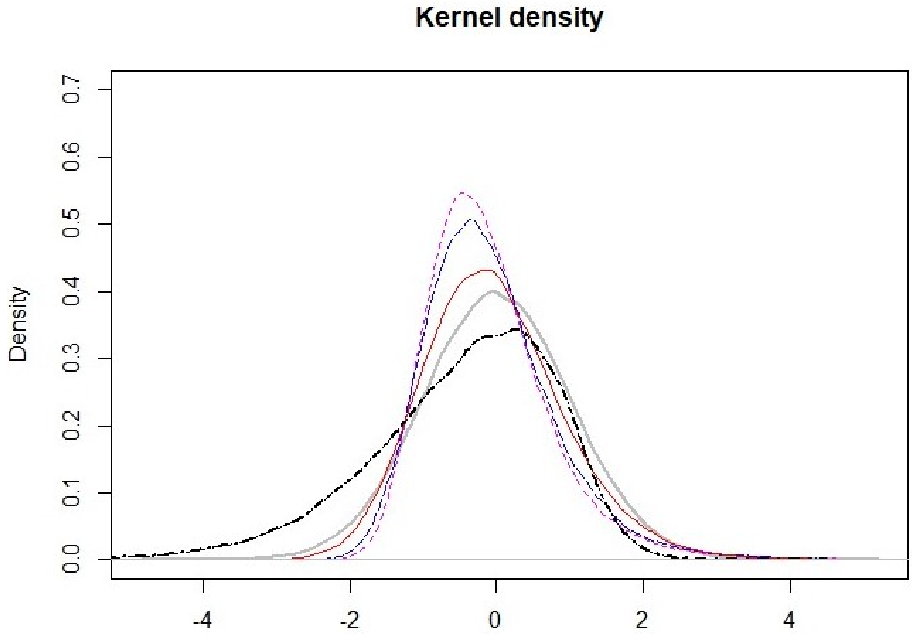

Figure 1: Kernel density functions for the statistics $Z_{\hat{\mathcal{L}}}$ and $S_{\hat{\mathcal{L}}}$ for the Theil index: S-M distribution $SM(a_0, b_0, c_0)$ with $(a_0, c_0) = (2.8, 1.7)$ and $\zeta = 4.76$. Gaussian density: ; Statistic $Z_{\hat{\mathcal{L}}}$, $N = 50$: ; Statistic $Z_{\hat{\mathcal{L}}}$, $N = 100$: ; Statistic $Z_{\hat{\mathcal{L}}}$, $N = 1000$: ; Statistic $S_{\hat{\mathcal{L}}}$, $N = 100$:

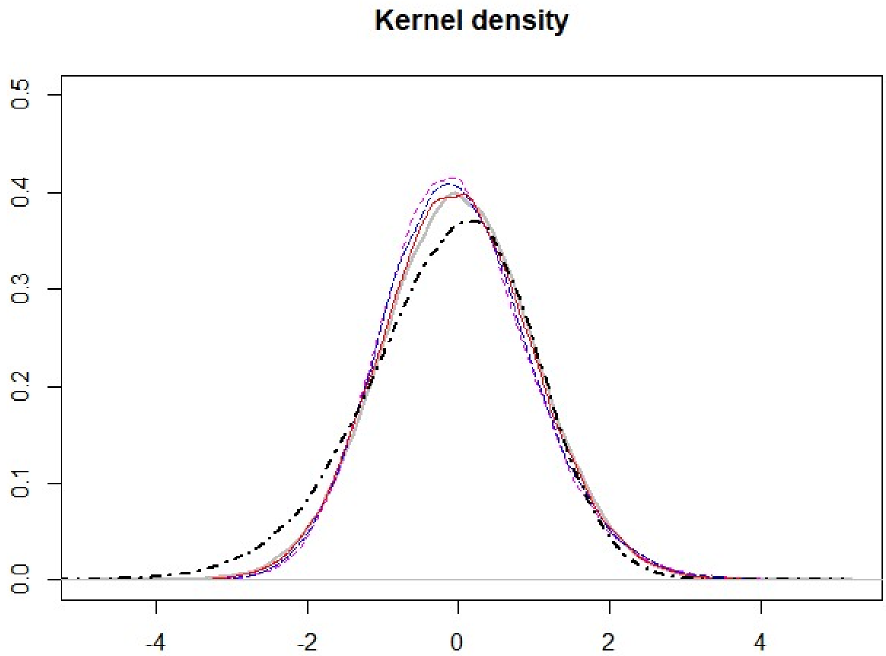

Figure 2: Kernel density functions for the statistics $Z_{\hat{\mathcal{L}}}$ and $S_{\hat{\mathcal{L}}}$ for the Gini index: S-M distribution $SM(a, b_0, c_0)$ with $(a_0, c_0) = (2.8, 1.7)$ and $\zeta = 4.76$. Gaussian density: ; Statistic $Z_{\hat{\mathcal{L}}}$, $N = 50$: ; Statistic $Z_{\hat{\mathcal{L}}}$, $N = 100$: ; Statistic $Z_{\hat{\mathcal{L}}}$, $N = 1000$: ; Statistic $S_{\hat{\mathcal{L}}}$, $N = 100$:

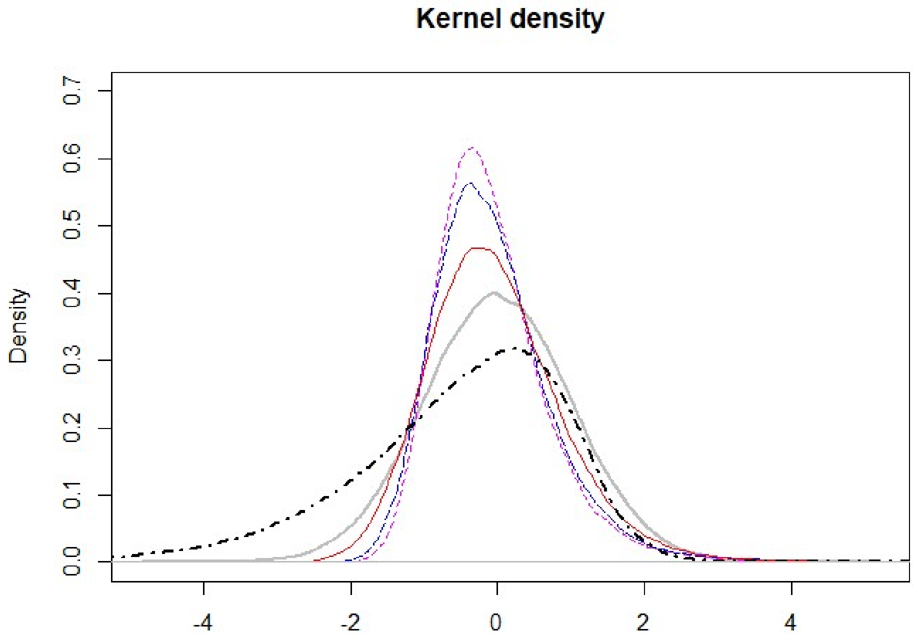

Figure 3: Kernel density functions for the statistics $Z_{\hat{\mathcal{L}}}$ and $S_{\hat{\mathcal{L}}}$ for the Gini index: S-M distribution $SM(a, b_0, c)$ with $(a, c) = (5.8, 0.4473111)$ and $\zeta = 2.59$. Gaussian density: ; Statistic $Z_{\hat{\mathcal{L}}}$, $N = 50$: ; Statistic $Z_{\hat{\mathcal{L}}}$, $N = 100$: ; Statistic $Z_{\hat{\mathcal{L}}}$, $N = 1000$: ; Statistic $S_{\hat{\mathcal{L}}}$, $N = 100$:

3.2 Two-sample problem: Inference on inequality comparing two populations

3.2.1 Inference in the two-sample problem: finite-sample distributions

In this section we focus on comparisons of the finite-sample performance of the $t-$statistic based approach to two-sample robust inference, with the permutation and bootstrap tests proposed by Dufour et al. (2019).

The true values of the inequality index $\mathcal{L}$ in two populations with cdfs F_I and F_Y are denoted $\mathcal{L}^I = \mathcal{L}(F_I)$ and $\mathcal{L}^Y = \mathcal{L}(F_Y)$. The full-sample estimators of $\mathcal{L}$ calculated using samples from the two populations are $\hat{\mathcal{L}}^I = \hat{\mathcal{L}}^I(I_1, ..., I_{N_I})$ and $\hat{\mathcal{L}}^Y = \hat{\mathcal{L}}^Y(Y_1, ..., Y_{N_Y})$. The group estimators of $\mathcal{L}$ in the two samples are $\hat{\mathcal{L}}^I_1, ..., \hat{\mathcal{L}}^I_{q_I}$ and $\hat{\mathcal{L}}^Y_1, ..., \hat{\mathcal{L}}^Y_{q_Y}$.

Asymptotic approaches to testing the hypothesis $H_0 : \mathcal{L}^I - \mathcal{L}^Y = d_0$ (d_0 can be 0 corresponding to the hypothesis of equality of the two inequality indices) against the two-sided alternative H_a : $\mathcal{L}^I - \mathcal{L}^Y \neq d_0$ are based on the normal approximation to the sampling distribution of the difference $\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y$; more precisely, on the standard normal approximation to the sampling distribution of the two-sample $t-$statistic $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y} = (\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y - d_0)/s.e._{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, where $s.e._{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ denotes the usual consistent standard error of the difference $\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y$ (see the formulae in Cowell and Flachaire, 2007, Davidson and Flachaire, 2007, Dufour et al., 2019).

As in the previous section, the validity of the $t-$statistic based robust approach to two-sample inference based on (2) requires weak convergence of the group estimators $\hat{\mathcal{L}}^I_j$, $\hat{\mathcal{L}}^Y_k$, to possibly heterogeneous Gaussian distributions. Again, asymptotic normality of the group estimators $\hat{\mathcal{L}}^I_j$, $j = 1, \ldots, q_I$, and $\hat{\mathcal{L}}_k^Y$, $k = 1, \ldots, q_Y$, holds under the same conditions as for the full-sample estimators $\hat{\mathcal{L}}^I$ and $\hat{\mathcal{L}}^Y$. We refer to the previous section for the assessment of finite-sample distributions of the full-sample estimators of inequality measures and their closeness to normality.

With $q_I = q_Y = q$, the validity of the $t-$statistic based robust approach to two-sample inference based on (3) - i.e., the one-sample $t-$statistic $\tilde{\tilde{t}}_{\mathcal{L}}$ in the q differences $\widehat{\mathcal{L}}_j^I - \widehat{\mathcal{L}}_j^Y$, $j = 1, ..., q$, (without Studentization/normalization) requires weak convergence of the differences $\widehat{\mathcal{L}}_j^I - \widehat{\mathcal{L}}_j^Y$, to possibly heterogeneous Gaussian distributions. Further, asymptotic normality of the differences $\hat{\mathcal{L}}_j^I - \hat{\mathcal{L}}_j^Y$ holds under the same conditions as in the case of the difference $\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y$ between the full-sample estimators.

We begin with assessments of finite-sample distributions of the difference $\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y$ between the (full-sample) inequality estimators, and of the (full-sample) $t-$statistics $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ calculated using them, examining their closeness to the Gaussian distribution. We focus on comparisons of the finite-sample distributions of the $t-$statistic $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, with the difference $\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y$ normalized by its true standard deviation, that is, of the statistic $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y} = (\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y)/\sigma_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, where $\sigma^2_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y} = Var(\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y)$.

In Figures 4-6, we present kernel estimates of the finite-sample densities of the statistics $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ and $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ for samples from two populations with the same S-M distribution. Figures 4-5 present, for Theil and Gini indices, the kernel densities of $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ (for $N_I = N_Y = N = 50, 100, 1000$) and $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ ($N = 100$)[20] for samples drawn from the S-M distribution $SM(a_0, b_0, c_0)$ with the parameters $a_0 = 2.8$, $b_0 = 100^{-1/2.8}$, $c_0 = 1.7$ and the corresponding tail index $\zeta = 4.76$. Figure 6 provides the analogous kernel densities for the Gini index when two samples are drawn from a more heavy-tailed S-M distribiion $SM(a, b_0, c)$ with $(a, c) = (5.8, 0.447)$ and the tail index $\zeta = 2.59$.

The finite-sample distributions of the statistic $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ and thus of the difference $\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y$ are approximately symmetric even with rather small samples and also under pronounced heavy-tailedness, with satisfactory Gaussian approximations as compared to the finite-sample distribution of the (full-sample) $t-$statistic $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$. This also holds when the sample sizes are not very different. [21]

3.2.2 Inference in two-sample problem: finite-sample size properties

Tables 1-3 provide the finite-sample size properties of the asymptotic, permutation, bootstrap and $t-$statistic based robust tests of equality of Theil and Gini indices. As before, we consider two samples, $I_1, ..., I_{N_I}$ and $Y_1, ..., Y_{N_Y}$, drawn from $SM(a_I, b_0, c_I)$ and $SM(a_Y, b_0, c_Y)$, with $b_0 = 100^{-1/2.8}$ and tail indices $\zeta_I = a_I c_I$, $\zeta_Y = a_Y c_Y$. In simulations, we consider a variety of settings with equal as well as different sample sizes N_I, N_Y; and a range of numbers of groups q_I, q_Y for $t-$statistic based robust tests. We focus on the important and empirically relevant case of samples drawn from different (income or wealth) distributions, with the results for the case of identical distributions presented and discussed in Appendix C.

Table 1 presents results for identical sample sizes $N_I = N_Y = N = 200$, and different distributions. The empirical sizes of the simple-to-use two-sample $t-$statistic based robust tests based on (2) with $q = 4, 8$ in the case of more heavy-tailed distributions, and based on (3) with $q = 4, 8, 12, 16$ in the

[20]Qualitatively similar results for other sample sizes N are omitted for brevity and available on request.

[21]If the sample sizes of two groups are very different, then different partition, q_I, q_Y should be used in applications of the t-statistic inference approach.

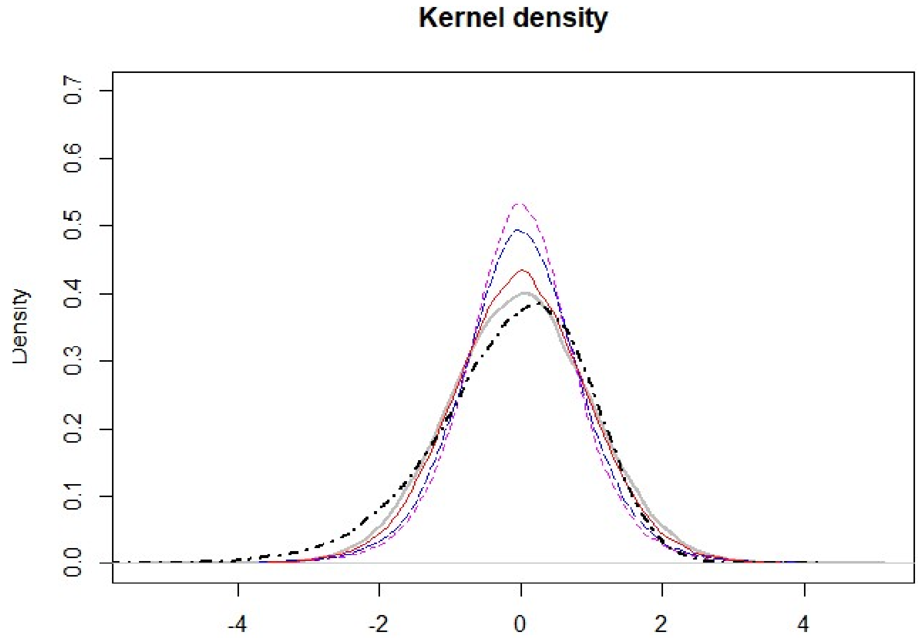

Figure 4: Kernel density functions for the statistics $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ and $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ for the difference between Theil indices: S-M distributions $SM(a_0, b_0, c_0)$ with $(a_0, c_0) = (2.8, 1.7)$ and $\zeta = 4.76$. Gaussian density: ; Statistic $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, $N = 50$: ; Statistic $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, $N = 100$: ; Statistic $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, $N = 1000$: ; Statistic $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, $N = 100$:

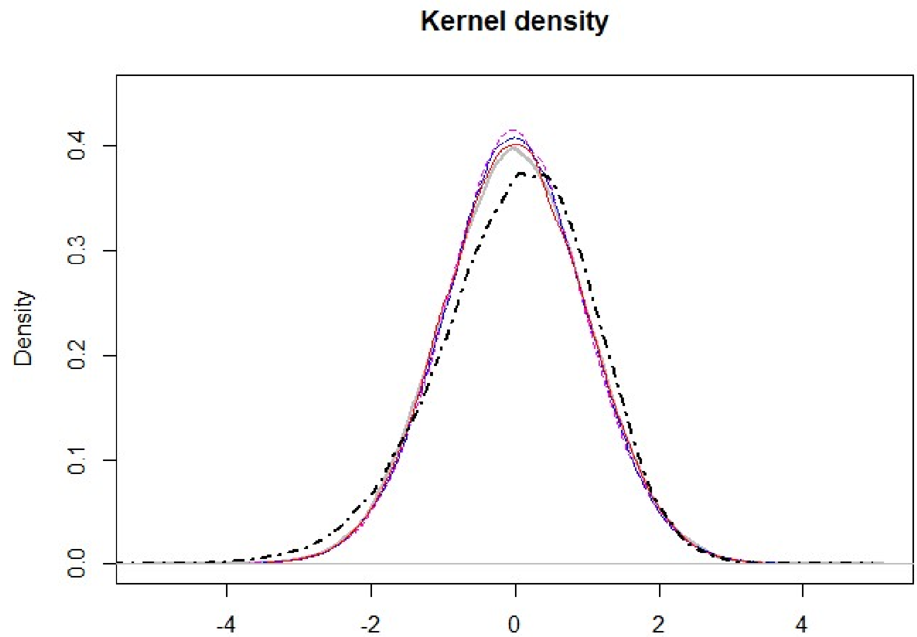

Figure 5: Kernel density functions for the statistics $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ and $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$ for the difference between Gini indices: S-M distributions $SM(a_0, b_0, c_0)$ with $(a_0, c_0) = (2.8, 1.7)$ and $\zeta = 4.76$.. Gaussian density: ; Statistic $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, $N = 50$: ; Statistic $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, $N = 100$: ; Statistic $Z_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, $N = 1000$: , Statistic $S_{\hat{\mathcal{L}}^I - \hat{\mathcal{L}}^Y}$, $N = 100$:

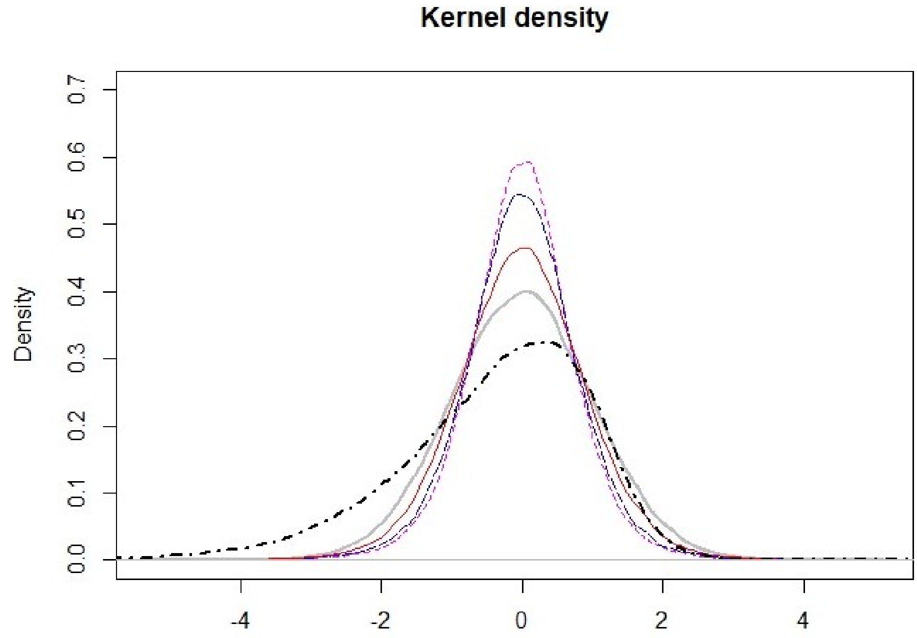

Figure 6: Kernel density functions for the statistics $Z_{\hat{\mathcal{L}}^I-\hat{\mathcal{L}}^Y}$ and $S_{\hat{\mathcal{L}}^I-\hat{\mathcal{L}}^Y}$ for the difference between Gini indices: S-M distributions $SM(a, b_0, c)$ with $(a, c) = (5.8, 0.4473111)$ and $\zeta = 2.59$. Gaussian density: ; Statistic $Z_{\hat{\mathcal{L}}^I-\hat{\mathcal{L}}^Y}$, $N = 50$: ; Statistic $Z_{\hat{\mathcal{L}}^I-\hat{\mathcal{L}}^Y}$, $N = 100$: ; Statistic $Z_{\hat{\mathcal{L}}^I-\hat{\mathcal{L}}^Y}$, $N = 1000$: ; Statistic $S_{\hat{\mathcal{L}}^I-\hat{\mathcal{L}}^Y}$, $N = 100$:

case of less heavy-tailed distributions, are comparable and in some cases are better than those of the computationally expensive permutation and bootstrap tests. The two-sample $t-$statistic based tests with the same number of groups $q_I = q_Y = q$ appear to have less over-rejections as compared to the one-sample $t-$statistic based tests for differences of the group estimators.

Table 2 provides results for different sample sizes N_I, N_Y and different distributions. We observe better size properties for the two-sample $t-$statistic based inference approach in comparison to permutation and bootstrap tests, based on $\tilde{t}_{\mathcal{L}}$ in (2) (with $q_I = q_Y = 4, 8, 12$ for all sample sizes, and also with $q = 16$ for large sample sizes) as well as those based on $\tilde{\tilde{t}}_{\mathcal{L}}$ in (3) (with $q_I = q_Y = 4, 8$ for all sample sizes) .

Finally, Table 3 provides the results for the case of samples with dependent observations, i.e., those with spatially dependent data. Each of the two samples consists of 192 observations with the standard lognormal distribution ($\mu = 0$ and $\sigma = 1$) located on a rectangular array of unit squares with 16 rows and 12 columns. The observations are generated such that the correlation between the logarithms of two observations is given by $\exp(-\phi d)$ for some $\phi > 0$, where d is the Euclidean distance between the two observations (see Section 3.4 in Ibragimov and Müller (2010) for the use of a similar spatially correlated setting in the analysis of finite sample size properties of one-sample $t-$statistic based approach for inference on the mean of Gaussian observations with spatial dependence). The case $\phi = \infty$ corresponds to samples of i.i.d. observations. More precisely, the observations in the samples are given by $I_{ij} = \exp(u_{ij})$, $Y_{ij} = \exp(v_{ij})$, $i = 1, ..., 16$, $j = 1, ..., 12$, where u_{ij} and v_{ij} are multivariate mean zero unit variance Gaussian with correlation between u_{ij} and u_{lk} and between v_{ij}

and v_{lk} equals $\exp(-\phi\sqrt{(i-l)^2+(j-k)^2})$.

The empirical size properties of $t-$statistic tests of equality of Theil and Gini indices in the two samples with spatial dependence are comparable (especially, for tests based on the two-sample $t-$statistic $\tilde{\tilde{t}}$ with $q_I = q_Y = q = 4$ groups and the one-sample $t-$statistic $\tilde{t}$ in differences with $q = 8$) to those of permutation and bootstrap procedures. Furthermore, the finite sample size properties of essentially all $t-$statistic based robust tests are better than those of bootstrap and permutation tests under pronounced spatial dependence with $\phi = 1$.

Table 1: Empirical size – identical sample sizes, different distributions

Theil$\backslash\zeta_Y$	6.26	3.94	2.9	Gini $\backslash\zeta_Y$	6.6	3.8	2.59
asy	5.1	5.4	12.3	asy	5.2	5.5	8.0
$\tilde{t}_{\mathcal{L}}(q=4)$	1.9	1.5	4.0	$\tilde{t}_{\mathcal{L}}(q=4)$	1.9	1.7	2.7
$\tilde{t}_{\mathcal{L}}(q=8)$	3.1	2.9	8.7	$\tilde{t}_{\mathcal{L}}(q=8)$	3.2	3.1	5.1
$\tilde{t}_{\mathcal{L}}(q=12)$	3.6	3.4	11.4	$\tilde{t}_{\mathcal{L}}(q=12)$	3.8	3.6	6.6
$\tilde{t}_{\mathcal{L}}(q=16)$	4.0	3.9	13.6	$\tilde{t}_{\mathcal{L}}(q=16)$	3.9	3.9	7.8
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	4.8	4.3	7.2	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	5.1	4.7	5.7
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	5.0	4.5	10.6	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	5.1	4.8	6.6
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	4.8	4.6	12.6	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	5.0	4.8	8.1
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	5.1	4.8	14.5	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	5.0	4.8	8.8
permutation	4.8	4.8	11.0	permutation	4.4	4.6	6.2
bootstrap	4.8	4.7	10.6	bootstrap	4.6	4.4	6.3

Notes: $N_I = N_Y = 200$, $(a_I, c_I) = (2.8, 1.7)$, $\zeta_I = 4.76$.

Theil index: (a_Y, c_Y)=$(2.5, 2.502), (3.2, 1.232), (5.8, 0.4996)$.

Gini coefficient: (a_Y, c_Y)=$(2.5, 2.640), (3.2, 1.1866), (5.8, 0.447)$.

Table 2: Empirical size – different distributions, different sample sizes

Theil, $\zeta_Y = 2.9\ \backslash N_Y$	50	200	500	1000	5000	Gini, $\zeta_Y = 2.59\backslash N_Y$	50	200	500	1000	5000
asymptotic	14.5	12.3	12.2	11.2	9.0	asymptotic	12.6	8.0	7.5	6.6	6.0
$\tilde{t}_{\mathcal{L}}(q=4)$	3.9	4.0	4.8	4.2	4.0	$\tilde{t}_{\mathcal{L}}(q=4)$	2.8	2.7	2.9	2.7	2.6
$\tilde{t}_{\mathcal{L}}(q=8)$	8.8	8.7	8.6	8.0	6.7	$\tilde{t}_{\mathcal{L}}(q=8)$	6.0	5.1	5.1	4.8	4.3
$\tilde{t}_{\mathcal{L}}(q=12)$	11.4	11.4	11.3	9.7	7.8	$\tilde{t}_{\mathcal{L}}(q=12)$	7.7	6.6	6.6	5.8	5.1
$\tilde{t}_{\mathcal{L}}(q=16)$	13.9	13.6	12.8	11.2	8.4	$\tilde{t}_{\mathcal{L}}(q=16)$	9.8	7.8	7.5	6.6	5.5
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	6.9	7.2	7.8	7.3	6.6	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	5.7	5.7	6.1	5.7	5.2
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	11.1	10.6	10.2	9.6	7.9	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	8.2	6.6	6.8	6.2	5.6
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	13.0	12.6	12.2	10.9	8.7	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	9.3	8.1	7.7	6.9	6.2
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	14.9	14.5	13.6	12.1	9.0	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	11.0	8.8	8.2	7.4	6.2
permutation	12.4	11.0	11.6	10.5	8.8	permutation	8.9	6.2	6.5	6.1	5.6
bootstrap	12.2	10.6	11.1	10.4	8.5	bootstrap	9.1	6.3	6.3	6.0	5.7

Notes: $N_I = 200$, $\zeta_I = 4.76$.

Table 3: Empirical size – spatial correlation

Theil \ ϕ	∞	2	1	Gini\ ϕ	∞	2	1
asymptotic	7.1	7.6	14.6	asymptotic	7.4	7.9	16.3
$\tilde{t}_{\mathcal{L}}(q=4)$	1.6	1.6	2.0	$\tilde{t}_{\mathcal{L}}(q=4)$	1.6	1.6	2.0
$\tilde{t}_{\mathcal{L}}(q=8)$	5.4	5.3	6.0	$\tilde{t}_{\mathcal{L}}(q=8)$	5.4	5.3	6.0
$\tilde{t}_{\mathcal{L}}(q=12)$	6.8	7.3	8.1	$\tilde{t}_{\mathcal{L}}(q=12)$	6.8	7.3	8.1
$\tilde{t}_{\mathcal{L}}(q=16)$	7.5	7.8	8.7	$\tilde{t}_{\mathcal{L}}(q=16)$	7.5	7.8	8.7
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	4.2	4.3	4.7	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	4.2	4.3	4.7
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	7.1	7.4	8.1	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	7.1	7.4	8.1
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	8.2	8.6	9.7	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	8.2	8.6	9.7
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	8.7	9.0	9.8	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	8.7	9.0	9.8
permutation	4.4	4.8	10.1	permutation	4.1	4.3	9.3
bootstrap	4.4	4.9	11.2	bootstrap	4.3	4.9	11.0

3.2.3 Inference in two-sample problem: Finite-sample power properties

We turn to the study of finite-sample, size-adjusted power of the two-sample $t-$statistic based test and the permutation test.[22] With size adjustment, the empirical size of a given t-statistic-based robust test and its permutation counterpart coincide under the null hypothesis, thereby enabling meaningful power comparisons. We consider a variety of simulation designs.

Table 4 presents the power and Table 5 presents the size-adjusted power when the two samples come from different S-M distributions. The results in the tables demonstrate that t-statistic approach is slightly undersized, leading to higher size-adjusted power in comparison to the non-size-adjusted case. The reason of using size-adjusted power is to compare the performance of the different tests with different sizes. The permutation test appears to be the most powerful with the two-sample t-statistic based tests (using $\tilde{t}_{\mathcal{L}}$ in 2) close behind (for $q = 12, 16$). The two-sample tests based on $\tilde{t}_{\mathcal{L}}$ (2) with $q_I = q_Y = q$ are always more powerful than those based on the one-sample $t-$statistic $\tilde{\tilde{t}}_{\mathcal{L}}$ (3) in the differences of the group estimators with the same number of groups.[23]

Tables 6 and 7 provide the results on finite sample power properties of different inference approaches in the case of more heavy-tailed distributions, standard (unadjusted) and size-adjusted, respectively. Similar to the previous case, the results in the tables show that t-statistic approach is slightly undersized, leading, as before, to higher size-adjusted power in comparison to the non-size-adjusted case. In the case of Theil or Gini indices, the $t-$statistic based tests based on $\tilde{t}_{\mathcal{L}}$ in (2) with

[22]Size adjustment is not performed for bootstrap tests as they are strongly dominated in terms of power by permutation test in all settings considered, see also Dufour et al. (2019).

[23]In inference on both Theil and Gini indices, the power of $t-$statistic based approach based on (2) is very similar across $q = 8, 12, 16$ if the second distribution is more light-tailed than the first one ($c > c_0$ and $\zeta_Y > \zeta_I$) and also very similar for $q = 12, 16$ if the second distribution is more heavy-tailed than the first one ($c < c_0$ and $\zeta_Y < \zeta_I$). In the former case the best power is returned by the $t-$statistic based test based on (2) with $q_I = q_Y = q = 8, 12$. In the latter case the most powerful $t-$statistic test for inference on Theil indices is the one based on (2) with $q_I = q_Y = 16$, and the second best test is the t-statistic based test based on (2) with $q_I = q_Y = 12$. Further, in this case the most powerful $t-$statistic test for inference on Gini indices is the test based on (2) with $q_I = q_Y = 12$.

Table 4: Power – identical sample sizes

Theil\ ζ_Y	1.96	3.08	4.76	7.56	88.76	Gini\ ζ_Y	1.96	3.08	4.76	7.56	88.76
asymptotic	68.5	22.3	8.4	30.2	97.8	asymptotic	98.3	44.4	5.9	27.4	93.2
$\tilde{t}_{\mathcal{L}}(q=4)$	21.8	4.7	2.0	8.6	63.2	$\tilde{t}_{\mathcal{L}}(q=4)$	64.5	14.7	2.0	10.0	57.3
$\tilde{t}_{\mathcal{L}}(q=8)$	42.9	9.4	3.2	15.8	88.3	$\tilde{t}_{\mathcal{L}}(q=8)$	91.1	27.1	3.5	16.6	80.4
$\tilde{t}_{\mathcal{L}}(q=12)$	47.1	11.1	3.7	17.9	90.3	$\tilde{t}_{\mathcal{L}}(q=12)$	93.7	29.4	3.7	18.0	81.5
$\tilde{t}_{\mathcal{L}}(q=16)$	49.9	11.7	4.0	19.0	90.3	$\tilde{t}_{\mathcal{L}}(q=16)$	94.5	30.8	4.3	18.7	81.4
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	27.7	9.0	4.6	14.0	64.1	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	65.7	21.3	4.9	15.6	60.3
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	45.2	11.6	4.9	18.8	87.4	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	90.3	30.4	4.8	19.9	79.4
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	49.0	12.4	4.8	20.0	89.4	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	93.2	31.8	5.1	20.1	80.8
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	50.6	12.9	4.8	20.2	90.0	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	94.2	32.1	5.2	19.8	81.3
permutation	48.2	13.6	4.8	19.6	94.2	permutation	97.6	39.7	4.5	21.6	90.0
bootstrap	43.1	12.3	3.9	18.4	91.1	bootstrap	95.2	39.0	4.6	21.8	90.2

Notes: $N_I = N_Y = 200$.

$q_I = q_Y = q$ for $t-$statistic based tests.

I sample distribution $SM(a, b_0, c)$, $a_0 = 2.8, c_0 = 1.7$, $\zeta_I = 4.76$.

II sample distribution $SM(a, b_0, c)$, $a_0 = 2.8$, $c = 0.7, 1.1, 1.7, 2.7, 31.7$; $\zeta_Y = 1.96, 3.08, 4.76, 7.56, 88.76$.

Table 5: Size-adjusted power – identical sample sizes

Theil\ ζ_Y	1.96	3.08	4.76	7.56	88.76	Gini\ ζ_Y	1.96	3.08	4.76	7.56	88.76
asymptotic	87.2	35.6	4.7	23.6	90.4	asymptotic	97.5	39.2	4.5	23.3	91.4
$\tilde{t}_{\mathcal{L}}(q=4)$	62.6	22.5	4.7	17.6	72.3	$\tilde{t}_{\mathcal{L}}(q=4)$	82.4	26.2	4.5	17.5	75.4
$\tilde{t}_{\mathcal{L}}(q=8)$	82.7	29.4	4.7	20.7	83.6	$\tilde{t}_{\mathcal{L}}(q=8)$	93.6	31.6	4.5	19.6	84.1
$\tilde{t}_{\mathcal{L}}(q=12)$	88.3	30.5	4.7	21.4	83.5	$\tilde{t}_{\mathcal{L}}(q=12)$	94.7	32.5	4.5	20.3	83.6
$\tilde{t}_{\mathcal{L}}(q=16)$	91.1	30.8	4.7	20.8	82.8	$\tilde{t}_{\mathcal{L}}(q=16)$	94.7	31.3	4.5	19.2	81.9
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	43.6	17.3	4.7	15.2	55.2	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	63.4	20.0	4.5	14.7	57.9
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	75.5	26.3	4.7	19.5	77.2	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	89.2	28.8	4.5	18.5	77.7
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	83.6	27.5	4.7	19.3	78.7	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	92.0	29.2	4.5	18.1	78.5
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	88.4	28.7	4.7	19.6	79.4	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	93.1	29.8	4.5	18.2	78.8
permutation	91.6	37.4	4.7	21.9	88.7	permutation	97.6	39.7	4.5	21.6	90.0
bootstrap	77.7	33.6	4.3	20.6	87.2	bootstrap	95.2	39.0	4.6	21.8	90.2

Notes: $N_I = N_Y = 200$.

$q_I = q_Y = q$ for $t-$statistic based tests.

I sample distribution $SM(a, b_0, c)$, $a_0 = 2.8, c_0 = 1.7$, $\zeta_I = 4.76$.

II sample distribution $SM(a, b_0, c)$, $a_0 = 2.8$, $c = 0.7, 1.1, 1.7, 2.7, 31.7$; $\zeta_Y = 1.96, 3.08, 4.76, 7.56, 88.76$.

$q_I = q_Y = 8, 12, 16$ are typically the most powerful among such tests with different choices of the number of groups (for not very light-tailed second distribution); they are typically more powerful than permutation tests. In the case of Theil indices, the best power properties are observed for the $t-$statistic based tests with $q = 16$, and the second best test is the t-statistic test based on (2) with $q_I = q_Y = 12$. The choice of $q = 12, 16$ also provides the best power properties for $t-$statistic tests based on $\tilde{t}_{\mathcal{L}}$ in (2) in inference on Gini indices.

Table 6: Power, $\zeta_I = 2.2$, identical sample sizes $N_I = N_Y = 200$

Theil\ ζ_Y	1.4	1.8	2.2	3	7.4	Gini\ ζ_Y	1.4	1.8	2.2	3	7.4
asymptotic	58.5	19.0	7.6	24.2	91.4	asymptotic	89.1	38.5	5.4	25.8	91.8
$\tilde{t}_{\mathcal{L}}(q=4)$	10.8	2.9	1.5	5.2	43.9	$\tilde{t}_{\mathcal{L}}(q=4)$	39.4	11.2	1.8	8.7	49.8
$\tilde{t}_{\mathcal{L}}(q=8)$	27.7	6.5	2.4	11.7	76.2	$\tilde{t}_{\mathcal{L}}(q=8)$	73.5	22.0	2.9	15.3	76.4
$\tilde{t}_{\mathcal{L}}(q=12)$	36.5	8.2	2.9	14.6	84.3	$\tilde{t}_{\mathcal{L}}(q=12)$	83.6	25.1	3.4	17.1	79.1
$\tilde{t}_{\mathcal{L}}(q=16)$	41.5	9.8	3.4	16.0	87.8	$\tilde{t}_{\mathcal{L}}(q=16)$	88.8	27.1	3.8	18.2	80.0
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	14.3	5.6	3.5	9.3	46.0	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	41.4	16.1	4.4	14.2	53.0
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	30.3	8.1	3.7	14.0	75.6	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	73.5	24.6	4.3	18.3	75.5
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	37.8	9.2	3.8	16.3	83.8	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	83.2	27.0	4.5	19.1	78.3
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	42.8	10.9	4.1	17.4	87.6	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	88.4	28.6	4.6	19.5	79.3
permutation	34.2	11.8	4.7	17.3	90.0	permutation	95.5	38.3	4.6	22.0	90.1
bootstrap	26.2	8.9	3.3	13.8	78.2	bootstrap	82.4	34.5	4.3	21.2	88.4

Notes: $N_I = N_Y = 200$.

I sample distribution: $SM(a, b_0, c)$, $a = 2, c = 1.1$ and $\zeta_I = 2.2$.

II sample distribution $SM(a, b_0, c)$ $a = 2$, $c = 0.7, 0.9, 1.1, 1.5, 3.7$; $\zeta_Y = 1.4, 1.8, 2.2, 3, 7.4$.

Table 7: Size-adjusted power, $\zeta_I = 2.2$, identical sample sizes $N_I = N_Y = 200$

Theil\ ζ_Y	1.4	1.8	2.2	3	7.4	Gini\ ζ_Y	1.4	1.8	2.2	3	7.4
asymptotic	49.76	13.71	4.71	17.72	85.99	asymptotic	59.93	15.61	4.75	21.05	95.03
$\tilde{t}_{\mathcal{L}}(q=4)$	28.37	9.25	4.71	15.83	73.85	$\tilde{t}_{\mathcal{L}}(q=4)$	39.79	10.82	4.75	17.89	83.77
$\tilde{t}_{\mathcal{L}}(q=8)$	40.72	11.3	4.71	18.76	87.21	$\tilde{t}_{\mathcal{L}}(q=8)$	51.02	12.83	4.75	20.52	92.41
$\tilde{t}_{\mathcal{L}}(q=12)$	45.08	11.97	4.71	19.95	90.05	$\tilde{t}_{\mathcal{L}}(q=12)$	52.19	13.36	4.75	21.2	92.33
$\tilde{t}_{\mathcal{L}}(q=16)$	47.94	12.4	4.71	20.1	91.22	$\tilde{t}_{\mathcal{L}}(q=16)$	52.51	13.08	4.75	20.71	91.51
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	20.04	7.72	4.71	12.81	56.42	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	28.35	9.3	4.75	14.38	64.9
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	34.93	9.77	4.71	16.56	80.37	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	44.19	11.25	4.75	18.05	86.69
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	40.99	10.53	4.71	18.18	86.32	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	48.82	12.43	4.75	19.97	89.29
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	45.19	11.98	4.71	19.17	89.14	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	50.57	12.85	4.75	20.14	89.97
permutation	34.17	11.79	4.71	17.33	89.95	permutation	48.24	13.6	4.75	19.63	94.18
bootstrap	26.18	8.92	3.32	13.78	78.16	bootstrap	43.1	12.25	3.87	18.39	91.1

Notes: $N_I = N_Y = 200$.

I sample distribution: $SM(a, b_0, c)$, $a = 2, c = 1.1$ and $\zeta_I = 2.2$.

II sample distribution $SM(a, b_0, c)$ $a = 2$, $c = 0.7, 0.9, 1.1, 1.5, 3.7$; $\zeta_Y = 1.4, 1.8, 2.2, 3, 7.4$.

Tables 8 and 9 provide the results on finite-sample power properties of different inference approaches for heavy-tailed distributions, including those considered in Table 7. The two-sample t-statistic based tests are typically much more powerful than permutation tests if the more heavy-tailed distribution has a larger sample size. Again, two-sample t-statistic based tests based on $\tilde{t}_{\mathcal{L}}$ (2) with the number of groups $q_I = q_Y = q$ are always more powerful than those based on the one-sample $t-$statistic $\tilde{\tilde{t}}_{\mathcal{L}}$ (3) in the differences of the group estimators with the same number of groups.

Table 8: Size-adjusted power– different sample sizes

Theil\ ζ_Y	1.4	1.8	2.2	3	7.4	Gini\ ζ_Y	1.4	1.8	2.2	3	4.4
asymptotic	62.61	20.09	4.47	13.74	87.24	asymptotic	73.49	21.82	4.2	21.41	75.46
$\tilde{t}_{\mathcal{L}}(q=4)$	40.75	14.61	4.47	9.84	68.67	$\tilde{t}_{\mathcal{L}}(q=4)$	57.1	16.06	4.2	16.47	56.77
$\tilde{t}_{\mathcal{L}}(q=8)$	58.48	19.39	4.47	8.33	80.52	$\tilde{t}_{\mathcal{L}}(q=8)$	72.14	21.62	4.2	18.05	67.1
$\tilde{t}_{\mathcal{L}}(q=12)$	65.39	22.06	4.47	5.08	78.17	$\tilde{t}_{\mathcal{L}}(q=12)$	75.29	22.9	4.2	15.64	63.4
$\tilde{t}_{\mathcal{L}}(q=16)$	70.03	23.41	4.47	3.06	73.4	$\tilde{t}_{\mathcal{L}}(q=16)$	78.8	24.63	4.2	14.53	62.81
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	28.92	11.94	4.47	7.94	49.73	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	39.41	12.4	4.2	11.21	36.8
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	51.49	17.78	4.47	7.41	72.84	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	64.51	19.31	4.2	14.92	57.43
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	59.87	20.39	4.47	4.48	71.01	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	72.52	22.26	4.2	15.32	60.2
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	65.75	22.16	4.47	2.96	68.15	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	75.23	22.96	4.2	13.53	58.66
permutation	45.47	13.75	4.47	23.59	96.1	permutation	58.98	16.22	4.2	27.71	82.51
bootstrap	28.62	9.72	3.23	20.1	89.22	bootstrap	50.06	14.59	3.85	25.86	78.94

Notes: $N_I = 200$, $N_Y = 400$.
I sample distribution: $SM(a, b_0, c)$, $a = 2, c = 1.1$ and $\zeta_I = 2.2$.
Theil index II sample distribution $SM(a, b_0, c)$ $a = 2$, $c = 0.7, 0.9, 1.1, 1.5, 3.7$; $\zeta_Y = 1.4, 1.8, 2.2, 3, 7.4$.
Gini index II sample distribution replaces the last column with $c = 2.2$ and $\zeta_Y = 4.4$.

Table 9: Size-adjusted power – different sample sizes

Theil ζ_Y	1.4	1.8	2.2	3	7.4	Gini ζ_Y	1.4	1.8	2.2	3	4.4
asymptotic	44.56	9.2	4.15	27.76	92.67	asymptotic	64.67	14.91	4.45	32.15	84.15
$\tilde{t}_{\mathcal{L}}(q=4)$	19.25	4.87	4.15	23.21	83.34	$\tilde{t}_{\mathcal{L}}(q=4)$	38.92	9.19	4.45	25.61	70.46
$\tilde{t}_{\mathcal{L}}(q=8)$	21.13	3.53	4.15	28.2	92.52	$\tilde{t}_{\mathcal{L}}(q=8)$	46.64	9.29	4.45	31.18	80.64
$\tilde{t}_{\mathcal{L}}(q=12)$	15.17	2.01	4.15	27.94	93.97	$\tilde{t}_{\mathcal{L}}(q=12)$	44	7.99	4.45	31.57	80.67
$\tilde{t}_{\mathcal{L}}(q=16)$	11.7	1.3	4.15	30.44	95.66	$\tilde{t}_{\mathcal{L}}(q=16)$	42.78	7.21	4.45	32.73	81.5
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	12.9	4.23	4.15	17.61	65.25	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	26.82	7.66	4.45	19.65	52.38
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	16.79	3.02	4.15	24.4	86.85	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	41.27	8.31	4.45	28.27	74.04
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	13.08	1.77	4.15	25.76	90.44	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	40.46	7.34	4.45	29.29	76.66
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	10.53	1.23	4.15	28.23	93.5	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	40.28	7.06	4.45	30.68	78.65
permutation	44.57	14.95	4.15	21.71	97.47	permutation	62.78	18.55	4.45	24.75	77.41
bootstrap	40.34	12.9	3.58	16.33	84.87	bootstrap	59.58	17.66	4.12	23.13	74.1

Notes: $N_I = 400$, $N_Y = 200$.
I sample distribution: $SM(a, b_0, c)$, $a = 2, c = 1.1$ and $\zeta_I = 2.2$.
Theil index II sample distribution $SM(a, b_0, c)$ $a = 2$, $c = 0.7, 0.9, 1.1, 1.5, 3.7$; $\zeta_Y = 1.4, 1.8, 2.2, 3, 7.4$.
Gini index II sample distribution replaces the last column with $c = 2.2$ and $\zeta_Y = 4.4$.

Table 10 gives the results on finite-sample size-adjusted power of different inference approaches in the same distributional settings as in Table 8 with sample sizes $N_I = 200$ and $N_Y = 800$. Table 11 provides the results on finite-sample size-adjusted power properties of the approaches in the same settings as in Table 9 with sample sizes $N_I = 800$ and $N_Y = 200$. We also consider different combinations of numbers q_I and q_Y of groups for the $t-$statistic based approach.

According to the results in Tables 10 and 11, if the smaller sample is more heavy-tailed then the power of all two-sample t-statistic based tests is dominated by that of permutation tests. Else, if the larger sample is more heavy-tailed then the power properties of two-sample t-statistic based tests (except the tests with very small q_I and q_Y) are typically considerably better than those of permutation test. For inference on Theil indices, the best (compromise) choice of the number of groups in $t-$statistic based approach is $q_I = 12$, $q_Y = 6$ or *vice versa* because this choice leads to correct size and good power in comparison to other size-controlled two-sample t-statistic based tests. For Gini indices, the finite-sample power properties are not very sensitive to choice q_I and q_Y. Interestingly, even if the samples differ by four times as in the tables, the choice $q_I = q_Y = 8, 12, 16$ leads to a very good size-adjusted power and seems to be one of the best across all combinations of q_I and q_Y. The choice $q_I = 12$ and $q_Y = 9$ is also good and leads to power properties of $t-$statistic inference approach that are comparable or slightly better than in the case $q_I = q_Y = 8, 12, 16$. The choice of the different number of groups q_I and q_Y may be useful if the sizes of two samples differ very much.

Table 10: Size-adjusted power, $\zeta_I = 2.2$, different sample sizes, $N_I = 200$, $N_Y = 800$

Theil $\backslash\zeta_Y$	1.4	1.8	2.2	3	7.4	Gini $\backslash\zeta_Y$	1.4	1.8	2.2	3	4.4
asymptotic	64.9	23.7	4.7	7.6	82.03	asymptotic	81.2	29.1	4.8	21.8	79.6
$\tilde{t}_{\mathcal{L}}(q=4)$	45.8	18.5	4.7	6.2	58.53	$\tilde{t}_{\mathcal{L}}(q=4)$	69.9	22.6	4.8	16.8	58.7
$\tilde{t}_{\mathcal{L}}(q=8)$	58.8	21.4	4.7	2.3	60.79	$\tilde{t}_{\mathcal{L}}(q=8)$	83.0	28.0	4.8	15.0	65.3
$\tilde{t}_{\mathcal{L}}(q=12)$	65.5	22.5	4.7	0.8	49.03	$\tilde{t}_{\mathcal{L}}(q=12)$	87.1	30.6	4.8	11.6	60.9
$\tilde{t}_{\mathcal{L}}(q=16)$	71.4	24.4	4.7	0.3	37.4	$\tilde{t}_{\mathcal{L}}(q=16)$	89.7	31.8	4.8	9.0	57.1
$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=3)$	41.9	16.8	4.7	8.6	56.32	$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=3)$	62.8	19.6	4.8	17.5	54.6
$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=6)$	58.9	21.2	4.7	4.1	64.72	$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=6)$	80.0	27.2	4.8	17.6	66.4
$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=9)$	67.1	22.8	4.7	2.2	62.79	$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=9)$	84.6	28.9	4.8	15.6	66.8
$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=12)$	72.3	25.0	4.7	0.9	55.22	$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=12)$	87.2	30.6	4.8	13.5	64.5
$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=2)$	29.1	11.4	4.7	13.9	52.99	$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=2)$	45.7	13.6	4.8	19.0	49.1
$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=4)$	57.5	20.1	4.7	8.6	66.29	$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=4)$	74.3	23.6	4.8	19.5	63.5
$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=6)$	66.1	22.4	4.7	5.7	69.66	$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=6)$	79.3	26.1	4.8	19.2	68.1
$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=8)$	71.4	23.9	4.7	4.2	70.21	$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=8)$	81.8	27.0	4.8	18.3	69.7
permutation	54.4	16.3	4.7	33.6	97.68	permutation	65.8	18.8	4.8	38.9	91.6
bootstrap	30.3	10.3	3.6	29.7	95.78	bootstrap	56.0	16.5	4.4	36.9	90.2

Summarizing these results, the easily applied two-sample t-statistic based approach to inference on equality of, and on the difference between, inequality in two populations appears to be very useful, and can complement the computationally expensive bootstrap and permutation-based inference methods. Finite-sample properties of the $t-$statistic based approach appear to be better in the case of the Gini index as compared to the Theil index as the former is more robust to heavy tails.

As discussed in Section 2.3, the simplest way to choose the number of groups in the case of population distributions that are not very different from each other is to have q_I/q_Y (approximately) equal to N_I/N_Y so that the sizes of all the groups considered are about the same. If the two population distributions have similar tail indices, then in the case of inference on Gini measures, q_I and q_Y may be taken to be equal. In general, the size of the groups in the sample from a more heavy-tailed distribution should be larger than the size of the groups from a less heavy-tailed distribution. Thus in the case of equally sized samples, one should take the number of groups in the more heavy-tailed sample to be less than the number of groups in the less heavy-tailed sample.

Table 11: Size-adjusted power, $\zeta_I = 2.2$, different sample sizes, $N_I = 800$, $N_Y = 200$

Theil $\backslash\zeta_Y$	1.4	1.8	2.2	3	7.4	Gini $\backslash\zeta_Y$	1.4	1.8	2.2	3	4.4
asymptotic	34.0	4.8	4.9	35.8	93.74	asymptotic	65.1	11.7	4.8	40.8	91.4
$\tilde{t}_{\mathcal{L}}(q=4)$	11.2	2.6	4.9	28.3	87.08	$\tilde{t}_{\mathcal{L}}(q=4)$	35.4	7.1	4.8	32.2	81.6
$\tilde{t}_{\mathcal{L}}(q=8)$	7.3	1.1	4.9	33.8	92.78	$\tilde{t}_{\mathcal{L}}(q=8)$	41.1	6.2	4.8	40.3	90.1
$\tilde{t}_{\mathcal{L}}(q=12)$	3.2	0.6	4.9	34.5	94.11	$\tilde{t}_{\mathcal{L}}(q=12)$	37.0	4.4	4.8	42.4	91.0
$\tilde{t}_{\mathcal{L}}(q=16)$	1.2	0.5	4.9	35.3	94.63	$\tilde{t}_{\mathcal{L}}(q=16)$	30.6	2.8	4.8	41.1	90.1
$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=3)$	14.9	4.1	4.9	26.7	85.5	$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=3)$	35.2	8.6	4.8	29.8	76.5
$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=6)$	11.3	1.8	4.9	33.4	93.34	$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=6)$	43.2	7.5	4.8	38.3	88.6
$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=9)$	8.4	1.0	4.9	35.7	95.69	$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=9)$	45.6	6.4	4.8	41.9	91.2
$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=12)$	4.3	0.6	4.9	36.2	96.56	$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=12)$	40.3	5.1	4.8	41.4	90.9
$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=2)$	20.6	7.4	4.9	20.0	77.76	$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=2)$	32.6	9.9	4.8	21.2	60.0
$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=4)$	17.3	3.9	4.9	33.5	94.34	$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=4)$	42.7	9.3	4.8	35.1	85.7
$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=6)$	15.3	2.4	4.9	35.9	96.79	$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=6)$	47.6	8.7	4.8	38.5	89.2
$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=8)$	13.5	1.8	4.9	37.4	97.96	$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=8)$	49.1	8.1	4.8	40.1	90.6
permutation	53.5	17.2	4.9	27.4	99.34	permutation	73.7	22.5	4.8	30.5	86.8
bootstrap	53.4	16.2	4.0	20.7	90.28	bootstrap	73.6	21.5	4.4	28.9	84.5

4 Empirical application: Income inequality across Russian regions

This section presents empirical results on inequality measured using the Gini coefficient, comparing Moscow with Russian regions, using the asymptotic, permutation, bootstrap and the $t-$statistic based robust inference approaches. This analysis is based on a large database compiled from household income surveys conducted by the Federal State Statistics Service of Russia (Rosstat) in 2017 (available at `https://www.gks.ru/free_doc/new_site/vndn-2017/index.html`; `https://www.gks.ru/free_doc/new_site/vndn-2017/OHousehold.html`). The database covers 160,000 households in Russian regions, and provides data on households' total income, among many other variables. Inference on income inequality presented in this section is based on the incomes of Russian households, normalized following Rosstat's methodology, by the total number of household members (i.e., the total household incomes per household member).

Table A.1 in the appendix provides the $p-$values for tests of the null hypothesis $H_0 : G_M = G_R$ against the alternative $H_a : G_M \neq G_R$, where G_M is the Gini coefficient in Moscow and G_R is the Gini coefficient in Russian region R. The entries in the table in bold are the $p-$values not greater than 0.05.

The table also reports the values of the Gini coefficients and the tail indices ζ of the income distributions in the regions estimated using (bias-corrected) log-log rank-size regression (with 5% tail truncation; see Gabaix and Ibragimov, 2011). In addition, the values of the ratio N_I/N_Y, where N_I, N_Y denote the number of households surveyed in Moscow and the regions considered is also reported. It should be noted that if q_I or $q_Y > 14$ for the number of groups in the samples for Moscow and the regions dealt with, we can use only the significance level less than 0.083 for the $t-$statistic approach to robust inference.

The estimated Gini coefficients range from 0.236 (Tambov Region) to 0.354 (the Republic of Ingushetia) indicating low to moderate inequality; Among the republics, Tyva Republic has the Gini coefficient of 0.423 indicating high inequality. The value of the Gini coefficient for Moscow is 0.264 indicating rather low inequality.

The point estimates ζ of tail indices of income distribution from log-log rank-size regressions lie in the interval $(3,6)$ for most of Russian regions, with the exception of Karachay-Cherkess ($\zeta = 2.08$) and Mari El ($\zeta = 2.29$) Republics, and Krasnodar ($\zeta = 2.6$), Kursk ($\zeta = 2.75$) and Tyumen ($\zeta = 2.71$) regions. The confidence intervals for tail indices of income distribution in most Russian regions lie to the right of 2 implying finite second moments and finite variances. The 95% confidence intervals for tail indices of income distributions in 33 regions,[24] and 13 republics[25] as well as 3 Autonomous Districts[26] intersect the interval $(1.5,3)$ which is where tail indices of income distribution in developed countries typically lie.

The 95% confidence intervals for tail indices of income distributions in 18 regions including Moscow[27], and 6 republics[28] as well as 3 Autonomous Districts[29] lie to the right of 3 thus implying finite third moments and variances. The tail index estimate for Moscow was 3.96 with the 95% confidence interval $(3.44, 4.48)$.

On the basis of all the tests, including the $t-$statistic based tests with a variety of values for q_I, q_Y, the null hypothesis $H_0 : G_M = G_R$ is rejected in favor of the alternative $H_a : G_M > G_R$ (at the level 2.5%) for the Republic of Tatarstan, Sevastopol City and Bryansk, Kostroma, Tambov and Tula regions. For Penza, Smolensk and Ulyanovsk regions and Udmurtia, $H_0 : G_M = G_R$ is rejected in favor of $H_a : G_M > G_R$ on the base of the asymptotic, bootstrap, permutation and the $t-$statistic based tests with some of the values q_I, q_Y in the table.

Further, according to all the tests conducted including the $t-$statistic based robust tests for most of the values q_I, q_Y, the null hypothesis $H_0 : G_M = G_R$ is rejected in favor of the alternative $H_a : G_M < G_R$ (at the level 2.5%) for Amur, Chelyabinsk, Irkutsk, Khabarovsk, Krasnodar, Krasnoyarsk, Kurgan, Moscow, Sakhalin and Jewish and Yamalo-Nenets Autonomous regions as well as for the Republics of Bashkortostan, Buryatia, Dagestan, Ingushetia, Kalmykia, Khakassia and Sakha (Yakutia); Altai, Chechen, Kabardino-Balkar, Karachay-Cherkess, Komi and Tyva Republics; Kamchatka, Primorskiy, Zabaykalsky Krays; Khanty-Mansi and Nenets Autonomous Okrugs and Chukotka Autonomous District. For Astrakhan, Kaliningrad, Kemerovo, Novosibirsk, Omsk, Penza, Smolensk, Sverdlovsk, Tomsk and Tyumen Regions, $H_0 : G_M = G_R$ is rejected in favor of $H_a : G_M < G_R$ on the base of the asymptotic, bootstrap, permutation and the $t-$statistic tests for some of the values q_I, q_Y

[24]Krasnodar, Krasnoyarsk, Stavropol, Khabarovsk, Arkhangelsk, Astrakhan, Belgorod, Vladimir, Volgograd, Vologda, Voronezh, Ivanovo, Tver, Kemerovo, Kurgan, Kursk, Lipetsk, Magadan, Murmansk, Novosibirsk, Omsk, Oryol, Penza, Pskov, Ryazan, Sakhalin, Sverdlovsk, Smolensk, Tambov, Tomsk, Tyumen, Ulyanovsk and Yaroslav

[25]Altai, Buryatia, Ingushetia, Kabardino-Balkar, Kalmykia, Karachay-Cherkess, Karelia, Komi, Mari El, Mordovia, North Osetia, Tyva and Sakha

[26]Chukotka, Khanty-Mansi and Nenets and Kamchatka Kray

[27]Amur, Bryansk, Chelyabinsk, Irkutsk, Kaliningrad, Kaluga, Kirov, Kostroma, Leningrad, Nizhny Novgorod, Novgorod, Orenburg, Perm, Rostov, Samara, Saratov, Sevastopol and Tula

[28]Adygeya, Bashkortostan, Chuvash, Dagestan, Khakassia and Tatarstan Republics

[29]Kamchatka, Primorsky and Zabaykalsky Krays

in the table.

Two interesting conclusions follow.

First, income inequality appears to be higher in most Russian Regions compared to Moscow.

Second, the conclusions of all the approaches - the asymptotic, bootstrap, permutation, and the $t-$statistic based robust tests - to testing equality of the Gini coefficients G_M and G_R agree among themselves. Two exceptions are Belgorod and Novgorod Regions, where $H_0 : G_M = G_R$ is not rejected in favor of $H_a : G_M < G_R$ on the base of the asymptotic, bootstrap, permutation, but is rejected on the base of the $t-$statistic based tests for some values of q_I, q_Y.

5 Conclusion and suggestions for further research

The $t-$statistic based robust approach to inference on inequality in two populations are computationally undemanding and has a wide range of applicability in econometric and statistical analysis under the problems of heterogeneity, dependence, and heavy-tailedness in observations. The approach does not require the estimation of limiting variances of estimators of interest, in contrast to inference methods based on consistent, e.g., HAC or clustered, standard errors that often have poor finite sample properties, especially under pronounced heterogeneity and dependence in observations. In addition, this inference approach can be used even in the presence of extremes and outliers in observations generated by heavy-tailedness with infinite variances, as also in settings where observations (e.g., on income or wealth) in each of the samples are ***dependent*** among themselves - for instance, due to spatial or clustered dependence, common shocks, or, in the case of time series or panel data on income or wealth levels, due to autocorrelation and dependence in observations over time. Further, when the two samples contain possibly dependent observations, the $t-$statistic inference approach may be used under ***arbitrary*** dependence ***between*** the samples as well as under possibly ***unequal*** sample sizes.

As discussed in this paper and in previous works on the t-statistic based robust inference approach and its applications, including Ibragimov and Müller (2010, 2016), and Section 3 in Ibragimov et al. (2015), the choice of the number of groups q is crucial for robustness. The numerical analysis of finite-sample performance presented here and in earlier studies provides guidance on selecting the number of groups under various distributional, heavy-tailedness, heterogeneity, and dependence settings. However, as highlighted in the aforementioned works and in Section 2.3, asymptotic efficiency results do not admit the use of data-dependent methods to determine the optimal number of groups q when the only assumptions are asymptotic normality and asymptotic independence of group estimators. Whether additional assumptions, such as on the degree of heavy-tailedness in the populations, enable derivation of data-driven optimal values for the number of groups used in robust t-statistic based inference is an interesting and important research question.

Future research should also explore different potential approaches to the formation groups in applications of the t-statistic based robust inference. This may include random splits and all possible splits along with inference procedures based on metrics such as the median, average, or quantiles of the t-statistics calculated from the corresponding group estimates. We thank an anonymous referee for this suggestion. Dagayev and Stoyan (2020) recently considered random sample splits in their empirical application, constructing t-statistics for group estimates of the parameters under analysis and basing the inference on quantiles, including the median and others, from the empirical distribution of the obtained t-statistics values for the splits considered. Investigating the finite sample and asymptotic properties of such inference procedures is an important area for further research.

The recent paper by Midões and de Crombrugghe (2023) citing our work presents extensive numerical results on the finite sample performance of the two-sample $t-$statistic-based robust inference approach in (2) with equal numbers of groups: $q_I = q_Y$ for selected inequality indices. Their study complements the analysis presented in this paper. Among other directions, the future research may focus on further analysis of finite sample performance, in the case of different dependence structures, of the robust inference approach using one-sample $t-$statistic (3) in differences of group estimates that provides a theoretically justified valid inference under arbitrary dependence between the samples.

In addition to inference on inequality indices that is the focus of this work, the $t-$statistic based robust inference approach may also be applied for inference on poverty and concentration indices where, as is well-known, the presence of extreme values, outliers, heavy-tailedness, and heterogeneity make the application of asymptotic methods problematic (see, among others, Appendix B.1 in Section E7 in Mandelbrot, 1997, Davidson and Flachaire, 2007, and Section 3.3.2 in Ibragimov et al., 2015). The method can also be used for inference on tail indices in power laws (1) and corresponding measures of top income or wealth inequality (see the discussion in the introduction, Section 3 and references therein). These and other applications of the $t-$statistic robust inference approach and its extensions are currently under development.

Acknowledgments

We thank the Editors, Esfandiar Maasoumi and Yuya Sasaki, an Associate Editor and two anonymous referees for very helpful comments and suggestions. We thank Emmanuel Flachaire for codes on permutation and bootstrap inference approaches discussed in the paper. We are also grateful to Mario Holzner, Clara Martínez-Toledano, Ulrich K. Müller, Artem Prokhorov, Alla Salmina, Yulong Wang and the participants at the seminars at Imperial College Business School, New Economic School, the Centre for Econometrics and Business Analysis (CEBA, St. Petersburg State University) seminar series, the GDN SEE and CIS Research Competition workshops, the 14th Annual Global Development Conference of the Global Development Network, the International Conferences on Computational and Financial Econometrics, the International Conferences on Modern Econometric Tools and Applications (META) and the International Conferences on Econometrics and Business Analytics (iCEBA) for helpful comments and discussions. Rustam Ibragimov thanks the Department of Theoretical Cybernetics and the Institute of Computational Mathematics and Information Technology at Kazan (Volga Region) Federal University for hospitality during his visits.

Data availability statement

The authors confirm that the data supporting the findings of this study are in its supplementary materials. These data were derived from the household income surveys conducted by the Federal State Statistics Service of Russia (Rosstat) (available in public domain at `https://www.gks.ru/free_doc/new_site/vndn-2017/index.html`; `https://www.gks.ru/free_doc/new_site/vndn-2017/OHousehold.html`).

Disclosure statement

The authors report there are no competing interests to declare.

Funding

Rustam Ibragimov's research for this paper was supported by a grant from the Russian Science Foundation (Project No. 22-18-00588).

References

Andrews, D. W. (2005), 'Cross-section regression with common shocks', *Econometrica* **73**, 1551–1585.

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**, 817—858.

Atkinson, A. B. (2008), Concentration among the rich, *in* J. B. Davies, ed., 'Personal Wealth from a Global Perspective', Oxford University Press, p. 64–89.

Bakirov, N. K. and Székely, G. J. (2006), 'Student's $t-$test for Gaussian scale mixtures', *Journal of Mathematical Sciences* **139**, 6497–6505.

Bhattacharya, D. (2007), 'Inference on inequality from household survey data', *Journal of Econometrics* **137**, 674—-707.

Blanchet, T., Garbinti, B., Goupille-Lebret, J. and Martínez-Toledano, C. (2018), 'Applying generalized Pareto curves to inequality analysis', *AEA Papers and Proceedings* **108**, 114–118.

Canay, I. A., Romano, J. P. and Shaikh, A. M. (2017), 'Randomization tests under an approximate symmetry assumption', *Econometrica* **85**, 1013–1030.

Conley, T. G. (1999), 'GMM estimation with cross sectional dependence', *Journal of Econometrics* **92**, 1—-45.

Cont, R. (2001), 'Empirical properties of asset returns: Stylized facts and statistical issues', *Quantitative Finance* **1**, 223–236.

Cowell, D. A. (1989), 'Sampling variance and decomposable inequality measures', *Journal of Econometrics* **42**, 27–41.

Cowell, F. A. (2000), Measurement of inequality, *in* A. B. Atkinson and F. Bourguignon, eds, 'Handbook of Income Distribution', Vol. 1, pp. 87–166.

Cowell, F. A. and Flachaire, E. (2007), 'Income distribution and inequality measurement: The problem of extreme values', *Journal of Econometrics* **141**, 1044–1072.

Davidson, R. and Flachaire, E. (2007), 'Asymptotic and bootstrap inference for inequality and poverty measures', *Journal of Econometrics* **141**, 141–166.

Davis, R. A. and Mikosch, T. (1998), 'The sample autocorrelations of heavy-tailed processes with applications to ARCH', *Annals of Statistics* **26**, 2049–2080.

den Haan, W. J. and Levin, A. (1997), A practitioner's guide to robust covariance matrix estimation, *in* 'Handbook of Statistics', Vol. 15, pp. 291–341.

Dufour, J.-M., Flachaire, E. and Khalaf, L. (2019), 'Permutation tests for comparing inequality measures', *Journal of Business & Economic Statistics* **37**, 457—-470.

Dufour, J.-M., Flachaire, E., Khalaf, L. and Zalghout, A. (2020), 'Identification-robust inequality analysis', *Working paper Cahier 03-2020, CIREQ* .

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer.

Esarey, J. and Menger, A. (2019), 'Practical and effective approaches to dealing with clustered data', *Political Science Research and Methods* **7**, 541–559.

Fontanari, A., Taleb, N. N. and Cirillo, P. (2018), 'Gini estimation under infinite variance', *Physica A: Statistical Mechanics and its Applications* **502**, 256–269.

Gabaix, X. (2009), 'Power laws in economics and finance', *Annual Review of Economics* **1**, 255–293.

Gabaix, X. and Ibragimov, R. (2011), 'Rank-1/2: A simple way to improve the OLS estimation of tail exponents', *Journal of Business & Economic Statistics* **29**, 24–39.

Gabaix, X., Lasry, J.-M., Lions, P.-L. and Moll, B. (2016), 'The dynamics of inequality', *Econometrica* **84**, 2071–2111.

Granger, C. W. J. and Orr, D. (1972), 'Infinite variance and research strategy in time series analysis', *Journal of the American Statistical Association* **67**, 275–285.

Hoeffding, W. (1948), 'A class of statistics with asymptotically normal distribution', *Annals of Mathematical Statistics* **19**, 293–325.

Hwang, J. (2021), 'Simple and trustworthy cluster-robust GMM inference', *Journal of Econometrics* **222**, 993–1023.

Ibragimov, M. and Ibragimov, R. (2018), 'Heavy tails and upper-tail inequality: The case of Russia', *Empirical Economics* **54**, 823–837.

Ibragimov, M., Ibragimov, R., Karimov, J. and Yuldasheva, G. (2013), Robust analysis of income inequality dynamics in russia: t-statistic based approaches.
URL: *https://wiiw.ac.at/robust-analysis-of-income-inequality-dynamics-in-russia-t-statistic-based-approaches-p-3189.html*

Ibragimov, M., Ibragimov, R. and Walden, J. (2015), *Heavy-Tailed Distributions and Robustness in Economics and Finance*, Vol. 214 of *Lecture Notes in Statistics*, Springer.

Ibragimov, R. (2007), 'Efficiency of linear estimators under heavy-tailedness: Convolutions of $\alpha-$symmetric distributions', *Econometric Theory* **23**, 501–517.

Ibragimov, R. (2009), 'Portfolio diversification and value at risk under thick-tailedness', *Quantitative Finance* **9**, 565—-580.

Ibragimov, R. and Müller, U. K. (2010), '$t-$statistic based correlation and heterogeneity robust inference', *Journal of Business and Economic Statistics* **28**, 453–468.

Ibragimov, R. and Müller, U. K. (2016), 'Inference with few heterogeneous clusters', *Review of Economics and Statistics* **98**, 83–96.

Ibragimov, R., Pedersen, R. S. and Skrobotov (2020), 'New approaches to robust inference on market (non-)efficiency, volatility clustering and nonlinear dependence', *Journal of Financial Econometrics* **24**, 1075–1097.

Kleiber, C. and Kotz, S. (2003), *Statistical Size Distributions in Economics and Actuarial Science*, Wiley Series in Probability and Statistics.

Koroljuk, V. S. and Borovskich, Y. V. (1994), *Theory of U-statistics*, Vol. 273 of *Mathematics and its Applications*, Kluwer Academic Publishers Group, Dordrecht.

Mandelbrot, B. (1997), *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, Springer.

Marshall, A. W., Olkin, I. and Arnold, B. C. (2011), *Inequalities: Theory of Majorization and Its Applications*, Springer Series in Statistics, second edn, Springer, New York.

McNeil, A. J., Frey, R. and Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools*, revised edition edn, Springer.

Mergane, P. D., Kpanzou, T. A., Ba, D. and Lo, G. S. (2018), 'A Theil-like class of inequality measures, its asymptotic normality theory and applications', *Afrika Statistika* **13**, 1699–1715.

Midões and de Crombrugghe, D. (2023), 'Assumption-light and computationally cheap inference on inequality measures by sample splitting: the Student t approach', *Journal of Economic Inequality* **21**, 899–924.

Mikosch, T. and Stărică, C. (2000), 'Limit theory for the sample autocorrelations and extremes of a GARCH $(1,1)$ process', *Annals of Statistics* **28**, 1427–1451.

Mills, J. A. and Zandvakili, S. (1997), 'Statistical inference via bootstrapping for measures of inequality', *Journal of Applied Econometrics* **12**, 133–150.

Phillips, P. C. B. (2005), 'Automated discovery in econometrics', *Econometric Theory* **21**, 3– 20.

Piketty, T. and Saez, E. (2003), 'Income inequality in the United States, 1913-1998', *Quarterly Journal of Economics* **118**, 1–39.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York.

Toda, A. A. (2012), 'The double power law in income distribution: Explanations and evidence', *Journal of Economic Behavior and Organization* **84**, 364–381.

Toda, A. A. and Wang, Y. (2021), 'Efficient minimum distance estimation of Pareto exponent from top income shares', *Journal of Applied Econometrics* **36**, 228–243.

APPENDICES

Appendix A: Tables

Table A.1: Empirical results: $p-$values, Gini measure

	Altai region	Krasnodar region	Krasnoyarsk region	Primorsky Krai	Stavropol region	Khabarovsk region	Amur region	Arkhangelsk region	Astrakhan region	Nenets Autonomous Okrug	Belgorod region	Bryansk region	Vladimir region	Volgograd region	Vologda Region	Voronezh region	Nizhny Novgorod Region
Gini	0.273	0.291	0.327	0.284	0.269	0.315	0.300	0.274	0.280	0.322	0.253	0.243	0.252	0.256	0.261	0.258	0.265
N_Y	2568	4392	2832	2160	2400	1608	1296	1488	1416	480	1752	1584	1752	2520	1512	2496	3360
N_I/N_Y	3.50	2.05	3.18	4.17	3.75	5.60	6.94	6.05	6.36	18.75	5.14	5.68	5.14	3.57	5.95	3.61	2.68
ζ	3.50	2.60	3.03	4.31	3.93	3.99	5.04	4.10	4.37	4.77	3.74	4.60	3.67	3.57	3.92	3.76	3.93
asymptotic	0.18	**0.00**	**0.00**	**0.00**	0.44	**0.00**	**0.00**	0.14	**0.02**	**0.00**	0.08	**0.00**	0.07	0.16	0.70	0.28	0.92
$q_I = q_Y = 4$	0.25	**0.04**	**0.00**	0.06	0.52	**0.00**	**0.02**	0.38	0.08	**0.04**	0.31	0.07	0.31	0.38	0.73	0.28	0.92
$q_I = q_Y = 8$	0.30	**0.02**	**0.00**	**0.01**	0.50	**0.00**	**0.00**	0.34	**0.02**	**0.00**	0.22	**0.02**	0.16	0.23	0.69	0.42	0.96
$q_I = q_Y = 12$	0.27	**0.01**	**0.00**	**0.01**	0.43	**0.00**	**0.00**	0.28	**0.02**	**0.00**	0.17	**0.00**	0.14	0.15	0.69	0.33	0.98
$q_I = q_Y = 16$	0.15	**0.00**	**0.00**	**0.01**	0.35	**0.00**	**0.00**	0.31	**0.02**	**0.00**	**0.02**	**0.00**	0.07	0.11	0.63	0.28	0.97
$q_I = 4, q_Y = 3$	0.33	0.13	**0.02**	**0.03**	0.57	**0.01**	**0.03**	0.50	0.22	**0.04**	0.37	0.06	0.35	0.34	0.68	0.29	0.90
$q_I = 8, q_Y = 6$	0.34	**0.03**	**0.00**	**0.03**	0.40	**0.00**	**0.00**	0.31	0.09	**0.01**	0.26	**0.01**	0.19	0.23	0.69	0.22	0.95
$q_I = 12, q_Y = 9$	0.23	**0.01**	**0.00**	**0.05**	0.30	**0.00**	**0.00**	0.23	**0.03**	**0.00**	**0.03**	**0.02**	0.17	0.19	0.71	0.37	0.97
$q_I = 16, q_Y = 12$	0.27	**0.01**	**0.00**	**0.01**	0.43	**0.00**	**0.00**	0.28	**0.02**	**0.00**	0.16	**0.00**	0.13	0.15	0.68	0.32	1.00
$q_I = 4, q_Y = 2$	0.48	0.20	0.10	0.19	0.61	0.09	0.18	0.46	0.15	0.12	0.55	0.11	0.21	0.60	0.60	0.47	0.95
$q_I = 8, q_Y = 4$	0.22	**0.04**	**0.00**	0.05	0.50	**0.00**	**0.02**	0.37	0.07	**0.04**	0.3	0.07	0.30	0.36	0.71	0.24	0.92
$q_I = 12, q_Y = 6$	0.33	**0.03**	**0.00**	**0.02**	0.38	**0.00**	**0.00**	0.30	0.08	**0.01**	0.26	**0.01**	0.19	0.22	0.68	0.20	0.94
$q_I = 16, q_Y = 8$	0.29	**0.02**	**0.00**	**0.01**	0.49	**0.00**	**0.00**	0.33	**0.02**	**0.00**	0.21	**0.02**	0.16	0.21	0.68	0.41	0.96
$q_I = 8, q_Y = 2$	0.47	0.19	0.09	0.18	0.60	0.09	0.18	0.45	0.13	0.12	0.55	0.09	0.18	0.60	0.54	0.45	0.95
$q_I = 12, q_Y = 3$	0.29	0.13	**0.02**	**0.01**	0.54	**0.01**	**0.02**	0.49	0.21	**0.04**	0.35	0.05	0.34	0.31	0.64	0.21	0.89
$q_I = 16, q_Y = 4$	0.21	**0.04**	**0.00**	**0.05**	0.48	**0.00**	**0.02**	0.37	0.07	**0.04**	0.29	0.06	0.29	0.35	0.70	0.21	0.92
permutation	0.15	**0.01**	**0.00**	**0.00**	0.48	**0.00**	**0.00**	0.15	**0.02**	**0.00**	0.08	**0.00**	0.08	0.19	0.72	0.26	0.92
bootstrap	0.20	**0.00**	**0.00**	**0.00**	0.43	**0.00**	**0.00**	0.15	**0.02**	**0.00**	0.08	**0.00**	0.10	0.17	0.73	0.33	0.87

Table A.1: Empirical results: p-values, Gini measure, Continued

	Ivanovo region	Irkutsk region	The Republic of Ingushetia	Kaliningrad region	Tver region	Kaluga region	Kamchatka Krai	Kemerovo region	Kirov region	Kostroma region	Samara Region	Kurgan region	Kursk region	Saint Petersburg city	Leningrad region	Lipetsk region
Gini	0.256	0.309	0.354	0.276	0.260	0.256	0.300	0.273	0.259	0.247	0.268	0.287	0.262	0.264	0.269	0.250
N_Y	1512	2448	600	1368	1704	1440	888	2808	1680	1248	3168	1368	1488	4344	1944	1488
N_I/N_Y	5.95	3.68	15.00	6.58	5.28	6.25	10.14	3.21	5.36	7.21	2.84	6.58	6.05	2.07	4.63	6.05
ζ	3.77	4.51	3.69	5.05	3.94	6.00	3.34	3.76	5.15	4.77	4.49	3.18	2.75	4.27	5.63	3.67
asymptotic	0.2	**0.00**	**0.00**	**0.05**	0.48	0.17	**0.00**	0.12	0.42	**0.01**	0.41	**0.01**	0.86	0.98	0.39	0.06
$q_I = q_Y = 4$	0.24	**0.01**	**0.01**	0.10	0.59	0.36	**0.02**	0.12	0.38	0.06	0.54	0.05	0.80	0.97	0.53	0.30
$q_I = q_Y = 8$	0.14	**0.00**	**0.00**	**0.02**	0.57	0.23	**0.01**	0.10	0.36	**0.02**	0.57	**0.02**	0.76	0.96	0.37	0.12
$q_I = q_Y = 12$	0.11	**0.00**	**0.00**	0.07	0.50	0.23	**0.03**	**0.03**	0.38	**0.01**	0.47	**0.02**	0.69	0.94	0.24	0.09
$q_I = q_Y = 16$	0.1	**0.00**	**0.00**	**0.02**	0.46	0.15	**0.00**	0.10	0.35	**0.01**	0.59	**0.01**	0.66	0.96	0.46	0.08
$q_I = 4, q_Y = 3$	0.27	**0.03**	**0.00**	0.18	0.76	0.47	0.11	0.16	0.34	0.07	0.72	**0.03**	0.82	0.98	0.61	0.43
$q_I = 8, q_Y = 6$	0.13	**0.00**	**0.00**	0.11	0.62	0.26	0.05	0.07	0.3	**0.04**	0.58	**0.05**	0.80	0.98	0.46	0.2
$q_I = 12, q_Y = 9$	0.15	**0.00**	**0.00**	**0.05**	0.57	0.25	**0.01**	**0.02**	0.46	**0.02**	0.59	**0.01**	0.80	0.96	0.42	0.13
$q_I = 16, q_Y = 12$	0.11	**0.00**	**0.00**	0.07	0.49	0.22	**0.03**	**0.03**	0.36	**0.01**	0.48	**0.02**	0.68	0.96	0.25	0.09
$q_I = 4, q_Y = 2$	0.26	0.08	0.07	0.35	0.47	0.44	0.17	0.31	0.47	0.23	0.59	0.19	0.80	0.98	0.72	0.42
$q_I = 8, q_Y = 4$	0.21	**0.01**	**0.01**	0.08	0.57	0.35	**0.02**	0.09	0.32	0.05	0.52	**0.05**	0.78	0.98	0.51	0.29
$q_I = 12, q_Y = 6$	0.12	**0.00**	**0.00**	0.10	0.62	0.24	0.05	0.05	0.27	**0.04**	0.56	**0.04**	0.80	0.97	0.44	0.19
$q_I = 16, q_Y = 8$	0.12	**0.00**	**0.00**	**0.02**	0.56	0.22	**0.01**	0.08	0.33	**0.01**	0.56	**0.02**	0.75	0.97	0.35	0.11
$q_I = 8, q_Y = 2$	0.21	0.07	0.07	0.33	0.42	0.42	0.16	0.28	0.42	0.21	0.57	0.18	0.78	0.98	0.72	0.42
$q_I = 12, q_Y = 3$	0.22	**0.03**	**0.00**	0.16	0.75	0.45	0.11	0.10	0.24	0.05	0.71	**0.02**	0.80	0.97	0.58	0.42
$q_I = 16, q_Y = 4$	0.19	**0.01**	**0.01**	0.07	0.56	0.34	**0.02**	0.07	0.28	**0.05**	0.5	**0.04**	0.76	0.98	0.5	0.28
permutation	0.22	**0.00**	**0.00**	**0.04**	0.49	0.18	**0.00**	0.12	0.46	**0.04**	0.39	**0.01**	0.85	0.94	0.33	0.08
bootstrap	0.18	**0.00**	**0.00**	**0.04**	0.50	0.19	**0.00**	0.11	0.44	**0.02**	0.44	**0.01**	0.88	0.99	0.36	0.09

Table A.1: Empirical results: p-values, Gini measure, Continued

	Magadan Region	Moscow city	Moscow region	Murmansk region	Novgorod region	Novosibirsk region	Omsk region	Orenburg region	Oryol Region	Penza region	Perm region	Pskov region	Rostov region	Ryazan Oblast	Saratov region	Sakhalin Region	Sverdlovsk region
Gini	0.303	0.264	0.298	0.290	0.252	0.305	0.294	0.270	0.257	0.248	0.273	0.259	0.262	0.258	0.258	0.326	0.278
N_Y	720	9000	5952	1296	1176	2760	2088	2160	1248	1656	2736	1248	3816	1488	2640	1080	4152
N_I/N_Y	12.50	1.00	1.51	6.94	7.65	3.26	4.31	4.17	7.21	5.43	3.29	7.21	2.36	6.05	3.41	8.33	2.17
ζ	4.14	3.96	4.70	4.21	4.77	3.50	3.45	4.53	3.11	3.06	4.73	3.97	4.72	3.06	4.24	3.59	3.24
asymptotic	**0.00**		**0.00**	**0.00**	0.06	**0.00**	**0.00**	0.27	0.43	**0.03**	0.07	0.49	0.57	0.47	0.22	**0.00**	**0.02**
$q_I = q_Y = 4$	**0.01**		**0.01**	0.06	0.11	**0.00**	**0.02**	0.41	0.47	0.19	0.23	0.37	0.64	0.55	0.34	**0.01**	0.12
$q_I = q_Y = 8$	**0.00**		**0.00**	**0.01**	**0.03**	**0.00**	**0.00**	0.35	0.42	0.09	0.18	0.43	0.58	0.48	0.23	**0.00**	**0.04**
$q_I = q_Y = 12$	**0.00**		**0.00**	**0.02**	0.07	**0.00**	**0.00**	0.32	0.44	0.08	0.11	0.4	0.57	0.44	0.24	**0.00**	**0.05**
$q_I = q_Y = 16$	**0.00**		**0.00**	**0.01**	**0.02**	**0.00**	**0.00**	0.31	0.39	**0.03**	0.08	0.35	0.52	0.37	0.17	**0.00**	0.06
$q_I = 4, q_Y = 3$	**0.03**		**0.04**	0.12	0.16	**0.02**	**0.02**	0.58	0.52	0.35	0.31	0.57	0.69	0.31	0.47	**0.03**	0.23
$q_I = 8, q_Y = 6$	**0.00**		**0.00**	**0.03**	0.11	**0.00**	**0.00**	0.38	0.49	0.14	0.17	0.40	0.59	0.49	0.31	**0.00**	0.07
$q_I = 12, q_Y = 9$	**0.00**		**0.00**	**0.04**	0.07	**0.00**	**0.00**	0.28	0.44	0.10	0.11	0.44	0.62	0.45	0.23	**0.00**	**0.04**
$q_I = 16, q_Y = 12$	**0.00**		**0.00**	**0.02**	0.07	**0.00**	**0.00**	0.32	0.43	0.08	0.11	0.39	0.56	0.43	0.23	**0.00**	0.05
$q_I = 4, q_Y = 2$	0.09		0.07	0.21	0.23	0.08	0.15	0.61	0.5	0.34	0.48	0.42	0.79	0.64	0.49	**0.04**	0.30
$q_I = 8, q_Y = 4$	**0.01**		**0.01**	0.06	0.09	**0.00**	**0.02**	0.39	0.46	0.18	0.21	0.31	0.61	0.54	0.31	**0.01**	0.11
$q_I = 12, q_Y = 6$	**0.00**		**0.00**	**0.02**	0.11	**0.00**	**0.00**	0.37	0.49	0.14	0.16	0.38	0.58	0.48	0.3	**0.00**	0.06
$q_I = 16, q_Y = 8$	**0.00**		**0.00**	**0.01**	**0.02**	**0.00**	**0.00**	0.33	0.41	0.09	0.16	0.42	0.55	0.47	0.2	**0.00**	**0.03**
$q_I = 8, q_Y = 2$	0.08		0.06	0.20	0.20	0.08	0.14	0.60	0.48	0.34	0.47	0.35	0.78	0.63	0.46	**0.03**	0.28
$q_I = 12, q_Y = 3$	**0.02**		**0.03**	0.11	0.13	**0.02**	**0.01**	0.57	0.50	0.35	0.29	0.55	0.66	0.25	0.45	**0.03**	0.21
$q_I = 16, q_Y = 4$	**0.01**		**0.00**	0.06	0.08	**0.00**	**0.02**	0.38	0.44	0.17	0.2	0.26	0.58	0.53	0.29	**0.01**	0.10
permutation	**0.00**		**0.00**	**0.00**	0.06	**0.00**	**0.00**	0.25	0.45	**0.02**	0.07	0.5	0.58	0.52	0.21	**0.00**	**0.03**
bootstrap	**0.00**		**0.00**	**0.00**	0.08	**0.00**	**0.00**	0.26	0.41	**0.03**	0.07	0.52	0.6	0.46	0.23	**0.00**	**0.02**

Table A.1: Empirical results: p-values, Gini measure, Continued

	Smolensk region	Sevastopol city	Tambov Region	Tomsk region	Tula region	Tyumen region	Khanty-Mansi Autonomous Okrug - Yugra	Ulyanovsk region	Yamalo-Nenets Autonomous District	Chelyabinsk region	Zabaykalsky Krai	Chukotka Autonomous District	Yaroslavskaya oblast	Republic of Adygea	Republic of Bashkortostan	The Republic of Buryatia	The Republic of Dagestan
Gini	0.251	0.240	0.236	0.315	0.248	0.324	0.293	0.245	0.335	0.283	0.317	0.328	0.260	0.267	0.292	0.327	0.318
N_Y	1440	792	1488	1416	1848	1560	1656	1560	1008	3408	1368	528	1632	936	3624	1272	1968
N_I/N_Y	6.25	11.36	6.05	6.36	4.87	5.77	5.43	5.77	8.93	2.64	6.58	17.05	5.51	9.62	2.48	7.08	4.57
ζ	4.41	5.78	4.32	3.13	5.40	2.71	4.11	4.36	5.46	4.00	5.17	4.64	4.19	5.46	4.48	3.53	4.26
asymptotic	**0.05**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.46	0.7	**0.00**	**0.00**	**0.00**
$q_I = q_Y = 4$	0.12	**0.03**	**0.02**	**0.01**	**0.03**	**0.01**	**0.01**	0.08	**0.00**	**0.02**	**0.02**	**0.00**	0.42	0.83	**0.01**	**0.01**	**0.03**
$q_I = q_Y = 8$	0.08	**0.01**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.03**	**0.00**	**0.02**	**0.00**	**0.00**	0.26	0.81	**0.00**	**0.00**	**0.00**
$q_I = q_Y = 12$	**0.03**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	0.28	0.97	**0.00**	**0.00**	**0.00**
$q_I = q_Y = 16$	**0.04**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.30	0.78	**0.00**	**0.00**	**0.00**
$q_I = 4, q_Y = 3$	0.24	0.05	**0.02**	**0.01**	0.08	**0.01**	**0.02**	0.12	**0.01**	**0.04**	0.07	**0.02**	0.33	0.57	0.06	**0.01**	0.06
$q_I = 8, q_Y = 6$	0.08	**0.02**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	0.06	**0.00**	**0.01**	**0.01**	**0.00**	0.3	0.8	**0.00**	**0.00**	**0.00**
$q_I = 12, q_Y = 9$	0.11	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**	**0.00**	0.33	0.8	**0.00**	**0.00**	**0.00**
$q_I = 16, q_Y = 12$	**0.03**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	0.27	0.98	**0.00**	**0.00**	**0.00**
$q_I = 4, q_Y = 2$	0.29	0.1	0.12	0.08	0.15	0.07	0.11	0.23	0.1	0.13	0.08	**0.04**	0.45	0.56	0.12	0.07	0.19
$q_I = 8, q_Y = 4$	0.11	**0.02**	**0.02**	**0.01**	**0.02**	**0.01**	**0.00**	0.07	**0.00**	**0.01**	**0.02**	**0.00**	0.38	0.82	**0.01**	**0.01**	**0.03**
$q_I = 12, q_Y = 6$	0.07	**0.02**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	0.05	**0.00**	**0.01**	**0.01**	**0.00**	0.27	0.79	**0.00**	**0.00**	**0.00**
$q_I = 16, q_Y = 8$	0.07	**0.01**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.03**	**0.00**	**0.01**	**0.00**	**0.00**	0.21	0.81	**0.00**	**0.00**	**0.00**
$q_I = 8, q_Y = 2$	0.28	0.08	0.11	0.08	0.13	0.07	0.1	0.22	0.1	0.12	0.08	**0.04**	0.39	0.5	0.11	0.07	0.19
$q_I = 12, q_Y = 3$	0.22	**0.04**	**0.02**	**0.01**	0.06	**0.01**	**0.01**	0.11	**0.00**	**0.03**	0.07	**0.02**	0.20	0.46	0.05	**0.01**	0.06
$q_I = 16, q_Y = 4$	0.10	**0.02**	**0.02**	**0.01**	**0.02**	**0.01**	**0.00**	0.06	**0.00**	**0.01**	**0.02**	**0.00**	0.35	0.82	**0.01**	**0.01**	**0.03**
permutation	0.08	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.51	0.70	**0.00**	**0.00**	**0.00**
bootstrap	0.07	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	0.47	0.63	**0.00**	**0.00**	**0.00**

Table A.1: Empirical results: p-values, Gini measure, Continued

	Kabardino-Balkar Republic	Altai Republic	Republic of Kalmykia	Republic of Karelia	Komi Republic	Mari El Republic	The Republic of Mordovia	Republic of North Ossetia-Alania	Karachay-Cherkess Republic	Republic of Tatarstan	Tyva Republic	Udmurtia	The Republic of Khakassia	Chechen Republic	Chuvash Republic - Chuvashia	The Republic of Sakha (Yakutia)	Jewish Autonomous Region
Gini	0.298	0.340	0.326	0.268	0.298	0.286	0.251	0.277	0.334	0.248	0.423	0.251	0.293	0.327	0.267	0.341	0.334
N_Y	960	600	720	1200	1392	1104	1248	912	720	3408	744	1704	1032	1032	1488	1320	600
N_I/N_Y	9.38	15.00	12.50	7.50	6.47	8.15	7.21	9.87	12.50	2.64	12.10	5.28	8.72	8.72	6.05	6.82	15.00
ζ	4.28	3.35	3.47	3.78	4.14	2.29	3.12	3.34	2.08	4.49	3.04	4.50	5.14	3.30	5.27	4.34	4.27
asymptotic	**0.00**	**0.00**	**0.00**	0.64	**0.00**	0.06	0.14	0.18	**0.00**	**0.00**	**0.00**	**0.03**	**0.00**	**0.00**	0.62	**0.00**	**0.00**
$q_I = q_Y = 4$	**0.04**	**0.00**	**0.01**	0.71	**0.03**	0.14	0.28	0.2	**0.03**	**0.03**	**0.00**	0.11	**0.03**	**0.02**	0.63	**0.00**	**0.05**
$q_I = q_Y = 8$	**0.01**	**0.00**	**0.00**	0.65	**0.00**	0.15	0.2	0.21	**0.01**	**0.01**	**0.00**	**0.02**	**0.01**	**0.00**	0.63	**0.00**	**0.01**
$q_I = q_Y = 12$	**0.00**	**0.00**	**0.00**	0.58	**0.00**	0.12	0.11	0.27	**0.01**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**	0.64	**0.00**	**0.00**
$q_I = q_Y = 16$	**0.00**	**0.00**	**0.00**	0.63	**0.00**	0.16	0.14	0.23	**0.01**	**0.00**	**0.00**	0.08	**0.00**	**0.00**	0.68	**0.00**	**0.00**
$q_I = 4, q_Y = 3$	**0.02**	**0.00**	**0.03**	0.68	0.06	0.13	0.22	0.43	**0.01**	0.13	**0.01**	0.08	0.09	0.05	0.55	**0.00**	0.10
$q_I = 8, q_Y = 6$	**0.00**	**0.00**	**0.00**	0.59	**0.00**	**0.03**	0.2	0.32	**0.01**	**0.03**	**0.00**	0.08	**0.01**	**0.01**	0.6	**0.00**	**0.01**
$q_I = 12, q_Y = 9$	**0.00**	**0.00**	**0.00**	0.62	**0.00**	0.14	0.12	0.33	**0.00**	**0.01**	**0.00**	0.07	**0.01**	**0.00**	0.8	**0.00**	**0.01**
$q_I = 16, q_Y = 12$	**0.00**	**0.00**	**0.00**	0.59	**0.00**	0.12	0.11	0.27	**0.01**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**	0.65	**0.00**	**0.00**
$q_I = 4, q_Y = 2$	0.16	**0.03**	0.14	0.48	0.16	0.16	0.18	0.21	**0.05**	0.18	0.13	0.28	0.12	**0.05**	0.54	0.06	0.14
$q_I = 8, q_Y = 4$	**0.04**	**0.00**	**0.01**	0.71	**0.03**	0.13	0.27	0.18	**0.03**	**0.02**	**0.00**	0.09	**0.03**	**0.02**	0.62	**0.00**	**0.05**
$q_I = 12, q_Y = 6$	**0.00**	**0.00**	**0.00**	0.58	**0.00**	**0.03**	0.19	0.31	**0.01**	**0.02**	**0.00**	0.07	**0.01**	**0.01**	0.57	**0.00**	**0.01**
$q_I = 16, q_Y = 8$	**0.01**	**0.00**	**0.00**	0.65	**0.00**	0.15	0.19	0.21	**0.01**	**0.01**	**0.00**	**0.01**	**0.00**	**0.00**	0.61	**0.00**	**0.01**
$q_I = 8, q_Y = 2$	0.15	**0.03**	0.13	0.41	0.15	0.15	0.14	0.19	**0.04**	0.16	0.13	0.26	0.12	**0.04**	0.48	0.06	0.14
$q_I = 12, q_Y = 3$	**0.01**	**0.00**	**0.03**	0.66	0.06	0.12	0.20	0.42	**0.01**	0.12	**0.01**	**0.05**	0.09	0.05	0.45	**0.00**	0.10
$q_I = 16, q_Y = 4$	**0.03**	**0.00**	**0.01**	0.71	**0.03**	0.13	0.26	0.17	**0.03**	**0.02**	**0.00**	0.08	**0.03**	**0.02**	0.61	**0.00**	**0.05**
permutation	**0.00**	**0.00**	**0.00**	0.59	**0.00**	**0.04**	0.18	0.18	**0.00**	**0.00**	**0.00**	**0.04**	**0.00**	**0.00**	0.6	**0.00**	**0.00**
bootstrap	**0.00**	**0.00**	**0.00**	0.63	**0.00**	**0.05**	0.18	0.17	**0.00**	**0.00**	**0.00**	0.05	**0.00**	**0.00**	0.64	**0.00**	**0.00**

Appendix B: Inequality measures and their sample analogues. Normal and heavy-tailed stable asymptotics.

In this section, we review the definitions of the widely used Gini and Theil inequality measures, sample analogues of the measures and their asymptotic properties (see, among others, Sections 2.1.3 and 2.1.4 in Kleiber and Kotz, 2003, Cowell and Flachaire, 2007, Davidson and Flachaire, 2007, Section 13.F, 17.C in Marshall et al., 2011, and references therein).

Let I be an (absolutely continuous) nonnegative r.v. (e.g., income or wealth level) with the finite first moment $\mu_I = E[I] < \infty$ and the cdf $F_I(x)$ representing income or wealth distribution in a population, and let $I_1, I_2, ..., I_N$ denote a sample of observations on the r.v. I.

As usual, we denote by $\overline{I}_N = N^{-1}\sum_{i=1}^N I_i$ and $s_N^2 = (N-1)^{-1}\sum_{i=1}^N (I_i - \overline{I})^2$ the sample mean and sample variance of the observations I_i.

Below, we provide the definitions of Theil and Gini inequality measures (denoted by $\mathcal{L}^I_{Theil}$ and $\mathcal{L}^I_{Gini}$ for the population considered) and discuss the standard results on their asymptotic normality.

Theil index The population Theil index is defined by

$$\mathcal{L}^I_{Theil} = \frac{E[I \log I]}{\mu_I} - \log(\mu_I).$$

The Theil index is the limiting case of the Generalized Entropy measures. Its sample analogue - sample Theil index - is given by

$$\hat{\mathcal{L}}^I_{Theil,N} = \frac{\frac{1}{N}\sum_{i=1}^N I_i \log(I_i)}{\overline{I}_N} - \log(\overline{I}_N).$$

Under i.i.d. observations $I_1, I_2, ..., I_N$, the Theil index is asymptotically normal if $E[I^2] < \infty$, $E[I^2 \log I] < \infty$ and $E[I^2 \log^2(I)] < \infty$. It is easy to see that these conditions are satisfied in the case of r.v.'s with power law distributions (1) (e.g., S-M distributions $SM(a, b, c)$ in (4) with $\zeta = ac$) if the tail index ζ is greater than 2: $\zeta > 2$.

Under the above conditions, one has

$$\sqrt{N}(\hat{\mathcal{L}}^I_{Theil,N} - \mathcal{L}^I_{Theil}) \to_w N(0, v^2_{Theil,I}),$$

where

$$v^2_{Theil,I} = \frac{E[I^2 \log^2 I]}{\mu_I^2} + \frac{E[I^2]}{\mu_I^2}\Big(\frac{E[I \log I]}{\mu_I} + 1\Big)^2 - \frac{2E[I^2 \log I]}{\mu_I^2}\Big(\frac{E[I \log I]}{\mu_I} + 1\Big) - 1$$

(see, among others, Mills and Zandvakili (1997), Cowell (1989, 2000), Cowell and Flachaire (2007) and Mergane et al. (2018) for the review of the results on asymptotic normality and the formulas for the liming and sampling variance of different estimators of inequality measures).

Gini coefficient The population Gini coefficient is defined by

$$\mathcal{L}^I_{Gini} = 0.5\frac{E|I' - I''|}{\mu_I},$$

where I' and I'' are independent copies of the r.v. I.

The most commonly used (nonparametric) estimator of the Gini coefficient $\mathcal{L}^I_{Gini}$ is given by its sample analogue (the sample Gini coefficient)

$$\hat{\mathcal{L}}^I_{Gini,N} = \frac{\sum_{1 \le i < j \le N} |I_i - I_j|}{(N-1)\sum_{i=1}^N I_i} = U_N/\overline{I}_N,$$

where U_N is the $U-$statistic $U_N = \frac{2}{N(N-1)}\sum_{1 \le i < j \le N} |I_i - I_j|$ (we refer to, among others, Hoeffding (1948), Ch. 5 in Serfling (1980) and Ch. 4 in Koroljuk and Borovskich (1994) for the asymptotic theory for general $U-$statistics).

From the results in the above references, it follows that asymptotic normality for the $U-$statistic U_N and the sample Gini coefficient holds if $I_1, I_2, ..., I_N$ are i.i.d. observations with finite second moment $E[I^2] < \infty$. This holds under power-law distributions (1) (e.g., for S-M distributions $SM(a,b,c)$ in (4) with $\zeta = ac$) if the tail index ζ is greater than 2: $\zeta > 2$. More precisely, under the above conditions (see Hoeffding (1948))

$$\sqrt{N}(\hat{\mathcal{L}}^I_{Gini,N} - \mathcal{L}^I_{Gini}) \to_w N(0, v^2_{Gini,I}),$$

where $v^2_{Gini,I} = (\mathcal{L}^I_{Gini})^2\sigma^2_I - 2\mathcal{L}^I_{Gini}E\{I'|I' - I''|\}/\mu^2_I + E(E_{I'}\{|I' - I''|\})/\mu^2_I$, and $E_{I'}(\cdot) = E_{I'}(\cdot) = E\{\cdot|I'\}$ denotes the expectation conditional on I'.

Naturally, the asymptotic normality of the sample Theil and Gini coefficients is lost under infinite second moments and variances: $E[I^2] = \infty$. For instance, from the results in Fontanari et al. (2018) it follows that under i.i.d. observations $I_1, I_2, ..., I_N$ that follow a power-law distribution (1) with the tail index $\zeta \in (1,2)$ (e.g., the S-M distribution $SM(a,b,c)$ in (4) with $1 < \zeta = ac < 2$) and have finite first and infinite second moments, the sample Gini coefficient $\hat{\mathcal{L}}_{Gini,N}$ has an asymptotic right-skewed stable distribution with the index of stability ζ. Using the standard generalized CLT and the delta-method, it is also not difficult to show that in the case of distributions exhibiting (double) power law behavior in both the lower and the upper (with the tail index ζ), similar to S-M distributions $SM(a,b,c)$ with $\zeta = ac$, the sample Theil index $\hat{\mathcal{L}}_{Theil,N}$ weakly converges to a function of stable r.v.'s with indices of stability that depend on ζ. The rate of convergence in the above asymptotic results is slower than $\sqrt{N}$ and depends on ζ. The fact that the tail index ζ is unknown in practice makes the results useless for (direct) asymptotic inference.[30]

On the other hand, from the above results it also follows that, for the two-sample problem of $t-$statistic inference in Section 2.2, in the case of identical distributions in the populations considered and the equal number of groups $q_I = q_Y = q$, as in Table C.3, the differences of group estimators of Theil and Gini indices in the two-samples are asymptotically symmetric stable and thus asymptotically scale mixtures of normals. As the $t-$statistic approach to robust inference in Ibragimov and Müller (2010, 2016) are asymptotically valid under weak convergence of group estimators of parameters dealt with to scale mixtures of normal distributions, this explains favorable performance of two-sample $t-$statistic approach in Table C.3 in Section 3.2.2.

[30]The situation is somewhat similar to the properties of autocorrelation functions of GARCH-type processes and their squares, where asymptotic normality is lost under tail indices smaller than 4 and infinite fourth moments, as is typically the case for financial returns and foreign exchange rates in real-world markets (see Davis and Mikosch (1998), Mikosch and Stărică (2000) and also Ibragimov et al. (2020) for asymptotically valid robust $t-$statistic approach to inference on measures of market (non-)efficiency and volatility clustering based on powers of absolute values of GARCH-type processes, e.g., financial returns).

Appendix C: Two-sample problem; the case of identical distributions.

In the appendix, we provide the results of the size of the tests considered in Section 3.2 in the case of identical distributions.

Table C.1 presents results for identical distributions, and sample sizes $N_I = N_Y = N = 200$. The results suggest that the sizes of all the tests, except the asymptotic, never exceed the nominal 5% level. In many cases, the sizes are quite close to the nominal level for the permutation, bootstrap, and the $t-$statistic based robust tests.

Tables C.2 and C.3 present results for different sample sizes N_I, N_Y and identical distributions. The finite-sample sizes of all the tests, except the asymptotic, appear to be good for all parameter settings. Essentially no over-rejections are observed for $t-$statistic based robust inference, including in settings with more pronounced heavy-tailedness and infinite variances (Table C.3). The finite-sample size properties of the simple-to-use $t-$statistic based robust tests based on (2) with $q = 4, 8$ in the case of more heavy-tailed distributions, and based on (3) with $q = 4, 8, 12, 16$ in the case of less heavy-tailed distributions, are comparable to those of the computationally expensive permutation and bootstrap tests.

are comparable to those of the bootstrap and permutation approaches. Asymptotic normality of estimators of Theil and Gini indices is lost under infinite variances, e.g., for tail indices $\zeta_I = \zeta_Y = 1.4$ in Table C.3 (see Fontanari et al. (2018) and the discussion in Appendix B). However, the $t-$statistic based approach has good finite sample size properties even in such heavy-tailed settings. This is due to the robustness of the approach to heavy-tailedness – they may be used under convergence of group estimators to scale mixtures of normals such as symmetric stable distributions. Symmetric stable asymptotics holds for differences of group estimators in the case of identical heavy-tailed power law distributions (1) with $\zeta < 2$ and infinite second moments in the populations, and equal number of groups, such as is used in $t-$statistic inference in the two-sample problem (Table C.3, see Appendix B).

Table C.4 is an analogue of Tables C.2 and C.3 with different numbers of groups used in two-sample $t-$statistic based robust inference approach. In the case of the Theil index, only the choice of $q_I = q_Y = 4$ leads to size control for all sample sizes. Size distortion is apparently due to skewness in finite-sample distributions of (group) Theil inequality estimates implying poor quality of normal approximations (see Section 3.1 and also the discussion in Appendix D). The solution is to use different numbers of groups q_I, q_Y for different sample size pairs N_I, N_Y. Good size properties are observed with $(q_I, q_Y) = (8, 8)$ for $N_I = 200, N_Y = 400$, $(q_I, q_Y) = (6, 8)$ for $N_I = 200, N_Y = 600$ and $(q_I, q_Y) = (6, 12)$ for $N_I = 200, N_Y = 800$.

The finite-sample distributions of (group) empirical Gini estimates are not as skewed and are better approximated by the normal when compared to the Theil estimates (see Appendix C). There is good size control for different combinations of q_I and q_Y in $t-$statistic based robust tests for Gini indices except for the cases $q_I = q_Y = 12$, $q_I = q_Y = 16$ and $q_I = 12, q_Y = 16$ with rather small number of observations in each of the group. To avoid very conservative size properties, the best choices for the number of groups in applications appear to be $(q_I, q_Y) = (8, 8)$, $(q_I, q_Y) = (9, 12)$ and $(q_I, q_Y) = (8, 16)$ for all sample sizes N_I, N_Y considered.

Table C.1: Empirical size – identical sample sizes, identical distributions

Theil $\backslash\zeta$	6.26	3.94	2.9	2.2	1.4	Gini$\backslash\zeta$	6.6	3.8	2.59	2.2	1.4
asymptotic	5.4	5.5	5.3	7.6	22.1	asy	5.7	6.1	7.0	8.4	18.8
$\tilde{t}_{\mathcal{L}}(q=4)$	2.0	1.7	1.5	1.5	1.3	$\tilde{t}_{\mathcal{L}}(q=4)$	2.1	1.9	1.9	2.0	2.0
$\tilde{t}_{\mathcal{L}}(q=8)$	3.2	2.7	2.2	2.4	2.4	$\tilde{t}_{\mathcal{L}}(q=8)$	3.5	3.4	3.0	3.2	3.4
$\tilde{t}_{\mathcal{L}}(q=12)$	3.6	3.2	2.7	2.9	3.0	$\tilde{t}_{\mathcal{L}}(q=12)$	3.8	3.7	3.4	3.7	3.7
$\tilde{t}_{\mathcal{L}}(q=16)$	4.0	3.7	3.1	3.4	3.5	$\tilde{t}_{\mathcal{L}}(q=16)$	4.4	4.2	4.1	4.0	4.1
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	4.6	4.4	3.5	3.5	2.9	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	5.0	4.9	4.4	4.6	4.6
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	4.7	4.1	3.3	3.7	3.7	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	4.9	4.8	4.5	4.9	4.9
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	4.9	4.3	3.5	3.8	4.1	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	5.1	4.9	4.6	4.8	4.9
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	5.0	4.5	3.8	4.1	4.3	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	5.3	5.1	4.7	4.8	5.0
permutation	4.8	4.7	4.9	4.7	4.7	permutation	4.4	4.5	4.8	4.8	4.5
bootstrap	4.5	4.2	3.4	3.3	2.8	bootstrap	4.8	4.4	4.1	3.9	3.9

Notes: parameter values (a, c) as in Dufour et al. (2019). $N_I = N_Y = 200$, $\zeta_I = \zeta_Y = \zeta$.

Theil index: $(a_I, c_I) = (a_Y, c_Y) = (2.5, 2.502), (3.2, 1.232), (5.8, 0.4996)$.

Gini coefficient: $(a_I, c_I) = (a_Y, c_Y) = (2.5, 2.64), (3.2, 1.1866), (5.8, 0.447)$.

Table C.2: Empirical size – identical distributions, different sample sizes

Theil, $\zeta = 2.9 \backslash N_Y$	50	200	500	1000	5000	Gini, $\zeta = 2.59 \backslash N_Y$	50	200	500	1000	5000
asymptotic	8.7	5.3	4.7	4.6	4.5	asymptotic	12.3	6.6	5.7	5.4	5.2
$\tilde{t}_{\mathcal{L}}(q=4)$	1.2	1.2	1.6	1.5	1.6	$\tilde{t}_{\mathcal{L}}(q=4)$	1.5	1.5	1.8	1.8	1.7
$\tilde{t}_{\mathcal{L}}(q=8)$	2.2	2.0	2.4	2.4	2.5	$\tilde{t}_{\mathcal{L}}(q=8)$	3.1	2.7	3.2	3.1	3.4
$\tilde{t}_{\mathcal{L}}(q=12)$	2.7	2.4	2.9	2.8	3.0	$\tilde{t}_{\mathcal{L}}(q=12)$	3.6	3.4	3.5	3.6	4.0
$\tilde{t}_{\mathcal{L}}(q=16)$	3.5	2.8	3.2	3.3	3.4	$\tilde{t}_{\mathcal{L}}(q=16)$	4.3	3.7	3.6	3.9	4.2
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	2.9	3.1	3.7	3.7	3.7	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	4.3	4.4	4.8	4.7	4.8
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	3.6	3.2	3.9	3.7	4.0	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	4.5	4.4	4.8	4.7	4.6
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	3.8	3.7	3.8	3.7	4.4	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	4.9	4.6	4.7	4.9	5.1
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	4.5	3.5	3.8	3.9	4.0	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	5.2	4.6	4.7	4.7	4.8
permutation	5.0	4.8	4.9	4.9	5.0	permutation	5.1	4.9	5.0	4.9	5.2
bootstrap	4.0	4.2	4.1	4.5	4.7	bootstrap	4.6	4.8	4.8	5.1	5.3

Notes:$N_I = 200$.

Theil index: $(a_I, c_I) = (a_Y, c_Y) = (5.8, 0.4996)$.

Gini coefficient: $(a_I, c_I) = (a_Y, c_Y) = (5.8, 0.447)$.

Appendix D: One-sample problem.

For illustrative properties, this appendix discusses $t-$statistic inference approach and its implementation in the one-sample case. In the context of one-sample inference on income or wealth inequality measured using an inequality index $\mathcal{L}$ (e.g., Theil index or Gini coefficient, among others) the $t-$statistic based inference is conducted in the following simple way **(see also Ibragimov et al., 2013).**

Following Ibragimov and Müller (2010), a (large) sample, $I_1,\ I_2,...,\ I_N$, of observations on income (or wealth) I, is partitioned into a fixed number $q \geq 2$ (e.g., $q = 2, 4, 8$) groups, and the inequality index $\mathcal{L}$ is estimated for each group; resulting in q group level income inequality estimates $\widehat{\mathcal{L}}_j,\ j = 1, ..., q$. The robust test of the null hypothesis

Table C.3: Empirical size – identical distributions, different sample sizes

Theil$\backslash N_Y$	50	200	500	1000	5000	Gini $\backslash N_Y$	50	200	500	1000	5000
asymptotic	31.5	21.9	16.7	13.8	8.7	asymptotic	28.9	18.3	14.0	12.3	8.6
$\tilde{t}_{\mathcal{L}}(q=4)$	1.0	1.1	1.3	1.1	1.1	$\tilde{t}_{\mathcal{L}}(q=4)$	1.7	1.7	1.9	1.6	1.6
$\tilde{t}_{\mathcal{L}}(q=8)$	2.5	2.2	2.3	2.1	2.0	$\tilde{t}_{\mathcal{L}}(q=8)$	3.3	3.0	3.1	3.2	2.7
$\tilde{t}_{\mathcal{L}}(q=12)$	3.5	3.0	3.1	2.9	2.7	$\tilde{t}_{\mathcal{L}}(q=12)$	3.8	3.5	3.8	3.7	3.5
$\tilde{t}_{\mathcal{L}}(q=16)$	4.0	3.3	3.3	3.5	3.0	$\tilde{t}_{\mathcal{L}}(q=16)$	4.3	3.8	4.1	4.0	4.0
$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	3.0	3.0	3.1	2.8	2.7	$\tilde{\tilde{t}}_{\mathcal{L}}(q=4)$	4.7	4.5	4.5	4.3	3.9
$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	4.0	3.5	3.6	3.5	3.2	$\tilde{\tilde{t}}_{\mathcal{L}}(q=8)$	5.2	4.8	4.9	4.7	4.3
$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	4.7	4.1	3.9	3.8	3.7	$\tilde{\tilde{t}}_{\mathcal{L}}(q=12)$	5.1	5.0	4.9	4.9	4.8
$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	5.2	4.2	4.1	4.2	3.7	$\tilde{\tilde{t}}_{\mathcal{L}}(q=16)$	5.4	4.9	4.7	5.0	4.6
permutation	4.9	4.8	5.1	5.1	4.9	permutation	5.1	4.7	4.9	5.0	5.0
bootstrap	3.3	3.4	3.6	3.6	3.6	bootstrap	4.5	4.2	4.3	4.4	4.1

Notes: $N_I = 200$, $\zeta_I = \zeta_Y = 1.4$.

Theil and Gini coefficient: $(a_I, c_I) = (a_Y, c_Y) = (2, 0.7)$.

Table C.4: Empirical size – identical distributions, different sample sizes, different numbers of groups

Theil $\backslash N_Y$	400	600	800	Gini$\backslash N_Y$	400	600	800
asymptotic	8.8	9.3	11.4	asymptotic	8.8	8.1	9.1
$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=4)$	1.9	2.3	3.2	$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=4)$	2.2	2.1	2.7
$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=8)$	4.2	6.3	8.6	$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=8)$	3.9	4.3	5.1
$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=12)$	6.8	11.1	15.1	$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=12)$	5.1	5.7	6.8
$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=16)$	9.4	15.9	21.6	$\tilde{t}_{\mathcal{L}}(q_I=16, q_Y=16)$	5.8	6.6	8.7
$\tilde{t}_{\mathcal{L}}(q_I=3, q_Y=4)$	0.7	0.7	1.1	$\tilde{t}_{\mathcal{L}}(q_I=3, q_Y=4)$	0.8	1.0	1.3
$\tilde{t}_{\mathcal{L}}(q_I=6, q_Y=8)$	2.5	3.6	5.5	$\tilde{t}_{\mathcal{L}}(q_I=6, q_Y=8)$	2.8	3.1	3.8
$\tilde{t}_{\mathcal{L}}(q_I=9, q_Y=12)$	4.1	6.3	9.4	$\tilde{t}_{\mathcal{L}}(q_I=9, q_Y=12)$	3.9	4.1	5.0
$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=16)$	6.0	9.5	14.0	$\tilde{t}_{\mathcal{L}}(q_I=12, q_Y=16)$	4.8	5.1	6.4
$\tilde{t}_{\mathcal{L}}(q_I=2, q_Y=4)$	0.1	0.1	0.1	$\tilde{t}_{\mathcal{L}}(q_I=2, q_Y=4)$	0.0	0.1	0.1
$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=8)$	1.1	1.1	1.9	$\tilde{t}_{\mathcal{L}}(q_I=4, q_Y=8)$	1.4	1.5	2.0
$\tilde{t}_{\mathcal{L}}(q_I=6, q_Y=12)$	2.1	2.9	4.6	$\tilde{t}_{\mathcal{L}}(q_I=6, q_Y=12)$	2.6	2.8	3.4
$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=168)$	3.1	4.5	6.9	$\tilde{t}_{\mathcal{L}}(q_I=8, q_Y=16)$	3.4	3.6	4.5
permutation	5.4	4.9	4.9	permutation	5.4	4.7	4.8
bootstrap	4.6	3.7	4.0	bootstrap	5.3	4.4	4.4

Notes: $N_I = 200$, $N_Y = 400, 600, 800$.

$(a_I, c_I) = (a_Y, c_Y) = (2, 1.1)$; $\zeta_I = \zeta_Y = 2.2$.

$H_0 : \mathcal{L} = \mathcal{L}_0$ against the two-sided alternative $H_a : \mathcal{L} \neq \mathcal{L}_0$ is based on the usual $t-$statistic $t^I_{\mathcal{L}}$ in the q group level inequality estimates $\widehat{\mathcal{L}}_j$, $j = 1, ..., q$:

$$t_{\mathcal{L}} = \sqrt{q}\frac{\overline{\widehat{\mathcal{L}}} - \mathcal{L}_0}{s_{\widehat{\mathcal{L}}}} \tag{D.1}$$

with $\overline{\widehat{\mathcal{L}}} = \frac{\sum_{j=1}^q \widehat{\mathcal{L}}_j}{q}$ and $s^2_{\widehat{\mathcal{L}}} = \frac{\sum_{j=1}^q \left(\widehat{\mathcal{L}}_j - \overline{\widehat{\mathcal{L}}}\right)^2}{q-1}$. The null hypothesis is rejected in favor of the alternative at level $\alpha \leq 0.083$

(which includes the conventional significance level $\alpha = 0.05$) if the absolute value $|t_{\mathcal{L}}|$ of the $t-$statistic in group estimates $\widehat{\mathcal{L}}_j$ exceeds the $(1-\alpha/2)-$quantile of the standard Student-t distribution with $q-1$ degrees of freedom: $|t_{\mathcal{L}}| > cv_{q,\alpha}$.

The test of H_0 against H_a at level $\alpha \leq 0.1$ can be conducted in the same way if $2 \leq q \leq 14$. Using the results in Bakirov and Székely (2006) and Ibragimov and Müller (2010), the $p-$values of the above $t-$statistic based tests can be calculated in the case of an arbitrary number of groups q. This enables robust tests on inequality index $\mathcal{L}$ at any chosen level. One-sided tests are conducted in a similar way; note that quantiles of Student-t distributions with $q-1$ degrees of freedom can also be used in one-sided tests of level $\alpha \leq 0.1$ if $q \in \{2, 3\}$.

By implication, for all $\alpha \leq 0.083$ (and all $\alpha \leq 0.1$ for $2 \leq q \leq 14$) a confidence interval for the inequality index $\mathcal{L}$ with asymptotic coverage of at least $1-\alpha$ may be constructed as $\widehat{\mathcal{L}}_j \pm cv_{q,\alpha} s_{\widehat{\mathcal{L}}}$. For example, the 95% confidence interval for $\mathcal{L}$ is given by $(\overline{\widehat{\mathcal{L}}} - cv_{q,0.05} s_{\widehat{\mathcal{L}}},\ \overline{\widehat{\mathcal{L}}} + cv_{q,0.05} s_{\widehat{\mathcal{L}}})$, where $cv_{q,0.05}$ is the 0.975-quantile of the Student-t distribution with $q-1$ degrees of freedom: $P(|T_{q-1}| > cv_{q,0.05}) = 0.05$.[31]

From the results in Ibragimov and Müller (2010), the $t-$statistic inference approach results in asymptotically valid inference under the assumption that the group level income inequality estimators $\widehat{\mathcal{L}}_j$, $j = 1, ..., q$, are asymptotically independent, unbiased and Gaussian, even if of different variances.

As discussed in Section 2.1, asymptotic validity of t-statistic based inference also holds under convergence of the group estimators to conditionally normal r.v.s. The limiting r.v.'s may further be unconditionally dependent through their second moments or have a common shock-type dependence. This implies that the approach can be applied to inference on $\mathcal{L}$ in the presence of extremes and outliers in observations generated by heavy-tailedness with infinite variances, as well as dependence structures that include models with multiplicative common shocks. Importantly, The inference approach does not require the estimation of limiting variances of estimators of interest, in contrast to inference methods based on consistent (e.g., HAC or clustered) standard errors.

The conditions for asymptotic validity of the $t-$statistic based approach to robust inference are typically satisfied in applications under the choice of groups that imply asymptotic unbiasedness and independence of group estimators of the inequality indices (see below). The asymptotic Gaussianity (and other weak convergence results such as convergence to heavy-tailed stable distributions under infinite second moments) of group inequality estimators ($\widehat{\mathcal{L}}_j$) follows from the same reasoning and holds under the same conditions as the asymptotic Gaussianity (and other relevant asymptotics) of the full-sample inequality estimator ($\widehat{\mathcal{L}}$). As discussed in Appendix B, asymptotic normality holds for estimators of Theil and Gini indices in the case of power law income distributions (1) with tail indices $\zeta > 2$ and finite second moments.

The condition that group estimators of inequality be asymptotically unbiased (and independent) places a natural restriction on the formation of groups (see also discussion of general $t-$statistic inference approach in Ibragimov and Müller, 2010). The grouping must be such that each group estimator is an unbiased estimate of inequality in the population, such that the mean of the group estimators of inequality will asymptotically be equal to inequality in the population. This requires the grouping to be such that there is no "between-group" inequality in expectation. Thus, for example, for one-sample inference on inequality in a country, groups cannot be the country's regions. This is because in that case, the group estimators would measure within-region inequality, and their mean will miss out the "between-region" component of inequality in the country as a whole.

The groups may be formed by simply partitioning the random sample. Given the random sample $I_1, I_2, ..., I_N$ of (i.i.d.) incomes that are not pre-grouped or ordered by regions or other markers, the q groups may be formed as $\{I_k, (i-1)N/q < k \leq iN/q\}$, $i = 1, ..., q,$. With this simple scheme of grouping, asymptotic unbiasedness and independence of group inequality estimators will hold due to i.i.d.ness of data – there is no "between-groups" inequality in expectation.

Table D.1 provides results on the empirical sizes of the asymptotic and the $t-$statistic based robust tests on the Theil and Gini indices for sample sizes $N = 200, 500, 1000$ drawn from the S-M distributions $SM(a, b_0, c)$ with, in the case of the Theil index, the parameters $(a, c) = (2.5, 2.502)$, $(3.2, 1.232)$, $(5.8, 0.4996)$, correspond to the tail

[31] Thus the width of the confidence interval (and their two-sample analogues) depends on sample standard deviations of group estimators. This is in contrast to confidence intervals constructed using consistent estimators of limiting variances of sample inequality measures whose width depends on the consistent standard errors.

indices $\zeta = 6.6, 3.94, 2.9$. In the case of the Gini coefficient, parameters $(a, c) = (2.5, 2.640), (3.2, 1.187)$, and $(5.8, 0.447)$ correspond to $\zeta = 6.26, 3.8, 2.9$.

Consistent with the finite-sample distributions of statistics Z and S such as those in Figures 1-3, the results show that the finite-sample sizes of both the asymptotic and $t-$statistic based tests become more distorted when the tail index is smaller, i.e., when the degree of heavy-tailedness is more pronounced. Importantly, size distortions for the Gini coefficient are not as large as for the Theil index which is more sensitive to the upper tail. One should note that the empirical size of the $t-$statistic based approach is sensitive to the choice of the number of groups q. At the same time, when the number of groups q is 4 or 8, the finite-sample size properties of robust tests based on the $t-$statistic in group estimates are generally better than those of the tests based on asymptotic normality of the (full-sample) $t-$statistics for the measures; and when each of the groups contains more than 100 observations (with $q = 4$), better than those of the asymptotic tests.[32]

Table D.1: Empirical size: Identical distributions with $\zeta_I = \zeta_Y = \zeta$ and sample sizes $N_I = N_Y = N$

Theil	$\zeta = 6.26$	$\zeta = 3.94$	$\zeta = 2.9$	Gini	$\zeta = 6.6$	$\zeta = 3.8$	$\zeta = 2.59$
$N = 200$							
asymptotic	8.2	14.5	25.5	asymptotic	6.2	7.5	13.0
q =4	6.9	10.6	18.0	$q = 4$	5.2	5.2	7.7
q =8	11.0	17.8	28.7	$q = 8$	5.2	6.0	11.3
q =12	15.9	24.9	37.3	$q = 12$	5.5	6.6	14.2
q =16	21.3	33.1	45.9	$q = 16$	5.5	6.9	16.9
$N = 500$							
asymptotic	6.9	11.9	20.2	asymptotic	5.7	6.5	10.8
q =4	5.8	8.2	13.5	$q = 4$	4.8	5.1	7.1
q =8	8.3	12.9	20.6	$q = 8$	5.3	5.9	9.5
q =12	10.0	16.2	25.6	$q = 12$	5.1	6.3	11.5
q =16	12.7	20.0	30.0	$q = 16$	5.1	6.4	13.0
$N = 1000$							
asymptotic	6.0	9.6	17.0	asymptotic	5.2	5.7	8.6
q =4	5.3	6.5	10.5	$q = 4$	4.8	4.9	5.8
q =8	6.1	9.3	16.3	$q = 8$	4.9	5.1	7.4
q =12	7.4	11.8	19.5	$q = 12$	4.9	5.2	8.5
q =16	8.7	14.2	22.6	$q = 16$	5.0	5.4	10.1

[32] These conclusions on the number accord with the numerical results presented in Ibragimov and Müller (2010) that indicate that, for many dependence and heterogeneity settings considered in the literature and typically encountered in practice for time series, panel, clustered and spatially correlated data, the choice of the number of groups $q = 4, 8$ or $q = 16$ leads to robust tests with attractive finite sample performance. One should emphasize that the asymptotic efficiency results for $t-$statistic based robust inference in Ibragimov and Müller (2010) imply that it is not possible to use data-dependent methods to determine the optimal number of groups q to be used in the approach when the only assumption imposed on the data generating process is that of asymptotic normality and asymptotic independence of group estimators of the parameter of interest.