CIREQ

Centre interuniversitaire de
recherche en économie quantitative

Cahier 03-2020

Identification-Robust Inequality Analysis

Jean-Marie Dufour, Emmanuel Flachaire, Lynda Khalaf and Abdallah Zalghout

CIREQ

Cahier 03-2020

Identification-Robust Inequality Analysis

Jean-Marie Dufour, Emmanuel Flachaire,
Lynda Khalaf and Abdallah Zalghout

CIREQ, Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal (Québec) H3C 3J7
Canada
Téléphone : (514) 343-6557
Télécopieur : (514) 343-7221
cireq@umontreal.ca
http://www.cireqmontreal.com

Dépôt légal - Bibliothèque nationale du Canada, 2020, ISSN 0821-4441
Dépôt légal - Bibliothèque et Archives nationales du Québec, 2020
ISBN-13 : 978-2-89382-757-5

Identification-robust inequality analysis [a]

Jean-Marie Dufour [b]
McGill University

Emmanuel Flachaire [c]
Aix-Marseille Université

Lynda Khalaf [d]
Carleton University

Abdallah Zalghout [e]
Carleton University

April 2020

[a] This work was supported by the William Dow Chair of Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and by project ANR-16-CE41-0005 managed by the French National Research Agency (ANR).

[b] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill Uniersity, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; email: jean-marie.dufour@mcgill.ca. Web page: http://www.jeanmariedufour.com

[c] Aix-Marseille Université, CNRS, EHESS, Centrale Marseille, AMSE. Mailing address: GREQAM-EHESS, Aix-Marseille Université, 2 rue de la charité, 13002, Marseille, France. Email: emmanuel.flachaire@univ-amu.fr.

[d] Corresponding author, Economics Department, Carleton University, K1S 5B6, ON, Canada. Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles (GREEN), Université Laval. Email: Lynda_Khalaf@carleton.ca.

[e] Economics Department, Carleton University. abdallahzalghout@cmail.carleton.ca.

ABSTRACT

We propose confidence sets for inequality indices and their differences, which are robust to the fact that such measures involve possibly weakly identified parameter ratios. We also document the fragility of decisions that rely on traditional interpretations of - significant or insignificant - comparisons when the tested differences can be weakly identified. Proposed methods are applied to study economic convergence across U.S. states and non-OECD countries. With reference to the growth literature which typically uses the variance of log per-capita income to measure dispersion, results confirm the importance of accounting for micro-founded axioms and shed new light on enduring controversies surrounding convergence.

Contents

List of Figures

List of Tables

1 Introduction

Economic inequality can be broadly defined in terms of the distribution of economic variables, which include income (predominantly), and other variables such as consumption or health. Using one or more samples, inequality can be measured in several ways, most of which are justified statistically as well as through theoretical axiomatic approaches. In this context, size-correct statistical inference is an enduring challenge. One reason is that the underlying distributions often have thick tails, which contaminate standard asymptotic and bootstrap-based procedures (Davidson and Flachaire, 2007; Cowell and Flachaire, 2007). Another reason is that two different distributions can yield equal measures, which complicates comparisons (Dufour et al., 2019).

An important additional difficulty is that common inequality measures – such as the generalized entropy (**GE**) and Gini indices – involve transformations of parameters (Cowell and Flachaire, 2015). Formally, denote by X the random variable with a typical realization representing say the income of a randomly chosen individual in the population, and let F_X refer to the CDF of X. Given a predetermined parameter – denoted γ – that characterizes the sensitivity to changes over different parts of the income distribution, the GE measure – denoted GE_γ – can be defined as a function of the ratio of two particular moments of F_X: the mean $\mu_X = \mathbb{E}_F(X)$ and $\nu_X(\gamma) = \mathbb{E}_F(X^\gamma)$.[1] Such nonlinear forms may easily be ill-conditioned or poorly identified, with non-trivial implications on the associated estimators and test statistics; see Dufour (1997). The *first* objective of this paper is to underscore and address resulting inference problems.

Identification broadly refers to our ability to recover objects of interest from available models and data (Dufour and Hsiao, 2008). In the context of income inequality, it was long believed that statistical measures of *precision* are not required, as researchers deal with large samples. The large standard errors reported in empirical studies suggest otherwise, stressing the importance of conducting inference valid for all sample sizes (Maasoumi, 1997). Yet standard errors, large or small, do not tell the whole story. In fact, the profession now recognizes that confidence intervals with bounded limits, which automatically result from inverting conventional t-type tests (based on standard errors), deliver false statistical decisions and undercut the reliability of related policies. Despite a sizable econometric literature on inequality, methods that take into account the irregularities underscored in the weak identification literature appear to be missing

[1] This definiton implies that GE_γ is more sensitive to differences in the top (bottom) tail with more positive (negative) γ.

in this context.[2]

More to the point from the index comparison perspective, most available approaches for this purpose focus on *significance* tests. The *second* objective of this paper is to document the fragility of decisions relying on traditional interpretations of – significant or insignificant – test results, when the difference under test can be weakly identified. In particular, when a zero difference cannot be rejected, we show that because of the definition of conventional inequality indices, one may also not be able to refute a large spectrum of possible values of this difference. From a policy perspective, this indicates that available samples are uninformative on inequality changes, which stands in sharp contrast to a no-change conclusion.

The *third* objective is to propose tractable identification-robust confidence sets for inequality indices – in particular, for differences between such indices – which require the same basic inputs as their standard counterparts. Whereas usual companion variances and covariances as well as critical values need to be computed, the alternative test statistics are formed and inverted analytically into confidence sets that will reflect the underlying identification status.

The *fourth* objective is to discuss challenges for empirical researchers and policy-makers in light of the above observations. We study evidence on economic convergence; see *e.g.* Romer (1994) for a historical critical perspective. We show that conflict in test decisions and uninformative confidence sets cannot be ruled out with standard measures and data sets. The fact that tests and confidence sets have different theoretical implications is not alarming. However, when these differences are empirically relevant, this can lead to severe economic and policy controversies. To the best of our knowledge, this problem and our proposed solution have escaped formal notice.

Indeed, the literature on statistical inference for inequality measures is relatively recent; see Cowell and Flachaire (2015) for a comprehensive survey. In particular, the standard bootstrap is known to fail, and alternative methods remain scarce for both the one-sample problem of analyzing a single index (Davidson and Flachaire, 2007; Dufour et al., 2018) and the two-sample problem of assessing differences between two indices (Dufour et al., 2019). It is important to note that the latter problem is much more challenging than the former. For testing the equality of two inequality measures from independent samples, Dufour et al. (2019) suggest a permutational approach for the two-sample problem which outperforms other asymptotic and bootstrap methods available in the literature. However, these results are limited to testing the equality

[2]See *e.g.* Dufour (1997), Andrews and Cheng (2013), Kleibergen (2005), Andrews and Mikusheva (2015), Beaulieu et al. (2013), Bertanha and Moreira (2016), and references therein; see also Bahadur and Savage (1956) and Gleser and Hwang (1987).

of two inequality measures and do not provide a way of making inference on a *possibly non-zero* difference nor building a confidence interval on the difference.

In the present paper, we propose Fieller-type methods for inference on the GE family of inequality indices. This family satisfies a set of key axiomatic principles and is widely used in practice.[3] We study the general comparison problem of testing any *possibly non-zero* difference between measures, with either independent or dependent samples. Moving from testing a zero difference to assessing the size of the difference is much more informative from both statistical and economic viewpoints, including potential policy recommendations.

The fact that inequality measures in general, and those considered in this paper in particular, can be expressed as ratios of moments or ratios of functions of moments, provides a strong motivation for our work since Fieller-type methods are typically used for inference on ratios. Fieller's original solution for the means of two independent normal random variables was extended to multivariate normals (Bennett, 1959), general exponential (Cox, 1967) and linear (Zerbe, 1978; Dufour, 1997) regression models, dynamic models with possibly persistent covariates (Bernard et al., 2007, 2019) and for simultaneous inference on multiple ratios (Bolduc et al., 2010). For a good review of inference on ratios, see Franz (2007).

On the GE class of inequality indices, this paper makes the following contributions. *First*, we provide analytical and tractable solutions for proposed confidence sets. *Second*, we show in a simulation study that the proposed solutions are more reliable than Delta counterparts. *Third*, we show that our approach outperforms most simulation-based alternatives including the permutation test of Dufour et al. (2019). *Fourth*, our solution covers tests for any given value of the difference [i.e. not just zero, in contrast with Dufour et al. (2019)], allowing the construction of confidence sets through test inversion. *Fifth*, we provide useful empirical evidence supporting the seemingly counter-intuitive bounds that Fieller-type methods can produce.

Key simulations results illustrate the superiority of Fieller-type methods across the board: (1) the improved level control (over the Delta method) is especially notable for indices that put more weight on the right tail of the distribution *i.e.* as γ increases; (2) size improvements preserve power; (3) results are robust to different assumptions on the shape of the null distributions; (4) tests based on the Fieller-type method outperform available permutation tests when the distributions under the null hypothesis are different. A permutational approach is not available (to date) for the general problem we consider here. Overall, while

[3] These include scale invariance, the Pigou-Dalton transfer, the symmetry and the Dalton population principle. It is also additively decomposable. See Cowell (2000) for a detailed discussion on these and other properties of indices.

irregularities arising from the right tail have long been documented, we find that left-tail irregularities are equally important in explaining the failure of standard inference methods for inequality measures.

Our empirical study on growth demonstrates the practical relevance of these theoretical results. Using per-capita income data for 48 U.S. states, we analyze the convergence hypothesis by comparing the inequality levels between 1946 and 2016. In contrast to the bulk of this literature, we depart from just testing and build confidence sets to document the economic and policy significance of statistical decisions. The empirical literature on growth relies on the variance of log incomes as a measure of dispersion in per-capita income distributions (Blundell et al., 2008). But this measure violates the Pigou-Dalton principle (Araar and Duclos, 2006). We use GE indices instead, since these satisfy the axioms suggested in the inequality measurement literature. We document specific cases where the variance of log incomes decrease while the GE_2 measure indicates the opposite. Empirically, accounting for micro-founded axioms is of first-order importance.

We find that inter-state inequality has declined over the 1946-2016 period indicating convergence across the states. For the GE_2 index, the Fieller-type and Delta methods lead to contradictory conclusions: in contrast to the former, the latter suggests that inequality declines are insignificant at usual levels. Results with non-OECD countries stress the severe consequences of ignoring identification problems: with the GE_2 index, the Fieller-type method produces an unbounded set, which casts serious doubts on the reliability of the no-change results using the Delta method.

The rest of the paper is organized as follows. Section 2 derives Fieller-type confidence sets. Section 3 reports the results of the simulation study. Section 4 contains the inter-state convergence application, and Section 5 concludes. Figures and tables are presented in the Appendix.

2 Fieller-type confidence sets for Generalized Entropy inequality measures

An inequality measure is a measure of dispersion in a distribution of a random variable. We shall find it convenient throughout the rest of this paper to refer to income distributions, though our results apply equally to other popular distributions considered in the area of inequality such as wage, wealth, and consumption distributions. Many inequality measures, including the GE_γ class, depend solely on the underlying distribution and can typically be written as a functional which maps the space of the cumulative distribution function (**CDF**) to the nonnegative real line $\mathbb{R}^0_+$.

Our aim is to make inference on the GE_γ measure for any given $\gamma \in (0,2)$. In particular, we wish to build an asymptotic Fieller-type confidence set (**FCS**) for the difference between two measures. We call this problem the *two-sample problem*, as opposed to the *one-sample problem* where the objective consists in testing and building a confidence interval for a single index. The crucial difference between a FCS and its standard counterpart based on the Delta method (**DCS**) is that the former reformulates the null hypothesis in a linear form. The method proceeds by inverting the square of the t-test associated with the reformulated linear hypothesis. Consequently, it avoids the irregularities which affect the validity of the Delta method as the denominator approaches zero.

A consequence of rewriting the null hypothesis in linear form is that the variance used by the Fieller-type statistic depends on the true value of the tested parameter, which leads to a quadratic inequality problem. The resulting confidence regions are not standard, in the sense that they may be asymmetric, consisting of two disjoint unbounded confidence intervals or the whole real line $\mathbb{R}$. Nevertheless, unbounded intervals are an attractive feature of the method which addresses coverage problems (Koschat et al., 1987; Gleser and Hwang, 1987; Dufour, 1997; Dufour and Jasiak, 2001; Dufour and Taamouti, 2005, 2007; Bertanha and Moreira, 2016). For a geometric comparison of the Fieller and Delta methods, see Hirschberg and Lye (2010).

Our notational framework is introduced below for presentation clarity. To set focus, we first extend the work of Dufour et al. (2018) on the one-sample problem for the Theil index in order to cover the GE_γ class. Our main contribution on the two-sample problem is next presented under three different dependence schemes.

2.1 Framework and notation

The $GE_\gamma(X)$ measure can be expressed as in Shorrocks (1980):

$$\begin{aligned} GE_\gamma(X) &= \frac{1}{\gamma(\gamma-1)}\left[\frac{\mathbb{E}_F(X^\gamma)}{[\mathbb{E}_F(X)]^\gamma}-1\right] \quad \text{for} \quad \gamma \neq 0, 1, \\ GE_0(X) &= \mathbb{E}_F[\log(X)] - \log[\mathbb{E}_F(X)], \\ GE_1(X) &= \frac{\mathbb{E}_F[X\log(X)]}{\mathbb{E}_F(X)} - \log[\mathbb{E}_F(X)]. \end{aligned} \tag{2.1}$$

The family nests several indices including two well-known ones introduced by Theil (1967): the Mean Logarithmic Deviation (*MLD*) which is the limiting value of the $GE_\gamma(X)$ as γ approaches zero, and the Theil index which is the limiting value of the $GE_\gamma(X)$ as γ approaches 1. When $\gamma = 2$, the index is equal to half of the coefficient of variation and is related to the Hirschman-Herfindahl (HH) index which is widely used in industrial organization (Schluter, 2012). The Atkinson index can be obtained from the $GE_\gamma(X)$ index using an appropriate transformation.

If $X_1,\ldots,X_n$ is a sample of i.i.d. observations, the empirical distribution function (**EDF**), denoted by $\hat{F}_X$, can be estimated by

$$\hat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(X_i \le x) \tag{2.2}$$

where n is the number of observations and $\mathbf{1}(\cdot)$ is the indicator function that takes the value of 1 if the argument is true, and 0 otherwise. We can consistently estimate $GE_\gamma(X)$ by

$$\widehat{GE}_\gamma(X) = \frac{1}{\gamma(\gamma-1)}\left[\frac{\hat{\nu}_X(\gamma)}{\hat{\mu}_X^\gamma} - 1\right], \tag{2.3}$$

$$\hat{\mu}_X = \int x\,d\hat{F}_X = \frac{1}{n}\sum_{i=1}^{n}X_i, \quad \hat{\nu}_X(\gamma) = \int x^\gamma d\hat{F}_X = \frac{1}{n}\sum_{i=1}^{n}X_i^\gamma. \tag{2.4}$$

For our two-sample analysis, we denote by X the random variable representing the incomes of individuals from the first population with CDF F_X, and by Y the incomes of individuals from the second population with CDF F_Y. We assume we have two i.i.d. samples $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_m$ from each population. EDF's denoted $\hat{F}_X(x)$ and $\hat{F}_Y(y)$ are obtained as in (2.2) and the inequality measures denoted $GE_\gamma(X)$ and $GE_\gamma(Y)$ and defined as in (2.1) are estimated conformably as in (2.3) with estimated counterparts denoted by $\widehat{GE}_\gamma(X)$ and $\widehat{GE}_\gamma(Y)$. Our analysis covers:

Assumption 1. *Samples are of equal sizes and independent.*

Assumption 2. *Samples have different sample size and are independent.*

Assumption 3. *Samples are of equal sizes and dependent.*

2.2 The one-sample problem

In this section we propose level $1-\alpha$ confidence sets for $GE_\gamma(X)$ using a single sample. The standard DCS is derived by inverting the square (or the absolute value) of the t-test associated with

$$H_D(\delta_0) : GE_\gamma(X) = \delta_0 \tag{2.5}$$

where δ_0 is a some admissible value of the index. By inverting a test statistic with respect to the parameter tested (δ_0 in this case), we mean collecting the values of the parameter for which the test cannot be rejected at a given significance level α. Assuming that the estimator is asymptotically normal, this can be carried out by solving the following inequality for δ_0:

$$\hat{T}_D(\delta_0)^2 = \left[\frac{\widehat{GE}_\gamma(X) - \delta_0}{\left(\hat{V}_D[\widehat{GE}_\gamma(X)]\right)^{1/2}}\right]^2 \leq z^2_{\alpha/2} \tag{2.6}$$

where $z_{\alpha/2}$ is the asymptotic two-tailed critical value at the significance level α (*i.e.*, $\mathbb{P}[Z \geq z_{\alpha/2}] = \alpha/2$ for $Z \sim N[0, 1]$) and $\hat{V}_D[\widehat{GE}_\gamma(X)]$ is the estimate of the asymptotic variance.

The Delta method implies:

$$V_D[GE_\gamma(X)] = \frac{1}{n}\begin{bmatrix} \frac{\partial GE_\gamma(x)}{\partial \mu_x} & \frac{\partial GE_\gamma(x)}{\partial \nu_x} \end{bmatrix} \begin{bmatrix} \sigma^2_x & \sigma_{x,x^\gamma} \\ \sigma_{x^\gamma,x} & \sigma^2_{x^\gamma} \end{bmatrix} \begin{bmatrix} \frac{\partial GE_\gamma(x)}{\partial \mu_x} & \frac{\partial GE_\gamma(x)}{\partial \nu_x} \end{bmatrix}' \tag{2.7}$$

where $\sigma^2_X, \sigma^2_{X^\gamma}$ and σ_{X,X^γ} represent the variance of X, the variance of X^γ and the covariance between X and X^γ respectively. The estimated variance of $\hat{\mu}_X$, the variance of $\hat{\nu}_X$ and the covariance between $\hat{\mu}_X$ and $\hat{\nu}_X$ are equal to $\hat{\sigma}^2_X/n$, $\hat{\sigma}^2_{X^\gamma}/n$ and $\hat{\sigma}_{X,X^\gamma}/n$. In our estimation, we use the sample counterparts for these population moments estimated using the EDF of the two samples. Solving (2.6) for δ_0 and plugging in the estimate $\hat{V}_D := \hat{V}_D[\widehat{GE}_\gamma(X)]$ of the variance in (2.7) we get:

$$\text{DCS}[GE_\gamma(X); 1-\alpha] = \left[\widehat{GE}_\gamma(X) - z_{\alpha/2}\hat{V}_D^{1/2},\ \widehat{GE}_\gamma(X) + z_{\alpha/2}\hat{V}_D^{1/2}\right]. \tag{2.8}$$

In contrast, we propose a Fieller-type set by inverting the square of the t-test associated with the linearized counterpart of $H_D(\delta_0)$, namely:

$$H_F(\delta_0) : \theta(\delta_0) = 0, \quad \text{where} \quad \theta(\delta_0) = \nu_X(\gamma) - \mu_X^\gamma - \gamma(1-\gamma)\mu_X^\gamma \delta_0\,. \tag{2.9}$$

Maintaining the asymptotic normality assumption as with the DCS, the proposed set can be formally characterized as follows.

Theorem 1. *Given a single sample of observations on a random variable X, consider the $GE_\gamma(X)$ index defined by* (2.1) *and its estimate given by* (2.3)*. The* $(1-\alpha)$*-level Fieller confidence set for this index is* *obtained as follows*

$$\mathrm{FCS}[GE_\gamma(X);1-\alpha]=\begin{cases}\left[\frac{-B-\sqrt{D}}{2A},\frac{-B+\sqrt{D}}{2A}\right] & \textit{if}\quad D>0 \textit{ and } A>0\\ \left]-\infty,\frac{-B+\sqrt{D}}{2A}\right]\cup\left[\frac{-B-\sqrt{D}}{2A},+\infty\right[& \textit{if}\quad D>0 \textit{ and } A<0\\ \mathbb{R} & \textit{if}\quad D<0,\ A<0\,,\end{cases} \tag{2.10}$$

$$D=B^2-4AC,\quad A=A_1-z_{\alpha/2}^2A_2\,,\quad B=B_1-z_{\alpha/2}^2B_2\,,\quad C=C_1-z_{\alpha/2}^2C_2\,, \tag{2.11}$$

$$A_1=\hat{\mu}_X^{2\gamma}[\gamma^2-\gamma]^2,\quad B_1=-2\hat{\mu}_X^{\gamma}[\gamma^2-\gamma][\hat{v}_X(\gamma)-\hat{\mu}_X^{\gamma}],\quad C_1=\left(\hat{v}_X(\gamma)-\hat{\mu}_X^{\gamma}\right)^2, \tag{2.12}$$

$$A_2=[\hat{\sigma}_X^2\gamma^2[\gamma^2-\gamma]^2\hat{\mu}_X^{2(\gamma-1)}]/n\,,\quad B_2=[2\hat{\sigma}_X^2\gamma^2[\gamma^2-\gamma]\hat{\mu}_X^{2(\gamma-1)}-2\gamma[\gamma^2-\gamma]\hat{\sigma}_{X,X^\gamma}\hat{\mu}_X^{\gamma-1})]/n\,, \tag{2.13}$$

$$C_2=[\hat{\sigma}_X^2\gamma^2\hat{\mu}_X^{2(\gamma-1)}-2\gamma\hat{\sigma}_{X,X^\gamma}\hat{\mu}_X^{\gamma-1}+\hat{\sigma}_{X^\gamma}^2]/n\,. \tag{2.14}$$

Proof. The standard t-statistic associated with (2.9) takes the form

$$\hat{T}_F(\delta_0)=\frac{\hat{\theta}(\delta_0)}{\left(\hat{V}[\hat{\theta}(\delta_0)]\right)^{1/2}}, \tag{2.15}$$

where $\hat{\theta}(\delta_0)$ is an estimate of $\theta(\delta_0)$ and $\hat{V}[\hat{\theta}(\delta_0)]$ is an estimate of the variance of $\hat{\theta}(\delta_0)$. Assuming asymptotic normality, inverting a test based on this statistic leads to the quadratic inequality

$$\hat{T}_F(\delta_0)^2\le z_{\alpha/2}^2\quad\Leftrightarrow\quad\frac{\hat{\theta}(\delta_0)^2}{\hat{V}[\hat{\theta}(\delta_0)]}\le z_{\alpha/2}^2\quad\Leftrightarrow\quad\hat{\theta}(\delta_0)^2-z_{\alpha/2}^2\hat{V}[\hat{\theta}(\delta_0)]\le 0\,. \tag{2.16}$$

A few algebraic manipulations give

$$\hat{\theta}(\delta_0)^2=A_1\delta_0^2+B_1\delta_0+C_1,\quad\hat{V}[\hat{\theta}(\delta_0)]=A_2\delta_0^2+B_2\delta_0+C_2. \tag{2.17}$$

On substituting (2.17) into (2.16), we obtain the following

$$\mathrm{FCS}[GE_\gamma(X);1-\alpha]=\left\{\delta_0: A\delta_0^2+B\delta_0+C\le 0\right\}, \tag{2.18}$$

the solution of which corresponds to (2.10). □

For more details, see Bolduc et al. (2010) and the references therein. Unlike the Delta method, the Fieller-type method satisfies the theoretical result which states that, for a confidence interval of a locally almost unidentified (**LAU**) parameter, or a parametric function, to attain correct coverage, it should allow for a non-zero probability of being unbounded (Koschat et al., 1987; Gleser and Hwang, 1987; Dufour, 1997; Dufour and Taamouti, 2005, 2007; Bertanha and Moreira, 2016).

2.3 The two-sample problem

The above rationale is extended here to $1-\alpha$ confidence sets for $\Delta GE_\gamma := GE_\gamma(X) - GE_\gamma(Y)$. Inverted tests thus focus on null hypotheses of the form

$$H_D(\Delta_0) : \Delta GE_\gamma = \Delta_0 \tag{2.19}$$

where Δ_0 is any known admissible value of ΔGE_γ, including possibly $\Delta_0 = 0$, for equality. We derive the DCS and FCS for each of Assumptions 1 - 3. These cases will actually differ only by the expression of the variance. Thus to avoid redundancy, we will derive the method in its most general form and state the restrictions required to obtain the relevant formulae in to each case.

The square of the asymptotic t-type statistic for $H_D(\Delta_0)$ is

$$\hat{W}_D(\Delta_0)^2 = \frac{[\Delta\widehat{GE}_\gamma - \Delta_0]^2}{\hat{V}_D[\Delta\widehat{GE}_\gamma]} \tag{2.20}$$

where $\Delta\widehat{GE}_\gamma = \widehat{GE}_\gamma(X) - \widehat{GE}_\gamma(Y)$ which upon inversion yields the confidence set:

$$\mathrm{DCS}(\Delta GE_\gamma; 1-\alpha) = \left[\widehat{GE}_\gamma(X) - z_{\alpha/2}\,[\hat{V}_D(\Delta\widehat{GE}_\gamma)]^{1/2},\; \widehat{GE}_\gamma(X) + z_{\alpha/2}\,[\hat{V}_D(\Delta\widehat{GE}_\gamma)]^{1/2}\right]. \tag{2.21}$$

The estimation of the variance $V_D(\Delta\widehat{GE}_\gamma)$ in (2.21) will differ according to the three cases stated above. The general form of the variance which encompasses the variances relevant for each of these cases can be written as:

$$V(\Delta\widehat{GE}_\gamma) = \begin{bmatrix} \frac{\partial\Delta GE_\gamma}{\partial\mu_x} \\ \frac{\partial\Delta GE_\gamma)}{\partial\nu_x} \\ \frac{\partial\Delta GE_\gamma)}{\partial\mu_y} \\ \frac{\partial\Delta GE_\gamma)}{\partial\nu_y} \end{bmatrix}' \begin{bmatrix} \Sigma_{XX}/n & \Sigma_{XY}/n \\ \Sigma_{YX}/n & \Sigma_{YY}/n \end{bmatrix} \begin{bmatrix} \frac{\partial\Delta GE_\gamma}{\partial\mu_x} \\ \frac{\partial\Delta GE_\gamma)}{\partial\nu_x} \\ \frac{\partial\Delta GE_\gamma}{\partial\mu_y} \\ \frac{\partial\Delta GE_\gamma}{\partial\nu_y} \end{bmatrix} = R \begin{bmatrix} \Sigma_{xx}/n & \Sigma_{xy}/n \\ \Sigma_{yx}/n & \Sigma_{yy}/n \end{bmatrix} R', \tag{2.22}$$

$$\Sigma_{XX} = \begin{bmatrix} \sigma_x^2 & \sigma_{x,x^\gamma} \\ \sigma_{x^\gamma,x} & \sigma_{x^\gamma}^2 \end{bmatrix},\; \Sigma_{YY} = \begin{bmatrix} \sigma_y^2 & \sigma_{y,y^\gamma} \\ \sigma_{y^\gamma,y} & \sigma_{y^\gamma}^2 \end{bmatrix},\; \Sigma_{XY} = \begin{bmatrix} \sigma_{x,y} & \sigma_{x,y^\gamma} \\ \sigma_{x^\gamma,y} & \sigma_{x^\gamma,y^\gamma} \end{bmatrix},\; \Sigma_{YX} = \Sigma_{XY}'. \tag{2.23}$$

The variance under Assumption 1 can be determined simply by setting Σ_{XY} in (2.22) equal to zero since the samples are assumed to be independent. Under Assumption 2, the variance is determined by setting Σ_{XY} in (2.22) equal to zero and by dividing Σ_{YY} by m instead of n Assumption 3 requires taking into account the dependence between the two samples. Since the samples are of equal size, the covariance matrix Σ_{XY} is not equal to zero. The variance actually corresponds to (2.22) without imposing earlier restrictions [*i.e.*, $V_3(\Delta\widehat{GE}_\gamma) = V(\Delta\widehat{GE}_\gamma)$].

Our proposed Fieller-type method proceeds by reformulating the null hypothesis in a linear form (without the ratio transformation). This can be obtained through multiplication of both sides of (2.19) by the common denominator $\gamma(\gamma-1)\mu_X^\gamma\mu_Y^\gamma$:

$$H_F(\Delta_0) : \Theta(\Delta_0) = 0 \qquad \text{where} \qquad \Theta(\Delta_0) = \nu_X(\gamma)\mu_Y^\gamma - \nu_Y(\gamma)\mu_X^\gamma - \gamma(\gamma-1)\mu_X^\gamma\mu_Y^\gamma\Delta_0 \,. \tag{2.24}$$

We then consider the acceptance region associated with the t-test of this linear hypothesis:

$$\hat{W}_F(\Delta_0)^2 = \left[\frac{\hat{\Theta}(\Delta_0)}{\left(\hat{V}[\hat{\Theta}(\Delta_0)]\right)^{1/2}}\right]^2 \leq z_{\alpha/2}^2 \tag{2.25}$$

where and $\hat{V}[\hat{\Theta}(\Delta_0)]$ is the relevant estimate of the variance of $\hat{\Theta}(\Delta_0)$.

Theorem 2. *Given a sample of observations on two random variable X and Y, consider two indices $GE_\gamma(X)$ and $GE_\gamma(Y)$ each defined as in* (2.1)*, and their difference $\Delta GE_\gamma := GE_\gamma(X) - GE_\gamma(Y)$ estimated using $\Delta\widehat{GE}_\gamma := \widehat{GE}_\gamma(X) - \widehat{GE}_\gamma(Y)$ each given as in* (2.3)*. The* $(1-\alpha)$*-level Fieller confidence set for ΔGE_γ, denoted* $\mathrm{FCS}[\Delta GE_\gamma; 1-\alpha]$*, takes the same form as in* (2.10) - (2.11) *with*

$$A_1 = \hat{\mu}_X^{2\gamma}\hat{\mu}_Y^{2\gamma}[\gamma^2-\gamma]^2, \quad B_1 = -2\hat{\mu}_X^\gamma\hat{\mu}_Y^\gamma[\gamma^2-\gamma][\hat{\nu}_X(\gamma)\hat{\mu}_Y^\gamma - \hat{\nu}_Y(\gamma)\hat{\mu}_X^\gamma],$$
$$C_1 = [\hat{\nu}_X(\gamma)\hat{\mu}_Y^\gamma - \hat{\nu}_Y(\gamma)\hat{\mu}_X^\gamma]^2, \tag{2.26}$$

and A_2, B_2 and C_2 are obtained using (2.22)*:*

$$A_2 = \left[\gamma[\gamma^2-\gamma]\hat{\mu}_X^{(\gamma-1)}\hat{\mu}_Y^\gamma\right]^2\frac{\hat{\sigma}_X^2}{n} + \left[\gamma[\gamma^2-\gamma]\hat{\mu}_Y^{(\gamma-1)}\hat{\mu}_X^\gamma\right]^2\frac{\hat{\sigma}_Y^2}{m} + 2\gamma^2[\gamma^2-\gamma]\hat{\mu}_X^{2\gamma-1}\hat{\mu}_Y^{2\gamma-1}\frac{\hat{\sigma}_{X,Y}}{n}, \tag{2.27}$$

$$\begin{aligned} B_2 \;=\; & 2\Big[\gamma^2[\gamma^2-\gamma]\hat{\mu}_X^{2(\gamma-1)}\hat{\mu}_Y^\gamma\hat{\nu}_Y(\gamma)\frac{\hat{\sigma}_X^2}{n} - \gamma[\gamma^2-\gamma]\hat{\mu}_X^{\gamma-1}\hat{\mu}_Y^{2\gamma}\frac{\hat{\sigma}_{X,X^\gamma}}{n} - \gamma^2[\gamma^2-\gamma]\hat{\mu}_Y^{2(\gamma-1)}\hat{\mu}_X^\gamma\hat{\nu}_X(\gamma)\hat{\sigma}_Y^2 \\ & + \gamma[\gamma^2-\gamma]\hat{\mu}_Y^{\gamma-1}\hat{\mu}_X^{2\gamma}\frac{\hat{\sigma}_{Y,Y^\gamma}}{m} + \gamma^2[\gamma^2-\gamma]\hat{\mu}_X^{\gamma-1}\hat{\mu}_Y^{\gamma-1}(\hat{\mu}_X^\gamma\hat{\nu}_Y(\gamma) - \hat{\mu}_Y^\gamma\hat{\nu}_X(\gamma))\frac{\hat{\sigma}_{X,Y}}{n} \\ & + \gamma[\gamma^2-\gamma]\hat{\mu}_X^{2\gamma-1}\hat{\mu}_Y^\gamma\frac{\hat{\sigma}_{X,Y^\gamma}}{n} - \gamma[\gamma^2-\gamma]\hat{\mu}_Y^{2\gamma-1}\hat{\mu}_X^\gamma\frac{\hat{\sigma}_{Y,X^\gamma}}{n}\Big], \end{aligned} \tag{2.28}$$

$$
\begin{aligned}
C_2 \;=\;& \left[\gamma\hat{\nu}_Y(\gamma)\hat{\mu}_X^{\gamma-1}\right]^2\frac{\hat{\sigma}_X^2}{n}+\hat{\mu}_Y^{2\gamma}\frac{\hat{\sigma}_{X^\gamma}}{n}-2\gamma\hat{\nu}_Y(\gamma)\hat{\mu}_X^{\gamma-1}\hat{\mu}_Y^{\gamma}\frac{\hat{\sigma}_{X,X^\gamma}}{n}+\left[\gamma\hat{\nu}_Y(\gamma)\hat{\mu}_Y^{\gamma-1}\right]^2\frac{\hat{\sigma}_Y}{m}+\hat{\mu}_X^{2\gamma}\frac{\hat{\sigma}_{Y^\gamma}}{m}\\
&-2\gamma\hat{\nu}_X(\gamma)\hat{\mu}_Y^{\gamma-1}\hat{\mu}_X^{\gamma}\frac{\hat{\sigma}_{Y,Y^\gamma}}{m}-2\gamma^2\hat{\mu}_X^{\gamma-1}\hat{\mu}_Y^{\gamma-1}\hat{\nu}_X(\gamma)\hat{\nu}_Y(\gamma)\frac{\hat{\sigma}_{X,Y}}{n}+2\gamma\hat{\mu}_X^{2\gamma-1}\hat{\nu}_Y(\gamma)\frac{\hat{\sigma}_{X,Y^\gamma}}{n}\\
&+2\gamma\hat{\mu}_Y^{2\gamma-1}\hat{\nu}_X(\gamma)\frac{\hat{\sigma}_{Y,X^\gamma}}{n}-2\hat{\mu}_X^{\gamma}\hat{\mu}_Y^{\gamma}\frac{\hat{\sigma}_{X^\gamma,Y^\gamma}}{n}, \qquad (2.29)
\end{aligned}
$$

imposing in turn: $\hat{\sigma}_{X,Y}$, $\hat{\sigma}_{X^\gamma,Y}$, $\hat{\sigma}_{X,Y^\gamma}$ and $\hat{\sigma}_{X^\gamma,Y^\gamma}$ equal to zero and $n=m$, under Assumption 1; $\hat{\sigma}_{X,Y}$, $\hat{\sigma}_{X^\gamma,Y}$, $\hat{\sigma}_{X,Y^\gamma}$ and $\hat{\sigma}_{X^\gamma,Y^\gamma}$ equal to zero under Assumption 2, and $n=m$ under Assumption 3.

Proof. To obtain a Fieller-type confidence set, we solve the inequality in (2.25) for Δ_0, as follows:

$$\hat{W}_F(\Delta_0)^2\leq z_{\alpha/2}^2 \quad\Leftrightarrow\quad \frac{\hat{\Theta}(\Delta_0)^2}{\hat{V}[\hat{\Theta}(\Delta_0)]}\leq z_{\alpha/2}^2 \quad\Leftrightarrow\quad \hat{\Theta}(\Delta_0)^2-z_{\alpha/2}^2\hat{V}[\hat{\Theta}(\Delta_0)]\leq 0. \qquad (2.30)$$

Here, $\hat{\Theta}(\Delta_0)^2$ and $\hat{V}[\hat{\Theta}(\Delta_0)]$ are quadratic functions of Δ_0 that can be expressed as follows:

$$\hat{\Theta}(\Delta_0)^2=A_1\Delta_0^2+B_1\Delta_0+C_1, \quad \hat{V}[\hat{\Theta}(\Delta_0)]=A_2\Delta_0^2+B_2\delta_0+C_2. \qquad (2.31)$$

As with Theorem 1, the above leads to

$$\mathrm{FCS}(\Delta GE_\gamma;1-\alpha)=\left\{\Delta_0 : A\Delta_0^2+B\Delta_0+C\leq 0\right\} \qquad (2.32)$$

which gives the proposed solution. □

The above presumes asymptotic normality of the underlying criteria. In fact, the considered measures are known transformations of two moments the estimators of which are asymptotically normal under standard regularity assumptions; see (Davidson and Flachaire, 2007; Cowell and Flachaire, 2007). These typically require that the first two moments exists and are finite. Asymptotic normality of the statistics in (2.6), (2.20), (2.15) and (2.25) thus follows straightforwardly. Nevertheless, convergence in this context is known to be slow, especially when the distribution of the data is heavy-tailed and with indices that are sensitive to the upper tail. Our simulations confirm these issues, yet the Fieller-based criteria perform better than the Delta method in finite samples because these eschew problems arising from the ratio.

3 Simulation evidence

This section reports the results of a simulation study designed to compare the finite-sample properties of FCS to the standard DCS in the one-sample and the two-sample problems. This will be done for the two popular inequality measures nested in the general entropy class of inequality measures: the Theil Index (GE_1), and half of the coefficient of variation squared (GE_2) which is related to the Hirschman-Herfindahl (HH) index.

We report the *rejection frequencies* of the tests underlying the proposed confidence sets, under both the null hypothesis (level control) and the alternative (power). Under the null hypothesis, these can also be interpreted as 1 minus the corresponding *coverage probability* for the associated confidence set. So we are studying here both the operating characteristics of tests used and the coverage probabilities of the confidence sets defined above. For further insight on confidence set properties, we also study the frequency of unbounded outcomes and the width of the bounded ones.

Since available inference methods perform poorly when the underlying distributions are heavy-tailed, we designed our simulation experiments to cover such distributions by simulating the data from the Singh-Maddala distribution, which was found to successfully mimic observed income distributions for developed countries such as Germany (Brachmann et al., 1995). Another reason to use the Singh-Maddala distribution is that it was widely used in the literature which makes our results directly comparable to previously proposed inference methods. The CDF of the Singh-Maddala distribution can be written as

$$F_X(x) = 1 - \left[1 + \left(\frac{x}{b_X}\right)^{a_X}\right]^{-q_X} \tag{3.1}$$

where a_X, q_X and b_X are the three parameters defining the distribution. a_X influences both tails, while q_X only affects the right tail. The third parameter (b_X) is a scale parameter to which we give little attention as the inequality measures considered in this paper are scale invariant. This distribution is a member of the five-parameter generalized beta distribution and its upper tail behaves like a Pareto distribution with a tail index equal to the product of the two shape parameters a_X and q_X ($\xi_X = a_X q_X$). The k-th moment exists for $-a_X < k < \xi_X$ which implies that a sufficient condition for the mean and the variance to exist is $-a_X < 2 < \xi_X$.

The moment of order γ of Singh-Maddala distribution have the following closed form:

$$\nu_X(\gamma) := \mathbb{E}(X^\gamma) = \frac{b_X^\gamma \Gamma\left(\gamma a_X^{-1} + 1\right) \Gamma\left(q_X - \gamma a_X^{-1}\right)}{\Gamma(q_X)} \tag{3.2}$$

where $\Gamma(\cdot)$ is the gamma function. For $\gamma = 1$, this yields the mean of X [$\mu_X = \nu_X(1) = \mathbb{E}(X)$] and, for $\gamma = 2$, the second moment of X [$\nu_X(2) = \mathbb{E}(X^2)$]. Similarly, replacing X by Y in the above expressions, we can compute μ_Y and $\nu_Y(2)$. Using the values of these moments, we compute analytical expressions for $GE_\gamma(X)$ and $GE_\gamma(Y)$. Each experiment involves 10000 replications and sample sizes of $n = 50,\ 100, 250,$ $500,\ 1000,\ 2000$. The nominal level α is set at 5%.

3.1 Simulation results: one-sample problem

Dufour et al. (2018) proposed Fieller-type confidence sets for the Theil index [GE_1] and showed, in a simulation study, that it improves coverage compared to the Delta method. In this section, we provide additional evidence on the superiority of the Fieller-type method by considering the GE_2 index. Following the literature in this area, we use a Singh-Maddala distribution with parameters $a_X = 2.8$, $q_X = 1.7$ as benchmark [$X \sim SM_X(a_X = 2.8, b, q_X = 1.7)$]. We study the finite-sample size and power behavior of Fieller-type method and the Delta method as we deviate from the benchmark case towards heavy-tailed distributions.

The tests reported here involve null hypotheses of the form $H_0 : GE_\gamma = \delta_0$ (where $\gamma = 1$ or 2) and can be performed in two different ways. For the Delta method, we can either use the critical region $\hat{T}_D(\delta_0)^2 > z^2_{\alpha/2}$, where $\hat{T}_D(\delta_0)^2$ is defined in (2.6), or check whether the confidence set $\mathrm{DCS}[GE_\gamma(X); 1-\alpha]$ defined in (2.8) contains the tested value δ_0. Similarly, for the Fieller method, we can either use the critical region $\hat{T}_F(\delta_0)^2 > z^2_{\alpha/2}$, where $\hat{T}_F(\delta_0)^2$ is defined in (2.15), or check whether the confidence set $\mathrm{FCS}[GE_\gamma(X); 1-\alpha]$ defined in (2.18) contains the tested value δ_0. Both approaches are numerically equivalent and yield the same results. If $\mathbb{P}[\hat{T}_F(\delta_0)^2 > z^2_{\alpha/2}] = p(\delta_0)$ for a distribution which satisfies $GE_\gamma = \delta_0$, then $\mathbb{P}\left[\delta_0 \in \mathrm{FCS}[GE_\gamma(X); 1-\alpha]\right] = 1 - p(\delta_0)$ is the coverage probability for δ_0 in this case, and similarly for $\hat{T}_D(\delta_0)^2$.

The left panel of Figure 1 plots the rejection frequencies of tests for the Theil index based on the Fieller and Delta methods, under the following Singh-Maddala null distribution: $X \sim SM_X(a_X = 1.1, q_X = 4.327273)$. For small sample (50 observations), the Fieller-type method reduces size distortions by about 3 percentage points. As n increases, size distortions shrink and both methods converge to the same level.

In contrast with the Theil index, the GE_2 index puts more emphasis on the right tail of the distribution. In this case, the Fieller-type method exhibits a greater advantage in terms of reliability [see the left panel of Figure 2]. For $n = 50$, the Delta method rejection frequency is 38%, while that of Fieller-type method are around 26.2%, thereby reducing the size distortion by more than 11%. The relative robustness of the Fieller method to the changes in the upper tail makes it an attractive alternative to the Delta method, which is known to perform poorly when the underlying distributions are characterized by thick right tails.

Another important observation about the Fieller-type method is that it is less distorted by the shape of the left tail. As we will show shortly, our results indicate that for small samples, size distortions caused by thick left tails are smaller with the Fieller-type method than with the Delta method. Table 1 reports the percentage difference of the rejection frequencies between both methods as the left tail becomes thicker. The simulation design behind the results starts with a lighter left tail ($a_X = 3.173$) and make it thicker by decreasing the value of a_X down to 1.1. To focus solely on the left tail, we fix the tail index ($\xi_X = a_X q_X$) at 4.76. We do so by increasing the parameter q_X sufficiently enough to offset the impact of a_X on ξ_X. In the second part of the table, we consider a smaller tail index ($\xi_X = 3.64$).

As we move down the table, the left tail becomes thicker, which negatively affects the performances of both the Delta and the Fieller-type methods, though the latter exhibits smaller level distortions. Thus, the Fieller method is less negatively affected by a thick left tail. This is true for the Theil index and the GE_2 index. As the left tail becomes more thick, the percentage difference of the rejection frequencies for the Theil index with $\xi_X = 4.67$ steadily increases from 1.53% to around 13.45% for very thick left tails. For GE_2, the percentage difference of the rejection frequencies is more prominent increasing from 5.91% to around 30.9%. Similar conclusions can be drawn from the lower part of the table which considers a thicker right tail (smaller tail index, $\xi_X = 3.64$).

To study power, we consider DGPs which deviate from the null hypothesis. We mainly change the shape parameter a_X as it affects both the left and the right tails. The right panels of Figures 1 and 2 plots the powers of both methods. To compare power, we focus on sample sizes at which the two methods have similar size performance (i.e., when the sample size is 500). As can be seen from the plots, both methods are equally powerful.

3.2 Simulation results: two-sample problem

We will now consider the problem of testing hypotheses of the form $H_0(\gamma) : GE_\gamma(X) - GE_\gamma(Y) = \Delta_0$, for each one of the inequality indices we focus on ($\gamma = 1$ or 2). Even though we emphasize the important problem of testing equality ($\Delta_0 = 0$), we also consider the problem of testing nonzero differences ($\Delta_0 \neq 0$). Our simulation experiments accommodate three possible cases which can arise in practice: (1) independent samples of equal sizes, (2) independent samples of unequal sizes, (3) dependent samples with equal sizes. More specifically, the study presented here covers the following three specifications for each of the three cases: (1) the underlying distributions are identical [$\Delta_0 = 0$ with $F_X = F_Y$]; (2) the two indices are equal and the underlying distributions under the null hypothesis are not identical [$\Delta_0 = 0$ with $F_X \neq F_Y$]; (3) the two indices are unequal [$\Delta_0 \neq 0$]. This leaves us with 9 possible cases, as follows.

1. Experiment I – Independent samples of equal sizes ($m = n$): (a) $\Delta_0 = 0$ with $F_X = F_Y$; (b) $\Delta_0 = 0$ with $F_X \neq F_Y$; and (c) $\Delta_0 \neq 0$ (hence $F_X \neq F_Y$).

2. Experiment II – Independent samples of unequal sizes ($m \neq n$):(a) $\Delta_0 = 0$ with $F_X = F_Y$; (b) $\Delta_0 = 0$ with $F_X \neq F_Y$; and (c) $\Delta_0 \neq 0$ (hence $F_X \neq F_Y$).

3. Experiment III – Dependent samples of equal sizes ($m = n$):
   (a) $\Delta_0 = 0$ with $F_X = F_Y$; (b) $\Delta_0 = 0$ with $F_X \neq F_Y$; and (c) $\Delta_0 \neq 0$ (hence $F_X \neq F_Y$).

As in the one-sample problem, the simulation results are presented graphically through plotting the rejection frequencies against the number of observations. When the number of observations is different between the two samples, we plot the rejection frequencies against the number of observations of the smallest sample.

As in the one-sample problem, the tests reported here can be performed in two different ways. For the Delta method, we can either use the critical region $\hat{W}_D(\Delta_0)^2 > z^2_{\alpha/2}$, where $\hat{W}_D(\Delta_0)^2$ is defined in (2.20), or check whether the confidence set $\mathrm{DCS}(\Delta GE_\gamma; 1-\alpha)$ defined in (2.21) contains the tested value Δ_0. Similarly, for the Fieller method, we can either use the critical region $\hat{W}_F(\Delta_0)^2 > z^2_{\alpha/2}$, where $\hat{W}_F(\Delta_0)^2$ is defined in (2.25), or check whether the confidence set $\mathrm{FCS}(\Delta GE_\gamma; 1-\alpha)$ defined in (2.32) contains the tested value Δ_0. Both approaches are numerically equivalent and yield the same results (a feature we did check). If $\mathbb{P}[\hat{W}_F(\Delta_0)^2 > z^2_{\alpha/2}] = p(\Delta_0)$ for a pair of distribution which satisfy $GE_\gamma(X) - GE_\gamma(Y) = \Delta_0$, then $\mathbb{P}\big[\Delta_0 \in \mathrm{FCS}[GE_\gamma(X); 1-\alpha]\big] = 1 - p(\Delta_0)$ is the coverage probability for Δ_0 in this case, and similarly for $\hat{W}_D(\Delta_0)^2$.

The powers of FCS and DCS are investigated by considering DGPs which do not satisfy the null hypothesis. We do so by considering DGPs with a lower value of the shape parameter a_X and a higher value of the shape parameter a_Y. Thus, we are deviating from the null hypothesis by assuming distributions with heavier left and right tails to draw the first sample, and distributions with less heavy left and right tails to draw the second sample. The rejection frequencies under the alternative are not size-controlled, yet we compare power when both methods have similar sizes.

Our extensive simulation study reveals several important results. First, the Fieller-type method outperforms the Delta method under most specifications, and when it does not, it performs as well as the Delta method. Put differently, the Fieller-type method was never dominated by Delta method. Second, the Fieller-type method is more robust to irregularities arising from both the left and right tails. Third, the Fieller-type method gains become more sizeable as the sensitivity parameter γ increases. Fourth, the performance of the Fieller-type method matches, and for some cases exceeds, the permutation method which is considered one of the best performing methods proposed in the literature so far for the two-sample problem. In the reminder of this section we take a closer look at the simulation evidence supporting the above findings.

Experiment I: Independent samples of equal sizes – The left panels of Figures 3 and 4 depict the rejection frequencies against the sample size for GE_1 and GE_2 respectively. Here the distributions are assumed identical $[F_X = F_Y]$. Comparing the two panels, we notice that better size control with the Fieller-type method is more noticeable for GE_2: the size gains are larger when the index used is more sensitive to the changes in the right tail of the underlying distributions. As the sample size increases the rejection probabilities of the two methods converge to the same level.

In the second specification, the indices are identical, but the underlying distributions are not $[\Delta_0 = 0 \text{ with } F_X \neq F_Y]$. The left panel of Figure 5 plots the FCS and DCS rejection frequencies for this scenario. Again, the results suggest that the Fieller-type method outperforms the Delta method in small samples in terms of size, and the gains are most prominent for GE_2. The gains are smaller in this scenario compared to the previous one. As we will show later, the Fieller-type method will not solve the over-rejection problem under all scenarios, but it will reduce size distortions in many cases, and when it does not, it performs as well as the Delta method.

We now move to the third scenario, where we consider different distributions under the null hypothesis and unequal inequality indices $[\Delta_0 \neq 0]$. In this scenario, the difference under the null hypothesis can take any admissible value (possibly different from zero). Testing a zero value, although informative, does not always translate into a confidence interval. Hence, one of our contributions lies in considering the non-zero null hypothesis which allows us to rely for inference on the more-informative confidence sets approach rather than testing the equality of the difference between the two indices to one specific value.

The results, as shown in the left panels of Figures 7 and 8, suggest a considerable improvement. In both panels, the Fieller-type method leads to size gains and almost achieves correct size. The improvements are more pronounced for the GE_2 index. The right panels of Figures 3 to 8 illustrate the power of FCS and DCS for both GE_1 and GE_2 under the three scenarios considered: $[\Delta_0 = 0 \text{ with } F_X = F_Y]$, $[\Delta_0 = 0 \text{ with } F_X \neq F_Y]$ and $[\Delta_0 \neq 0]$ respectively. The results show that the Fieller-type method is as powerful as the Delta method when compared at sample sizes where both FCS and DCS have similar empirical rejection frequencies.

Experiment II: Independent samples of unequal sizes – Empirically, when comparing inequality levels spatially or over time, it is unlikely one encounters samples with the same size. Thus, it is useful to assess the performance of our proposed method when the sample sizes are unequal. To do so, we adjust our simulation design by setting the number of observations of the second sample to be as twice as large as the first sample. If we denote the size of the first sample by n and that of the second by m, then $n = 2m$.[4] The results are analogous to those obtained in the first experiment, under which sample sizes were equal, in the sense that the Fieller-type method improves level control for both GE_1 and GE_2, with a larger improvement for GE_2. The size and power simulation results for the three scenarios considered here are available in the online appendix.

Experiment III: Dependent samples of equal sizes – Another interesting case is the one where the samples are dependent. This occurs mostly when comparing inequality levels before and after a policy change, such as comparing pre-tax and post-tax income inequality levels, or comparing the distributional impact of a macroeconomic shock. To accommodate for such dependencies, we modify the simulation design as follows: the samples are drawn in pairs from the joint distribution, which we denote F_{XY}, where the correlation between the two marginal distributions is generated using a Gumbel copula with a high Kendall's correlation coefficient of 0.8. For this case, results are in line with the independent cases, in small samples and when larger γ is used. Size and power plots are available in the online appendix.

Comparing the Fieller-type method with the permutation method – As outlined in the introduction, the permutation-based Monte-Carlo test approach proposed in Dufour et al. (2019) stands out as one of the best performing nonparametric inference method for testing the equality of two inequality indices. The authors focus on the Theil and the Gini indices. The permutation testing approach provides exact inference when the null distributions are identical $(F_X = F_Y)$ and it leads to a sizable size distortion reduction when the null distributions are sufficiently close $(F_X \approx F_Y)$. However, as the null distributions differ, the performance of the method deteriorates.

Figures 9 and 10 plot size and power of the permutation Fieller-type methods against the tail index of F_Y. As in Dufour et al. (2019), we fix the tail index of the null distribution F_X to 4.76. When the distributions under the null hypothesis are identical, the permutation method is exact and thus it is important to compare methods when exactness does not hold. Our results point to two main advantages of the Fieller-type method over the permutation method: for the Theil index, the Fieller-type method is more powerful and these power gains are magnified as the difference between the indices becomes larger. On the other hand,

[4]The results presented here are not sensitive to choice of the ratio between n and m

when considering the GE_2, there are size gains mainly when the tail index is relatively small (*i.e.*, when the right tail is heavier). These size gains are not associated with power loss as the right panel of the same figure illustrates.

The attraction of the Fieller-type method with respect to the permutation approach goes beyond the superior performance highlighted above. Unlike the Fieller-type method, its applicability is restricted to the null hypothesis of equality ($\Delta_0 = 0$), and further theoretical developments would be needed to test more general hypotheses. Building confidence intervals using a permutation-based or another simulation-based method (such as the bootstrap) would also require a computationally intensive numerical inversion (*e.g.*, through a grid search). So another appealing feature of the Fieller-type approach comes from the fact that it is computationally easy to implement.

Behavior with respect to the tails – To better understand under what circumstances does the Fieller-type method improves level control, we assess the performance of the proposed method to different tail shapes. The literature has focused on the role of heavy right tails in the deterioration of the Delta method confidence sets. However, as our results indicate, heavy left tails also contribute to the under-performance of the standard inference procedures. The Fieller-type method is less prone to such irregularities arising from both ends of the distributions and thus it reduces size distortions whether the cause of the under-performance is arising from the left tail or the right tail. This is supported by the results reported below in Tables 2 and 3. The results in these tables rely on samples of 50 observations. Table 2 reports the percentage difference of the rejection frequencies as the right tails of the two distributions become thicker. The right-tail shape is determined by the tail index ($\xi_X = a_X q_X$). The smaller the tail index, the thicker is the right tail of the distribution under consideration. The reliability advantage of the Fieller-type method (over the Delta method) increases as the right tail of the distributions gets thicker.

To study the impact of the left tail, the parameters of the first distribution are fixed at $a_X = 2.8$ and $q_X = 1.7$, while a_Y and q_Y are varied such that the left tail becomes thicker and the right tail is left unchanged. This is done by decreasing a_Y, and increasing q_Y enough to keep the tail index fixed ($\xi_X = \xi_Y = 4.76$). The last column of Table 3 shows the percentage difference of the rejection frequencies between the Fieller-type and Delta methods. As the left tail thickens, the performance of the Delta method deteriorates relative to the Fieller-type method, and thus the Fieller method better captures irregularities in the left tail. This conclusion holds regardless of whether the left tail of the second distribution is lighter or thicker than the left tail of the first distribution.

Fieller-type method and the sensitivity parameter γ – A consistent conclusion from our results is that the Fieller's-induced size gains are more prominent for GE_2 compared to GE_1, that is, when the sensitivity parameter γ increases from 1 to 2. This might suggest that as γ increases, size gains from the Fieller-type method increase. Such generalization is indeed supported by simulation evidence illustrated by Figure 11. The left panel plots rejection frequencies of DCS and FCS for $\gamma \in [0, 3.5]$ for independent samples. The right panel considers dependent samples. As γ becomes larger, FCS outperforms DCS at an increasing rate. The superiority of the Fieller-type method in this context is unaffected by the independence assumption as shown in the right panel where the rejection frequencies are plotted against γ for dependent samples with Kendall's correlation of 0.8.

Recall that the parameter γ characterizes the sensitivity of the index to changes at the tails of the distribution. For instance, the index becomes more sensitive to changes at the upper tails as γ increases (assuming positive γ). Thus, relative to the Delta method, the performance of the Fieller-type method in the two-sample problem improves as the right tail of the underlying distributions becomes heavier. This conclusion, as we saw from the results above, is robust to the assumptions about the independence of the samples and to the distance between the two null distributions.

The identical performance of the Fieller-type method and Delta method at $\gamma = 0$ is expected as the underlying t-tests inverted in the process of building FCS and DCS are identical since the null hypothesis is no longer a ratio. To see that, recall that the limiting solution for $GE_\gamma(\cdot)$ at $\gamma = 0$ is equal to $\mathbb{E}_F[\log(X)] - \log[\mathbb{E}_F(X)]$. Graphically, we can see that both methods start off at the same rejection frequencies when $\gamma = 0$, and then diverge as γ increases.

Robustness to the shapes of the null distributions – So far, our simulation experiments have focused on comparing the finite-sample performance of FCS and DCS by studying their behavior as the number of observation increases, holding the parameters of the two underlying null distributions constant. Here we try to check the robustness of our results by fixing the number of observations at 50 and allowing the parameters $(a_X, q_X$, a_Y and $q_Y)$ to vary. This type of analysis highlights the (in)sensitivity of our conclusions regarding the Fieller-type method to the shape of the null distributions. In left panel of Figure 12, we plot the rejection frequencies of both methods against the sensitivity parameter ξ_X for the Theil index. We set ξ_X equal to 4.76 and allow ξ_Y to vary between 3.05 and 6.255. In the right panel, we focus on the GE_2 index. Here ξ_X is fixed at 4.76 again and the parameter ξ_Y ranges between 3.293 and 5.7107.

For small samples, the gains of the Fieller-type method are maintained regardless the shape of the distribution. The gains are more pronounced for GE_2 compared to GE_1. These two graphs show that the gains attained by the Fieller-type method are not arbitrary and that they hold for various parametric assumptions of the underlying distributions.

Slow convergence – Inequality estimates are characterized by slow convergence when underlying distributions are heavy-tailed. This problem has in fact motivated most of the proposed asymptotic refinements in this literature [see Davidson and Flachaire (2007); Cowell and Flachaire (2007)]. Our results in Table 4 and Table 5 corroborate this fact, as over-rejections remain even with samples as large as 200000, particularly with the GE_2 which puts more weight on the upper tail of the distribution. On balance, our main finding is the superiority of the Fieller method in finite samples.

Widths of the confidence sets – The last two columns of Table 5 show the average widths of the FCS and the DCS for the two sample problem. Since the Fieller's method can produce unbounded confidence sets, we take the average of the widths based on the bounded confidence sets. In general, compared to the FCS widths, the DCS widths are shorter with small samples, *i.e.* they are shorter when the Delta method rejection frequencies are higher than those of Fieller. This suggest that the DCS are too short and thus they tend to undercover the true difference between the indices. As the sample size increases, the two methods exhibit similar performance and the widths coincides. This is true as well for the one-sample problem as the last two columns of table 4 illustrate.

4 Application: Regional income convergence

In this section, we present empirical evidence on the relevance of our theoretical results to applied economic work. We assess economic convergence across the U.S. states between 1946 and 2016. One of the motivating factors behind the choice of the convergence question is the small number of observations, which represents an ideal opportunity to assess the empirical value of our theoretical findings as our simulation results have shown that improvements via the Fieller-type method are most prominent when sample sizes are small. In what follows, unless stated otherwise, tests and confidence sets are at the 5% level.

The late 1980's witnessed a new wave of interest in economic convergence that was spurred by the revival of growth models. The convergence hypothesis, first theorized by the popular Solow growth model, postulates that in the long run, economies will converge to similar per-capita income levels. The convergence question is important from theoretical and policy perspectives.

Theoretically, Romer (1994) and Rebelo (1991) argue that the rejection of the convergence hypothesis provides empirical support for the endogenous growth model and evidence against the neoclassical growth model. In the latter models, per-capita income convergence results from the diminishing return to capital assumption. This assumption implies that the return to capital increases in economies with low level of capital and decreases in capital-abundant economies. Moreover, since the rate of return on capital is higher in poorer economies, investments will migrate from rich economies to poorer ones, further enhancing growth and reducing the gap between them. On the other hand, in endogenous growth models as in Romer (1994) and Rebelo (1991), the diminishing rate of return on capital is considered implausible once knowledge is assumed to be one of the production factors. Thus, the model does not predict convergence, but on the contrary predicts that divergence might occur.

Empirically, policy-makers are interested in learning about the dynamics of income dispersion across regions/states so they can engage in redistributive policies when needed or to assess the distributional impact of a specific policy. Among the various definitions of convergence provided in the literature, two definitions appear to dominate the work on this topic: β-convergence and σ-convergence (Barro, 2012; Barro and Sala-i Martin, 1992; Quah, 1996; Sala-i Martin, 1996; Higgins et al., 2006). Although related, these two measures might lead to different conclusions as they capture different dimensions of economic convergence. For an analytical treatment of the relationship between the two measures, see Higgins et al. (2006).

β-convergence occurs when there is a negative relationship between the growth rate and the initial level of per-capita income, that is, when poor economies grow at a faster rate than the rich ones. The σ-convergence concept focuses on the dispersion of the income distribution which is typically measured in this literature by the variance of the logs. The variance of logs is scale-independent and thus multiplying the per-capita incomes by a scale k has no impact on the dispersion level. Alternative scale-independent measures of dispersion such as inequality measures have generally not been utilized in convergence analysis. The only exception is Young et al. (2008) which reported the Gini coefficient for comparison purposes with reference to the variance of logs.

One feature of inequality measures such as the Gini coefficient and the GE measures is that they respect the Pigou-Dalton principle, which states that a rank preserving transfer from a richer individual/state to a poorer individual/state should make the distribution at least as equitable. In the context of economic convergence, this principle is particularly relevant. For instance, if the US government makes a transfer from a richer state to a poorer one, one would expect dispersion between states to decline. The Gini and GE measures would capture this decline, whereas the variance of logs might indicate no change or even an increase in dispersion. The fact that the variance of logs violates the Pigou-Dalton principle is usually neglected in the literature on the grounds that the problem occurs only at the extreme right tail of the distribution. However, Foster and Ok (1999) show that disagreement between the variance of logs and inequality measures can result from changes in incomes in other parts of the distribution including the left tail. The following example (Foster and Ok, 1999) underscores the importance of the Pigou-Dalton principle and its implications for convergence. Consider two income distributions defined by the following incomes $(2, 5, 10, 28, 40)$ and $(2, 5, 10, 34, 34)$ where the latter is associated with a transfer from the richest [40 to 34] incomes to poorer ones [28 to 34]. The resulting change in the variance of logs, from 1.5125 to 1.5154, suggests an increase of inequality. In contrast, the GE_2 index declines from 0.3696 to 0.3446, thereby capturing the expected distributional impact of such a transfer.

Our empirical analysis of per-capita income dispersion across the US is motivated by comparably peculiar statistics. Consider the publicly available per-capita income at the state level for 48 out of the 50 states (as the data for Alaska and Hawaii is not available). The variance of logs between the years 2000 and 2016 indicates a 3% increase in dispersion, whereas GE_2 indicates a decline in dispersion by 0.3%. This provides a compelling basis for the more comprehensive inferential analysis reported next.

Using the same data source, we first compute the Theil index for the per-capita income distributions of 1946 and 2016. Then we construct the Delta and Fieller confidence sets for the difference between the two indices. A standard interpretation of differences between the two confidence intervals (at the considered level) implies that one will reject the null hypothesis $\Delta GE_\gamma = \Delta_0$ for a given Δ_0 while the other fails to reject it. Special attention should be paid to the $\Delta_0 = 0$ case, as decisions might reverse the conclusion on whether convergence holds or not.

Using the Theil index, our results in the first column of Table 6 indicate that per-capita income inequality across states has declined between 1946 and 2016. The decline in inequality implies convergence. This is compatible with the general convergence trend reported in the literature (Barro and Sala-i Martin, 1992; Bernat Jr, 2001; Higgins et al., 2006). Although the Fieller and Delta-method confidence sets are not identical, they still lead to the same conclusion which is that the decline of inequality is statistically different from zero at the level used.

In the second column of Table 6, we consider the same problem using GE_2 index rather than the Theil one. This index puts more weight on the right tail of the distribution. In this case, the results also indicate a decline of inequality across states. Inequality in 1946 was 0.02679 and declined by -0.01163 by 2016. The confidence sets based on the Delta and Fieller-type methods lead to opposite conclusions about the statistical significance of the decline in inequality. DCS fails to reject the null hypothesis of no change in inequality, thus the decline in inequality based on DCS is not statistically different from zero. On the other hand, the Fieller-type methods rejects the hypothesis of no inequality change, which entails that the decline is significant.

In addition to DCS and FCS, we report the permutational p-values. For the GE_2, the p-value is less than 5% and thus we reject the null hypothesis of no change in inequality contradicting the conclusion based on the Delta method. This constitutes an empirical evidence supporting the findings of Dufour et al. (2019).

Two conclusions can be drawn from our findings. First, the Fieller-type and the Delta methods can lead to different confidence sets in practice which documents the empirical relevance of our theoretical findings. Second, disparities between both sets can lead to spurious conclusions about inequality changes if one set includes zero while the other does not. From a policy point of view, this disparity is crucial, especially if important policy actions are motivated by underlying analysis.

We next turn to non-OECD countries between 1960 and 2013. Table 7 presents estimates and confidence sets for the difference of inequality measures between the two periods. The main finding here is that the Fieller-type confidence set based on the GE_2 index is the whole real line $\mathbb{R}$. These results confirm that decisions based on Delta-method are spurious, and that a no-change conclusion is flawed: data and measure are, instead, uniformative.

The permutational method leads results similar to Delta and the Fieller-type methods for non-OECD countries. Available permutation tests although preferable size-wise to their standard counterparts, are difficult to invert to build confidence sets. Instead, the confidence sets proposed here can be unbounded and thus avoid misleading statistical inferences and policy decisions, in particular from seemingly insignificant tests. The econometric literature on inequality has long emphasized the need to avoid over-sized tests. Rightfully, spurious rejections are misleading. Our results document a different although related problem: even with adequately sized no-change tests, weak identification can undercut the reliability of policy advice resulting form insignificant no-change test outcomes. Far more attention needs to paid to confidence sets. Moreover, sets that can be unbounded make empirical and policy work far more credible than it can be using bounded alternatives or no-change tests that cannot be inverted.

5 Conclusion

This paper introduces a Fieller-type method for inference on the GE class of inequality indices, in the one and two-sample problem with a focus on the latter. Simulation results confirm that the Fieller-type method outperforms standard counterparts including the permutation test, over all experiments considered. Size gains are most prominent when using indices that put more weight on the right tail of the distribution and results are robust to different assumptions about the shape of the null distributions. While irregularities arising from the right tail have long been documented, we find that left tail irregularities are equally important in explaining the failure of standard inference methods. On recalling that permutation tests are difficult to invert, our results underscore the usefulness of the Fieller-type method for evidence-based policy. An empirical analysis of economic convergence reinforces this result, and casts a new light on traditional controversies in the growth literature.

Fieller's approach is frequently applied in medical research and to a lesser extent in applied economics despite its solid theoretical foundations (Srivastava, 1986; Willan and O'Brien, 1996; Johannesson et al., 1996; Laska et al., 1997). This could be due to the seemingly counter-intuitive non-standard confidence sets it produces which economists often find hard to interpret. Consequently, many applied researchers encountering the estimation of ratios avoid using it and opt to use methods that yield closed intervals regardless of theoretical validity. This paper illustrates serious empirical and policy flaws that may result from such practices in inequality analysis.

References

Andrews, D. W. and Cheng, X. (2013). Maximum likelihood estimation and uniform inference with sporadic identification failure. *Journal of Econometrics*, 173(1):36–56.

Andrews, I. and Mikusheva, A. (2015). Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics*, 6(1):123–152.

Araar, A. and Duclos, J.-Y. (2006). *Poverty and equity: Measurement, policy, and estimation with DAD*. Springer, New York.

Bahadur, R. R. and Savage, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *The Annals of Mathematical Statistics*, 27(4):1115–1122.

Barro, R. J. (2012). Convergence and modernization revisited. Technical report, National Bureau of Economic Research.

Barro, R. J. and Sala-i Martin, X. (1992). Convergence. *Journal of Political Economy*, 100(2):223–251.

Beaulieu, M.-C., Dufour, J.-M., and Khalaf, L. (2013). Identification-robust estimation and testing of the zero-beta CAPM. *Review of Economic Studies*, 80(3):892–924.

Bennett, B. (1959). On a multivariate version of Fieller's theorem. *Journal of the Royal Statistical Society: Series B (Methodological)*, 21(1):59–62.

Bernard, J.-T., Chu, B., Khalaf, L., Voia, M., et al. (2019). Non-standard confidence sets for ratios and tipping points with applications to dynamic panel data. *Annals of Economics and Statistics*, (134):79–108.

Bernard, J.-T., Idoudi, N., Khalaf, L., and Yélou, C. (2007). Finite sample inference methods for dynamic energy demand models. *Journal of Applied Econometrics*, 22(7):1211–1226.

Bernat Jr, G. A. (2001). Convergence in state per capita personal income, 1950-99. *Survey of Current Business*, 81(6):36–48.

Bertanha, M. and Moreira, M. J. (2016). Impossible inference in econometrics: Theory and applications. *arXiv preprint arXiv:1612.02024*.

Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption inequality and partial insurance. *American Economic Review*, 98(5):1887–1921.

Bolduc, D., Khalaf, L., and Yélou, C. (2010). Identification robust confidence set methods for inference on parameter ratios with application to discrete choice models. *Journal of Econometrics*, 157(2):317–327.

Brachmann, K., Stich, A., and Trede, M. (1995). Evaluating parametric income distribution models. *Allge-*

meines Statistisches Archiv, 80:285–298.

Cowell, F. A. (2000). Measurement of inequality. In *Handbook of Income Distribution*, volume 1, pages 87–166. Elsevier, Amsterdam.

Cowell, F. A. and Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics*, 141(2):1044–1072.

Cowell, F. A. and Flachaire, E. (2015). Statistical methods for distributional analysis. In *Handbook of Income Distribution*, volume 2, pages 359–465. Elsevier, Amsterdam.

Cox, D. R. (1967). Fieller's theorem and a generalization. *Biometrika*, 54(3-4):567–572.

Davidson, R. and Flachaire, E. (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics*, 141(1):141–166.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65(6):1365–1387.

Dufour, J.-M., Flachaire, E., and Khalaf, L. (2019). Permutation tests for comparing inequality measures. *Journal of Business & Economic Statistics*, 37(3):457–470.

Dufour, J.-M., Flachaire, E., Khalaf, L., and Zalghout, A. (2018). Confidence sets for inequality measures: Fieller-type methods. In Green, W., Khalaf, L., Makdissi, P., Sickles, R., Veall, M., and Voia, M., editors, *Productivity and Inequality*. Springer, Cham.

Dufour, J.-M. and Hsiao, C. (2008). Identification. In *The New Palgrave Dictionary of Economics*, The New Palgrave Dictionary of Econometrics. Palgrave MacMillan, 2nd edition.

Dufour, J.-M. and Jasiak, J. (2001). Finite sample limited information inference methods for structural equations and models with generated regressors. *International Economic Review*, 42(3):815–844.

Dufour, J.-M. and Taamouti, M. (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica*, 73(4):1351–1365.

Dufour, J.-M. and Taamouti, M. (2007). Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. *Journal of Econometrics*, 139(1):133–153.

Foster, J. E. and Ok, E. A. (1999). Lorenz dominance and the variance of logarithms. *Econometrica*, 67(4):901–907.

Franz, V. H. (2007). Ratios: A short guide to confidence limits and proper use. *arXiv preprint arXiv:0710.2024*.

Gleser, L. J. and Hwang, J. T. (1987). The nonexistence of 100 (1-α)% confidence sets of finite expected diameter in errors-in-variables and related models. *The Annals of Statistics*, 15(4):1351–1362.

Higgins, M. J., Levy, D., and Young, A. T. (2006). Growth and convergence across the United States: Evidence from county-level data. *The Review of Economics and Statistics*, 88(4):671–681.

Hirschberg, J. and Lye, J. (2010). A geometric comparison of the delta and Fieller confidence intervals. *The American Statistician*, 64(3):234–241.

Johannesson, M., Jönsson, B., and Karlsson, G. (1996). Outcome measurement in economic evaluation. *Health economics*, 5(4):279–296.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, 73(4):1103–1123.

Koschat, M. A. et al. (1987). A characterization of the Fieller solution. *The Annals of Statistics*, 15(1):462–468.

Laska, E. M., Meisner, M., and Siegel, C. (1997). Statistical inference for cost–effectiveness ratios. *Health economics*, 6(3):229–242.

Maasoumi, E. (1997). Empirical analyses of inequality and welfare. *Handbook of applied econometrics*, 2:202–245.

Quah, D. T. (1996). Empirics for economic growth and convergence. *European Economic Review*, 40(6):1353–1375.

Rebelo, S. (1991). Long-run policy analysis and long-run growth. *Journal of Political Economy*, 99(3):500–521.

Romer, P. M. (1994). The origins of endogenous growth. *Journal of Economic Perspectives*, 8(1):3–22.

Sala-i Martin, X. X. (1996). The classical approach to convergence analysis. *The Economic Journal*, 106(437):1019–1036.

Schluter, C. (2012). On the problem of inference for inequality measures for heavy-tailed distributions. *The Econometrics Journal*, 15(1):125–153.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica*, 48(3):613–625.

Srivastava, M. (1986). Multivariate bioassay, combination of bioassays, and Fieller's theorem. *Biometrics*, 42(1):131–141.

Theil, H. (1967). *Economics and information theory*. North-Holland, Amsterdam.

Willan, A. R. and O'Brien, B. J. (1996). Confidence intervals for cost-effectiveness ratios: An application of Fieller's theorem. *Health Economics*, 5(4):297–305.

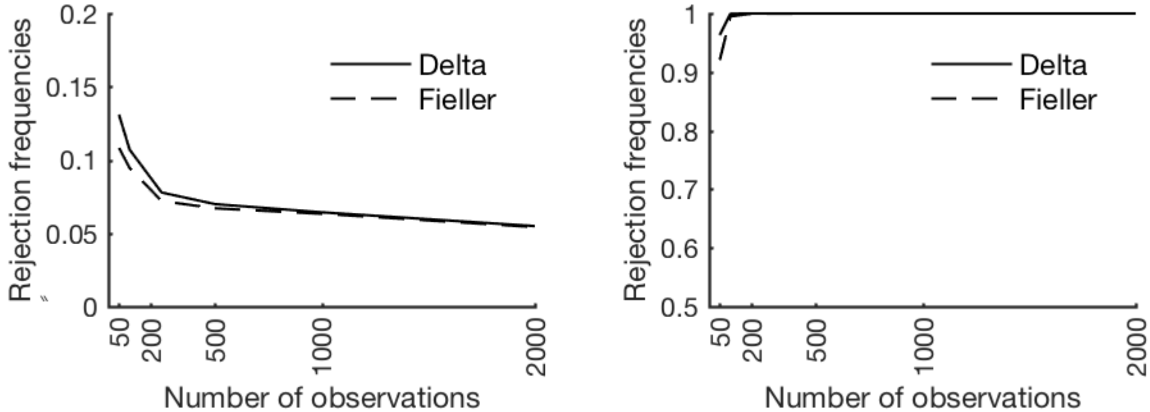
Young, A. T., Higgins, M. J., and Levy, D. (2008). Sigma convergence versus beta convergence: Evidence from US county-level data. *Journal of Money, Credit and Banking*, 40(5):1083–1093.

Zerbe, G. O. (1978). On Fieller's theorem and the general linear model. *The American Statistician*, 32(3):103–105.

Appendix

A Figures

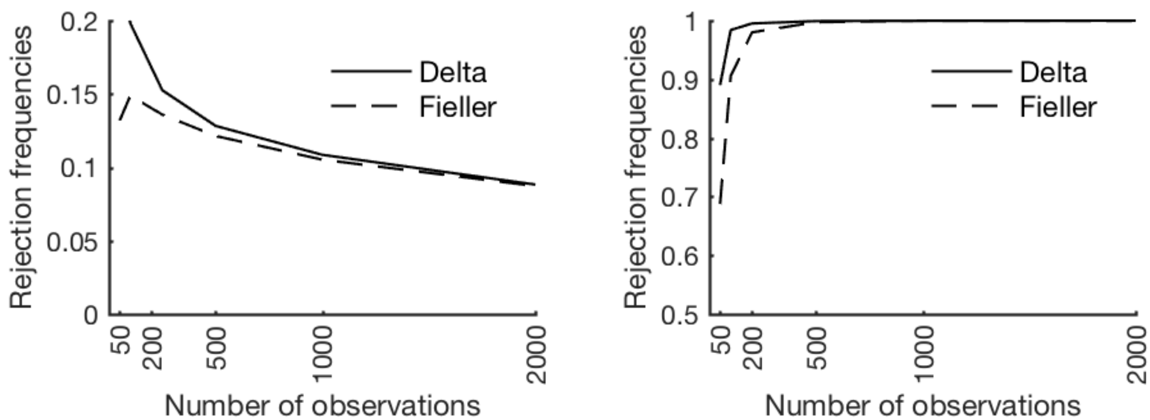
A.1 One-sample problem



GE_1: Size

GE_1: Power

Figure 1: Size and power of Delta(-method) and Fieller-type tests for GE_1 (Theil) index
H_0: $GE_1(X) = 0.4929$
Left panel DGP: $SM_X(a_X = 1.1, q_X = 4.327273)$. $GE_1(X) = 0.4929$
Right panel DGP: $SM_X(a_X = 1.7, q_X = 2.8)$. $GE_1(X) = 0.27137$



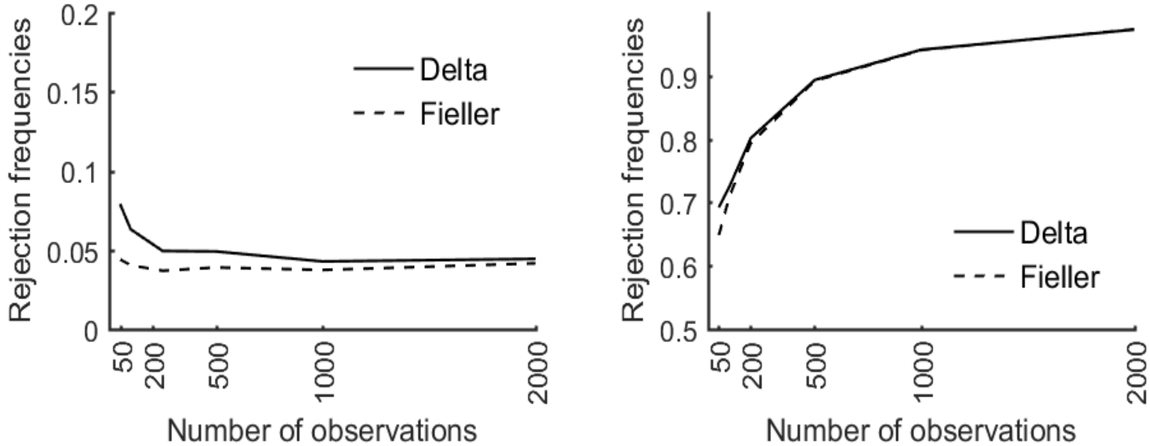
GE_2: Size

GE_2: Power

Figure 2: Size and power of Delta(-method) and Fieller-type tests for GE_2 index
H_0: $GE_2(X) = 0.71578$
Left panel DGP: $SM_X(a_X = 1.1, q_X = 4.327273)$. $GE_2(X) = 0.71578$
Right panel DGP: $SM_X(a_X = 1.7, q_X = 2.8)$. $GE_2(X) = 0.33503$

A.2 Two-sample problem

A.2.1 Experiment I: Design (I-a) – Independent samples: $n = m$, $F_X = F_Y$, $\Delta_0 = 0$



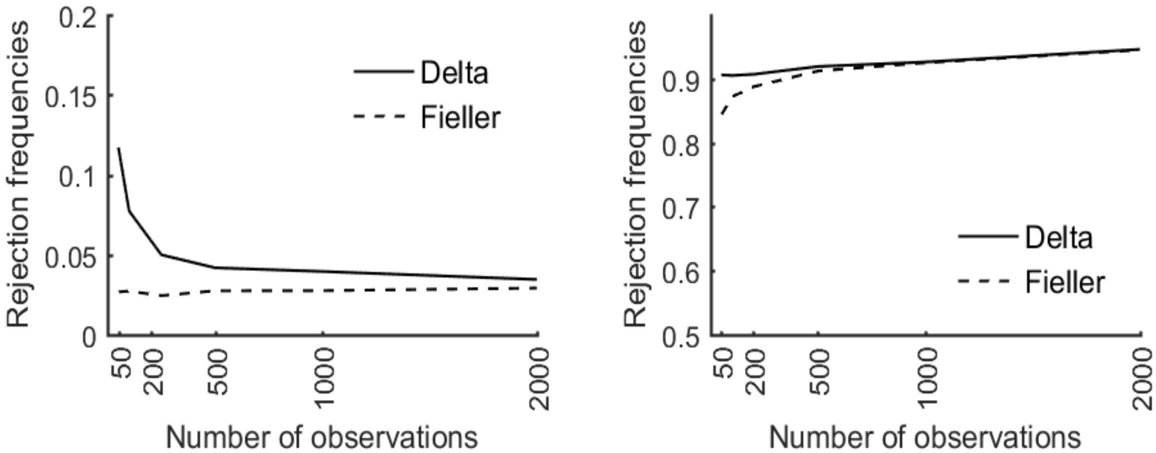
GE_1: Size GE_1: Power

Figure 3: Design (I-a) – Size and power of Delta and Fieller-type tests for GE_1 comparisons (used to derive confidence sets)

H_0: $GE_1(X) = GE_1(Y)$

Left panel: $SM_X(a_X = 5.8, q_X = 0.499616)$, $SM_Y(a_Y = 5.8, q_Y = 0.499616)$. $GE_1(X) = GE_1(Y) = 0.14011$

Right panel: $SM_X(a_X = 4.8, q_X = 0.499616)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$. $GE_1(X) = 0.22857$, $GE_1(Y) = 0.09514$



GE_2: Size GE_2: Power

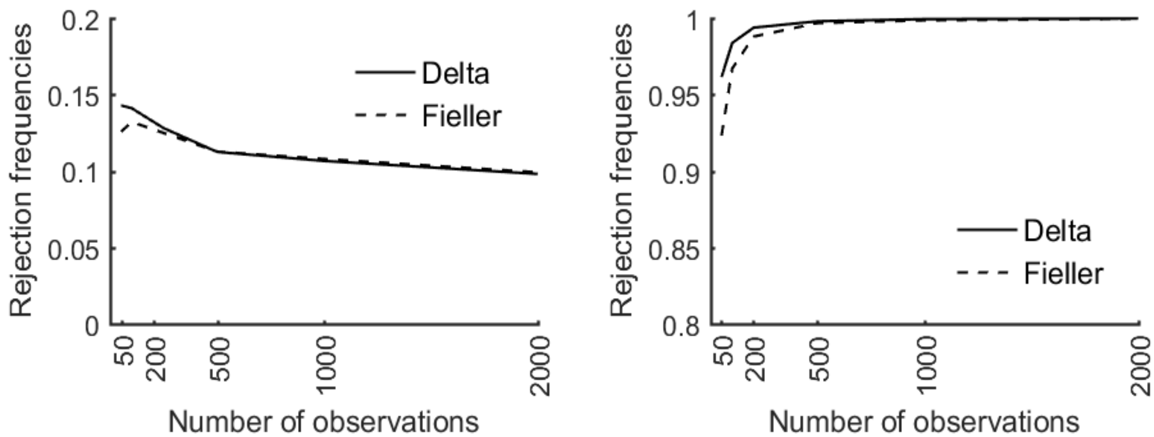
Figure 4: Design I(a) – Size and power of Delta and Fieller-type tests for GE_2 comparisons (used to derive confidence sets)

H_0: $GE_2(X) = GE_2(Y)$

Left panel: $SM_X(a_X = 5.8, q_X = 0.499616)$, $SM_Y(a_Y = 5.8, q_Y = 0.499616)$. $GE_2(X) = GE_2(Y) = 0.24396$

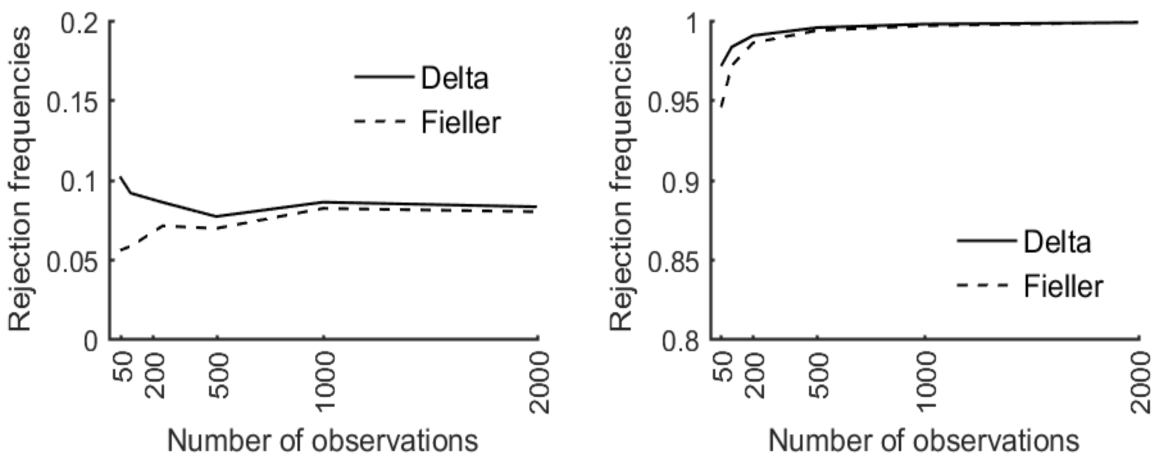
Right panel: $SM_X(a_X = 4.8, q_X = 0.499616)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$. $GE_2(X) = 0.63705$, $GE_2(Y) = 0.13806$

A.2.2 Experiment I: Design (I-b) – Independent samples: $n = m$, $F_X \neq F_Y$, $\Delta_0 = 0$



GE_1: Size GE_1: Power

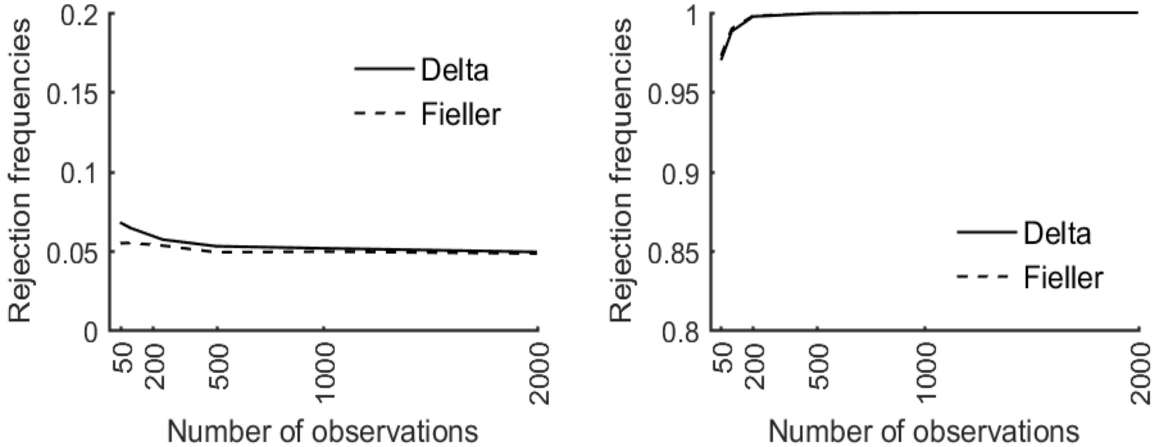
Figure 5: Design (I-b) – Size and power of Delta and Fieller-type tests for GE_1 comparisons
H_0: $GE_1(X) = GE_1(Y)$
Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 5.8, q_Y = 0.499616)$. $GE_1(X) = GE_1(Y) = 0.14011$
Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$. $GE_1(X) = 0.33830$, $GE_1(Y) = 0.09514$



GE_2: Size GE_2: Power

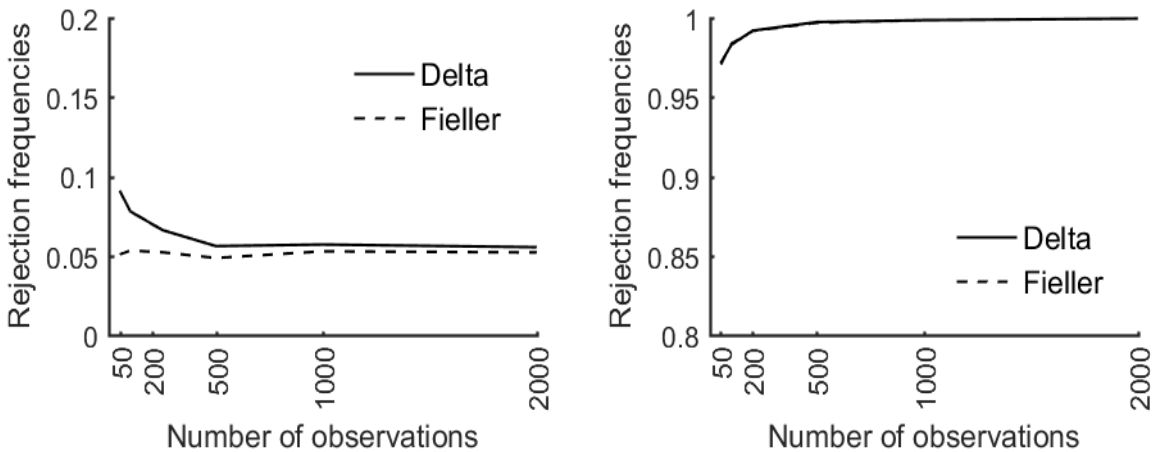
Figure 6: Design I(b) – Size and power of Delta and Fieller-type tests for GE_2 comparisons
H_0: $GE_2(X) = GE_2(Y)$
Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 3.8, q_Y = 0.9831)$. $GE_2(X) = GE_2(Y) = 0.16204$
Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4.8, q_Y = 0.9831)$. $GE_2(X) = 0.5479$, $GE_2(Y) = 0.08835$

A.2.3 Experiment I: Design (I-c) – Independent samples: $n = m$, $F_X \neq F_Y$, $\Delta_0 \neq 0$



GE_1: Size $\quad$ GE_1: Power

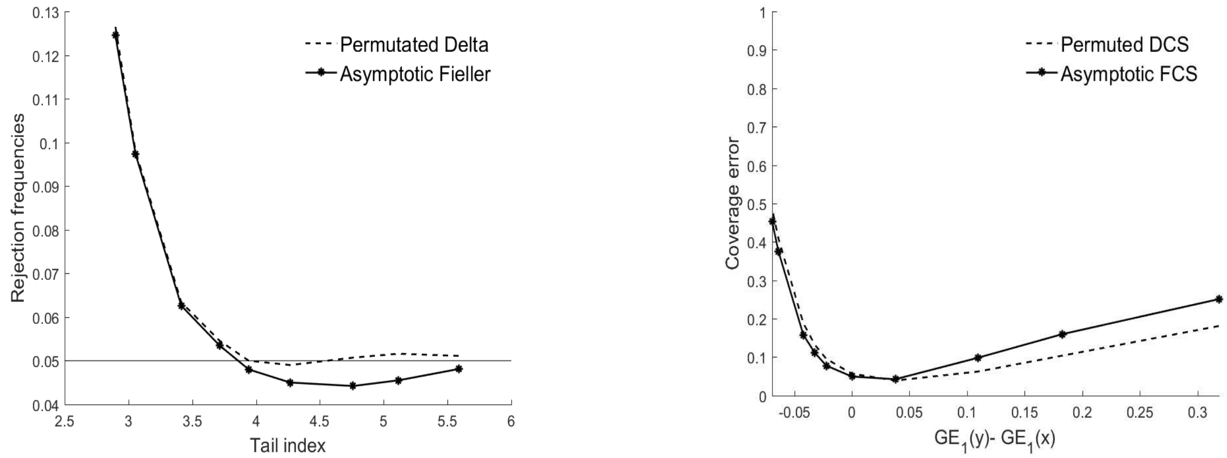
Figure 7: Design (I-c) – Size and power of Delta and Fieller-type tests for GE_1 comparisons
H_0: $GE_1(X) - GE_1(Y) = 0.04670$
Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 3.8, q_Y = 1.3061)$. $GE_1(X) = 0.14011$, $GE_1(Y) = 0.09340$
Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4.8, q_Y = 1.3061)$. $GE_1(X) = 0.33829$, $GE_1(Y) = 0.05839$



GE_2: Size $\quad$ GE_2: Power

Figure 8: Design (I-c) – Size and power of Delta and Fieller-type tests for GE_2 comparisons
H_0: $GE_2(X) - GE_2(Y) = 0.05401$
Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 3.8, q_Y = 1.2855)$. $GE_2(X) = 0.16203$, $GE_2(Y) = 0.10802$
Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4.8, q_Y = 1.2855)$. $GE_2(X) = 0.54790$, $GE_2(Y) = 0.06367$

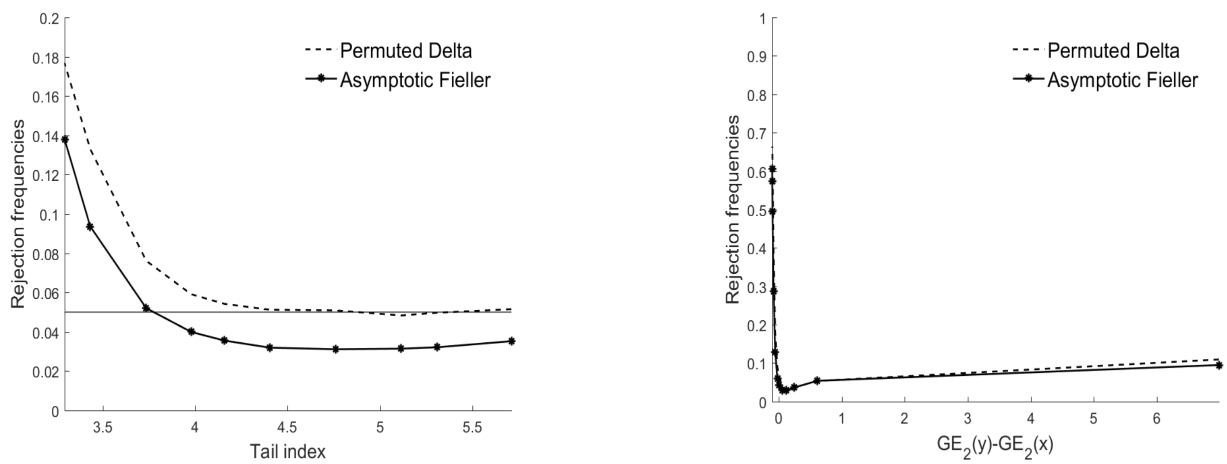
A.2.4 Comparing Fieller's method and the permutation method



GE_1: Size GE_1: Power

Figure 9: Size and Power of two-sample tests
Rejection frequencies of asymptotic Fieller-type and permuted delta methods

Note – Samples are independent and $n = m$. $F_X = F_Y$ and $GE_1(X) = GE_1(Y)$. The left panel pertains to the size analysis and it plots the Rejection frequencies of asymptotic the Fieller-types and Permuted Delta method against the tail index: $\xi = [2.897, 6.256]$. Power analysis is presented in the right panel where rejection frequencies are plotted against the difference between the two indices: $GE_1(Y) - GE_1(X)$. For power, we set $q_Y = 10$.



GE_2: Size GE_2: Power

Figure 10: Size and Power of two-sample tests
Rejection frequencies of asymptotic Fieller-type and permuted delta method

Note – Samples are independent and $n = m$. $F_X = F_Y$ and $GE_\gamma(X) = GE_\gamma(Y)$. The left panel pertains to the size analysis and it plots the Rejection frequencies of asymptotic Fieller-type and Permuted delta methods against the tail index: $\xi = [2.897, 6.256]$. Power analysis is presented in the right panel where rejection frequencies are plotted against the difference between the two indices: $GE_2(Y) - GE_2(X)$. For power, we set $q2 = 10$.

A.2.5 Behavior with respect to the sensitivity parameter γ

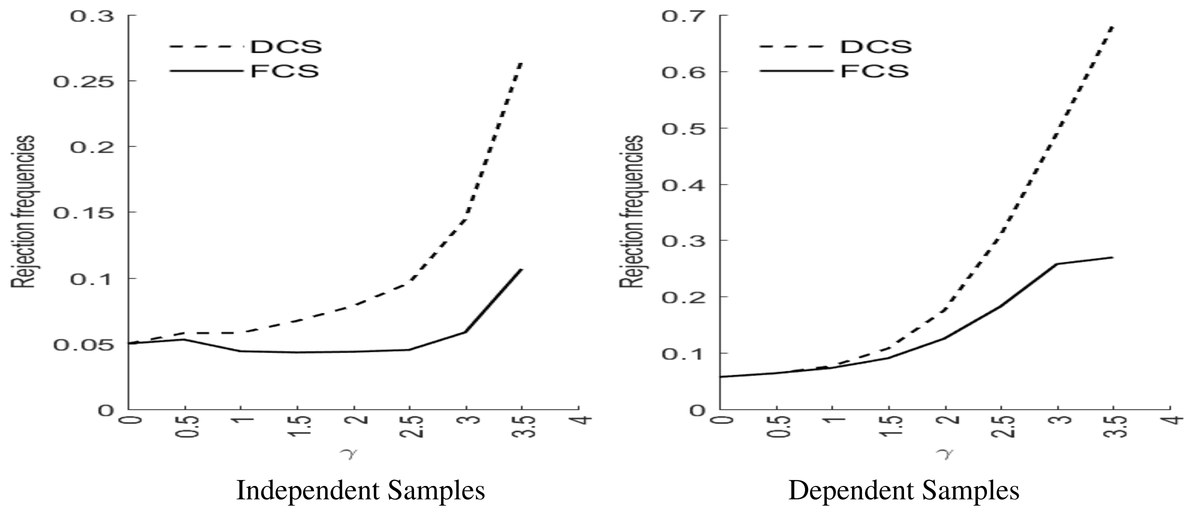


Figure 11: Rejection frequencies of the tests inverted to derive the Delta method and Fieller's confidence sets over the sensitivity parameter γ

Note – The distributions under the null hypothesis are identical and defined by: $SM_X(a_X = 2.8, q_X = 1.7)$ and $SM_Y(a_Y = 2.8, q_Y = 2)$. $n = m = 50$

A.2.6 Robustness to the shape of the null distributions

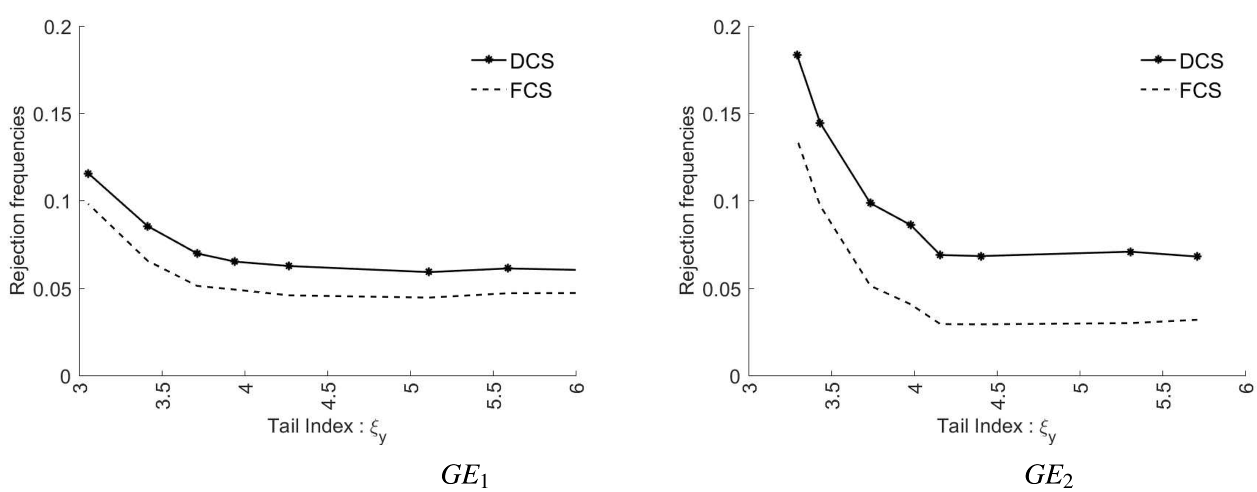


Figure 12: Rejection frequencies of the tests inverted to derive the Delta method and Fieller's confidence sets over the tail index ξ_y

Note – In the left panel, we consider the Theil index where ξ_X is fixed at 4.76 and $\xi_Y = [3.055, 6.255]$. In the right panel, we consider GE_2 with ξ_X is fixed at 4.76 and $\xi_Y = [3.293, 5.7107]$. $n = m = 50$

B Tables

Table 1: Rejection frequencies of Delta and Fieller methods: effect of left-tail thickness

a	q	$\xi = aq$	GE_1	GE_2	PDL - GE_1	PDL - GE_2
3.173333	1.5	4.76	0.11991	0.13808	**1.53**	**5.91**
2.8	1.7	4.76	0.14012	0.16204	**4.08**	**8.67**
2.38	2	4.76	0.17279	0.20215	**4.10**	**11.45**
2	2.38	4.76	0.21774	0.26039	**6.23**	**15.65**
1.5	3.173333	4.76	0.32206	0.4112	**10.47**	**21.87**
1.1	4.327273	4.76	0.4929	0.71578	**13.45**	**30.91**
2.5	1.456	3.64	0.19888	0.26379	**2.95**	**4.48**
2.3	1.582609	3.64	0.21891	0.29233	**2.83**	**5.83**
2.1	1.733333	3.64	0.24375	0.32895	**3.23**	**7.34**
1.9	1.915789	3.64	0.27516	0.37727	**5.30**	**9.47**
1.7	2.141176	3.64	0.31577	0.44332	**6.30**	**11.03**
1.5	2.426667	3.64	0.36978	0.53789	**8.11**	**14.14**
1.3	2.8	3.64	0.44414	0.682	**9.15**	**17.33**

Note- PDL stands for the percentage difference of the levels of the Delta and the Fieller-type method. The results in this table pertain to the percentage difference of the DCS and FCS levels as the left tail of the underlying distribution gets thicker. The right tail is fixed ($\xi_X = 4.76$). Column 6 reports the percentage difference associated with the null hypothesis $H01$: $GE_1 = 0$ and column 7 reports the percentage difference associated with the null hypothesis $H02$: $GE_2 = 0$.

Table 2: Rejection frequencies of Delta and Fieller methods:
effect of right-tail thickness in the two sample problem; n=50.

a_X	q_X	q_Y	a_Y	$\xi_X = \xi_Y$	$GE_1(X) = GE_1(Y)$	$GE_2(X) = GE_2(Y)$	PDL - GE_1	PDL - GE_2
5	2.1	5	2.1	10.5	0.04075	0.04096	2.84	10.58
5	1.9	5	1.9	9.5	0.04268	0.04326	4.47	13.99
5	1.7	5	1.7	8.5	0.04524	0.04639	5.19	20.59
5	1.5	5	1.5	7.5	0.0488	0.05084	8.81	24.2
5	1.3	5	1.3	6.5	0.05401	0.05763	13.60	31.96
5	1.1	5	1.1	5.5	0.06230	0.06906	16.88	42.72
5	0.9	5	0.9	4.5	0.07708	0.09155	29.70	56.74
5	0.7	5	0.7	3.5	0.10877	0.15046	36.87	66.55
5	0.5	5	0.5	2.5	0.20464	0.49151	53.75	84.86

Note – PDL stands for the percentage difference of the levels of the Delta and the Fieller-type method. The results in this table pertain to the percentage difference of the DCS and FCS levels as the right tails of both distributions gets thicker. The left tails of both distributions are fixed (a_X and a_Y are fixed) and the right tails gets thicker (with smaller ξ_X and ξ_Y). Column 8 reports the percentage difference associated with the null hypothesis H_{01}: $GE_1(X) = GE_1(Y)$ and column 9 reports the percentage difference associated with the null hypothesis H_{02}: $GE_2(X) = GE_2(Y)$.

Table 3: Rejection frequencies of Delta and Fieller methods: effect of left-tail thickness in the two sample problem; n=50.

a_X	q_X	a_Y	q_Y	$\xi_X = \xi_Y$	$GE_1(X)$	$GE_2(X)$	$GE_1(Y)$	$GE_2(Y)$	$\Delta_{0,1}$	$\Delta_{0,2}$	PDL - GE_1	PDL - GE_2
2.8	1.7	5.8	0.821	4.76	0.14012	0.16204	0.0628	0.07347	0.07732	0.08857	**18.74**	**38.6**
2.8	1.7	5.2	0.915	4.76	0.14012	0.16204	0.06957	0.08095	0.07055	0.08109	**20.80**	**39.95**
2.8	1.7	4.8	0.992	4.76	0.14012	0.16204	0.07524	0.0872	0.06488	0.07484	**19.78**	**41.72**
2.8	1.7	4.2	1.133	4.76	0.14012	0.16204	0.08666	0.09998	0.05346	0.06206	**21.35**	**45.9**
2.8	1.7	3.8	1.253	4.76	0.14012	0.16204	0.09685	0.11148	0.04327	0.05056	**23.41**	**48.41**
2.8	1.7	3.2	1.488	4.76	0.14012	0.16204	0.11866	0.13661	0.02146	0.02543	**23.82**	**52.19**
2.8	1.7	3	1.587	4.76	0.14012	0.16204	0.12848	0.14816	0.01164	0.01388	**26.22**	**56.03**
2.8	1.7	2.6	1.831	4.76	0.14012	0.16204	0.15401	0.17888	-0.01389	-0.01684	**25.37**	**56.94**
2.8	1.7	2.4	1.983	4.76	0.14012	0.16204	0.17092	0.19982	-0.0308	-0.03778	**27.01**	**58.79**
2.8	1.3	1.3	2.8	3.64	0.17535	0.23133	0.44414	0.682	-0.26879	-0.45067	**30.08**	**49.28**
2.8	1.3	1.5	2.426667	3.64	0.17535	0.23133	0.36978	0.53789	-0.19443	-0.30656	**32.65**	**52.58**
2.8	1.3	1.7	2.141176	3.64	0.17535	0.23133	0.31577	0.44332	-0.14042	-0.21199	**32.77**	**54.64**
2.8	1.3	1.9	1.915789	3.64	0.17535	0.23133	0.27516	0.37727	-0.09981	-0.14594	**32.98**	**59.81**
2.8	1.3	2.1	1.733333	3.64	0.17535	0.23133	0.24375	0.32895	-0.0684	-0.09762	**37.37**	**62.30**
2.8	1.3	2.3	1.582609	3.64	0.17535	0.23133	0.21891	0.29233	-0.04356	-0.061	**40.25**	**66.83**
2.8	1.3	2.5	1.456	3.64	0.17535	0.23133	0.19888	0.26379	-0.02353	-0.03246	**41.10**	**71.34**

Note- PDL stands for the percentage difference of the levels of the Delta and the Fieller's method. The results in this table pertain to the percentage difference of the DCS and FCS levels as the left tails of both distributions gets thicker. The right tails of both distributions are fixed ($\xi_X = \xi_Y = 4.76$)while the left tail of the second distribution gets thicker (with smaller a_Y). Column 12 reports the percentage difference associated with the null hypothesis $H01$: $GE_1(X) - GE_1(Y) = \Delta_{0,1}$ and column 13 reports the percentage difference associated with the null hypothesis $H02$: $GE_1(X) - GE_1(Y) = \Delta_{0,2}$. The values of $\Delta_{0,1}$ and $\Delta_{0,1}$ are given in columns 10 and 11 respectively.

Table 4: Rejection probabilities and widths of confidence sets
based on the Delta and Fieller-type methods: One-sample problem

n	Rejection Delta	Rejection Fieller	Bounded	Union of two disjoint sets	Unbounded	Width Fieller	Width Delta
50	0.3758	0.2616	9841	105	54	1.4339	0.6316
100	0.3211	0.2773	9983	16	1	0.7616	0.6026
200	0.2707	0.258	9998	2	0	0.6324	0.5462
500	0.2219	0.2244	10000	0	0	0.4482	0.4325
1000	0.1764	0.1796	10000	0	0	0.3635	0.3575
2000	0.1626	0.167	10000	0	0	0.2746	0.2726
10000	0.1077	0.1095	10000	0	0	0.1474	0.1472
20000	0.0990	0.1006	10000	0	0	0.1098	0.1097
100000	0.0753	0.0756	10000	0	0	0.0544	0.0544
200000	0.0686	0.0698	10000	0	0	0.0395	0.0395

Note – The coverage rate of the confidence set is equal to 1−(Rejection probability). The results in this table pertains to the GE_2 index with $SM_X(a_X = 1.1,\ q_X = 4.327273)$. H_0: $GE_2 = 0.71577$

Table 5: Rejection probabilities and widths of confidence sets
based on the Delta and Fieller-type methods: Two-sample problem

n	Rejection Delta	Rejection Fieller	Bounded	Union of two disjoint sets	Unbounded	Width Fieller	Width Delta
50	0.1843	0.1161	9955	35	10	0.1031	0.0655
100	0.1666	0.1293	9997	3	0	0.0642	0.0548
200	0.1468	0.1297	9999	1	0	0.0461	0.0436
500	0.1316	0.125	10000	0	0	0.032	0.0313
1000	0.1187	0.1168	10000	0	0	0.0239	0.0237
2000	0.1049	0.1047	10000	0	0	0.0179	0.0179
10000	0.0790	0.0787	10000	0	0	0.0090	0.0090
20000	0.0761	0.0766	10000	0			0.0066
100000	0.0663	0.0663	10000	0	0	0.0032	0.0032
200000	0.0616	0.0617	10000	0	0	0.0023	0.0023

Note – The coverage rate of the confidence set is equal to 1−(Rejection probability). The results in this table pertains to GE_2 index with $SM_X(a_X = 2.8, q_X = 1.7)$ and $SM_Y(a_Y = 3.8, q_Y = 1.2855)$. H_0: $GE_2(X) - GE_2(Y) = 0.05401$.

Table 6: Estimates and confidence intervals of the change in inequality across U.S. states between 1946 and 2016.

	Theil Index / GE_1	GE_2
First sample - 1946	0.02743	0.02679
Second sample -2016	0.0144	0.01516
$GE_\gamma(2016) - GE_\gamma(1946)$	-0.01303	-0.01163
Delta C.I.	$[-0.02486, -0.001204]$ Inequality decreases	$[-0.02349, 0.00024]$ No change in Inequality
Fieller's C.I.	$[-0.02531, -0.00155]$ Inequality decreases	$[-0.02456, -0.00043]$ Inequality decreases
Permutation test $\mathrm{p-Value}$	0.014 Inequality decreases	0.014 Inequality decreases
Number of states	48	48

Table 7: Estimates and confidence intervals of the change in inequality across non-OECD countries

	Theil Index / GE_1	GE_2
First sample - 1960	0.717621	1.46631
Second sample -2013	0.78726	1.45076
$GE_\gamma(2013) - GE_\gamma(1960)$	0.06964	-0.01554
Delta C.I.	$[-0.35694, 0.49623]$	$[-1.15143, 1.120337]$
Fieller's C.I.	$[-0.40436, 0.63075]$	$\mathbb{R}$
Permutation test $\mathrm{p-value}$	0.886	0.992
Number of countries	72	72

Récents cahiers de recherche du CIREQ
Recent Working Papers of CIREQ

La liste complète des cahiers de recherche du CIREQ est disponible à l'adresse suivante : http://www.cireqmontreal.com/cahiers-de-recherche .

The complete list of CIREQ's working papers is available at the following address : http://www.cireqmontreal.com/cahiers-de-recherche .

02-2019 Mueller, H., C. Rauh, "The Hard Problem of Prediction for Conflict", avril 2019, 46 pages

03-2019 Galindo da Fonseca, J.A., P. Pannella, "The Housing Boom and Selection into Entrepreneurship", mai 2019, 30 pages

04-2019 Galindo da Fonseca, J.A., "Unemployment, Entrepreneurship and Firm Outcomes", mai 2019, 96 pages

05-2019 Galindo da Fonseca, J.A., C. Berubé, "Spouses and Entrepreneurship", mai 2019, 45 pages

06-2019 Galindo da Fonseca, J.A., I. Snoddy, "Entrepreneurship, Outside Options and Constrained Efficiency", mai 2019, 22 pages

07-2019 Benchekroun, H., A.R. Chaudhuri, D. Tanseem, "On the Impact of Trade in a Common Property Renewable Resource Oligopoly", mai 2019, 28 pages

08-2019 Duffy, J., J.H. Jiang, H. Xie, "Experimental Asset Markets with an Indefinite Horizon", juillet 2019, 58 pages

09-2019 Rauh, C., "Measuring Uncertainty at the Regional Level Using Newspaper Text", août 2019, 28 pages

10-2019 Aguirregabiria, V., M. Marcoux, "Imposing Equilibrium Restrictions in the Estimation of Dynamic Discrete Games", septembre 2019, 45 pages

11-2019 Pelli, M., J. Tschopp, N. Bezmaternykh, K.M. Eklou, "In the Eye of the Storm : Firms and Capital Destruction in India", octobre 2019, 43 pages

12-2019 Amodio, F., G. Chiovelli, S. Hohmann, "The Employment Effects of Ethnic Politics", décembre 2019, 65 pages

13-2019 Amodio, F., Martinez-Carrasco, M.A., "Inputs, Incentives, and Self-selection at the Workplace", décembre 2019, 55 pages

14-2019 Degan, A., M. Li, H. Xie, "Persuasion Bias in Science : An Experiment on Strategic Sample Selection", novembre 2019, 47 pages

01-2020 Gupta, R., M. Pelli, "Electrification and Cooking Fuel Choice in Rural India", janvier 2020, 34 pages

02-2020 Doĝan, B., L. Ehlers, "Robust Minimal Instability of the Top Trading Cycles Mechanism", mars 2020, 21 pages