

Three Essays on Statistical Inference on Inequality Measures

by

Abdallah Zalgout

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial
fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Economics

Carleton University

Ottawa, Ontario

©2019

Abdallah Zalgout

Abstract

Providing reliable inference on inequality measures is an enduring challenge, mainly due to the complications arising from nonlinearities in their definitions and from the complex nature of the underlying distributions which are typically characterized by extremely heavy tails. The thesis is concerned with proposing non-standard asymptotic and simulation-based inference procedures for moment-based inequality measures (general entropy family of inequality indices) and quantile-based measures (quantile ratio index). Inference on both types of measures is prone to heavy-tailed distributions complications and to the ratio-induced identification issues. In addition to that, moment-based measures are subject to the so-called Bahadur-Savage impossibility problem while quantile-based measures are not. On the other hand, the main difficulty with inference on quantile-based measures is the dependence of the quantile variance on the underlying density function which involves kernel estimation and bandwidth selection.

The first chapter of my thesis introduces a Fieller-type method for the Theil Index and assess its finite-sample properties by a Monte Carlo simulation study. The fact that almost all inequality indices can be written as a ratio of functions of moments

and that a Fieller-type method does not suffer from weak identification as the denominator approaches zero, makes it an appealing alternative to the available inference methods. The simulation results exhibit several cases where a Fieller-type method improves coverage. This occurs in particular when the Data Generating Process (DGP) follows a finite mixture of distributions, which reflects irregularities arising from low observations (close to zero) as opposed to large (right-tail) observations. Designs that forgo the interconnected effects of both boundaries provide possibly misleading finite-sample evidence.

The second chapter proposes confidence set procedures for inequality indices (the one-sample problem) and for differences of indices (the two-sample problem), that do not require identifying these measures nor their differences. The paper documents the fragility of decisions (at usual levels) relying on traditional interpretations of - significant or insignificant - comparisons when the difference under test can be weakly identified. Proposed robust procedures are analytical and require basically the same inputs as their standard counterparts. In particular, the chapter introduces Fieller-type confidence sets for the Generalized Entropy family. Extensive simulations allowing for possibly dependent samples demonstrate the superiority of the proposed methods relative to the standard Delta method and when relevant, to the permutation test method.

Simulation results show the following. (1) Inference problems stem from irregularities arising from both tails of the underlying distribution, and not - as has been long believed in this literature - from the right tail only. (2) Proposed methods outperform the Delta method across the board. (3) Size gains are most prominent with indices

that place more weight on the right tail of the underlying distribution. (4) Proposed tests for differences of indices outperform the permutation test when the distributions under the null are different. The latter cannot be inverted, a disadvantage we aim to strongly underscore. To further demonstrate that our theoretical results are empirically relevant, an empirical study is conducted.

As an illustrative discussion of outcomes that challenge conventional judgements, the economic convergence hypothesis across the U.S. states and across non-OECD countries is revisited. With reference to the growth literature which typically uses the variance of log per-capita income to measure dispersion, one of the contributions of the paper is endorsing, instead, the Generalized Entropy family. Unlike the former, Generalized Entropy satisfies the Pigou-Dalton principle. The results confirm the importance of accounting for micro-founded axioms and cast doubt on traditional no-change test decisions, which sheds new light on enduring controversies surrounding convergence.

Unlike the first two chapters, the third chapter focuses on inference on quantile-based measures, mainly the quantile ratio index. Three inference methods were proposed: the standard Fieller method and Bootstrap-based alternative Wald and Fieller studentized procedures. In particular, the latter two circumvent complications arising from the dependence of the quantile variance on the underlying density function. Simulation results show that the standard Wald-type confidence sets as well as their Fieller-based counterparts have levels that deviate arbitrarily from their nominal ones, as well as very low power. In contrast, the proposed studentization which relies on bootstrapping the ratio directly restores coverage and improves power remarkably

even with samples of small sizes drawn from extremely heavy-tailed distributions. This suggests that robustness in conjunction with scale-invariance jointly justify the success of our proposed methodology.

Acknowledgements

First and foremost, I would like to thank my supervisor Lynda Khalaf for her patience, trust, motivation, immense knowledge, and invaluable support; a better supervisor and mentor for my doctoral studies would be hard to find. I have benefited tremendously from Lynda's wisdom and guidance and I consider it a privilege to have worked under her supervision.

I would also like to thank Jean-Marie Dufour (McGill University) and Emmanuel Flachaire (Aix-Marseille University) for their guidance with the first two chapters. Working with them has helped me sharpen my research skills and enrich my knowledge of econometrics and income inequality.

Sincere thanks also go to my dissertation committee: Paul Makdissi, Marcel Voia, Matthew D. Webb, and Myra Yazbeck for their guidance, expertise, enthusiastic encouragement, and constructive feedback. I would also like to thank my fellow Ph.D. peers and friends, and the faculty and staff of Carleton University for their assistance and cooperation.

Last but not least I would like to thank my family and friends who showered me with consistent unconditional love and support.

Table of Contents

Abstract	ii
Acknowledgements	vi
1 Confidence sets for inequality measures: Fieller-type methods	1
1.1 Introduction	1
1.2 Fieller-type inference for inequality measures	5
1.2.1 General functional ratios	6
1.2.2 Fieller-type inference for the Theil Index	9
1.3 Simulation results	11
1.4 Conclusion	18
2 Identification-robust inequality analysis	19
2.1 Introduction	19
2.2 Fieller-type confidence sets for Generalized Entropy inequality measures	27
2.2.1 One-sample problem	30
2.2.2 Two-sample problem	34
2.3 Simulation evidence	40
2.3.1 Simulation results: one-sample problem	42
2.3.2 Simulation results: two-sample problem	45
2.4 Application: Regional income convergence	56
2.5 Conclusion	62
3 Reliable non-parametric inference for quantile ratios	95
3.1 Introduction	95
3.2 Methodology	100
3.2.1 Method 1 - Wald-type Confidence Sets (WSC)	102
3.2.2 Method 2- Density-free Bootstrap Wald Confidence Sets (WCS-B)	105

3.2.3	Method 3 - Fieller's Confidence Sets (FCS)	106
3.2.4	Method 4 - Density-free Bootstrap Fieller-type method FTT-B	108
3.3	Simulation Results	110
3.4	Conclusion	116
References		124

List of Tables

2.1	Rejection frequencies of Delta and Fieller methods: effect of left-tail thickness	88
2.2	Rejection frequencies of Delta and Fieller methods: effect of right-tail thickness in the two sample problem; $n = 50$	89
2.3	Rejection frequencies of Delta and Fieller methods: effect of left-tail thickness in the two sample problem; $n = 50$	90
2.4	Rejection probabilities and widths of confidence sets based on the Delta and Fieller-type methods: One-sample problem	91
2.5	Rejection probabilities and widths of confidence sets based on the Delta and Fieller-type methods: Two-sample problem	92
2.6	Estimates and confidence intervals of the change in inequality across U.S. states between 1946 and 2016.	93
2.7	Estimates and confidence intervals of the change in inequality across non-OECD countries	94
3.1	Coverage error of the proposed inference methods for the 99/50 ratio	118
3.2	Coverage error of the proposed inference methods for the 80/20 ratio	119
3.3	Coverage error of the proposed inference methods for the 99/20 ratio	120
3.4	Empirical Power of the covered inference methods for the 99/20 ratio	121

List of Figures

1.1	Rejection probabilities for Delta and Fieller methods	13
1.2	Rejection probabilities for Delta and Fieller methods	14
1.3	Rejection probabilities for Delta and Fieller methods	16
1.4	Rejection probabilities for Delta and Fieller methods	17
2.1	Size and power of Delta(-method) and Fieller-type tests for GE_1 (Theil) index	64
2.2	Size and power of Delta(-method) and Fieller-type tests for GE_2 index	65
2.3	Design I(a) – Size and power of Delta and Fieller-type tests for GE_1 comparisons (used to derive confidence sets)	66
2.4	Design I(a) – Size and power of Delta and Fieller-type tests for GE_2 comparisons (used to derive confidence sets)	67
2.5	Design I(b) – Size and power of Delta and Fieller-type tests for GE_1 comparisons	68
2.6	Design I(b) – Size and power of Delta and Fieller-type tests for GE_2 comparisons	69
2.7	Design I(c) – Size and power of Delta and Fieller-type tests for GE_1 comparisons	70
2.8	Design I(c) – Size and power of Delta and Fieller-type tests for GE_2 comparisons	71
2.9	Design II(a) – Size and power of delta and Fieller-type tests for GE_1 comparisons	72
2.10	Design II(a) – Size and power of delta and Fieller-type tests for GE_2 comparisons	73
2.11	Design II(b) – Size and power of delta and Fieller-type tests for GE_1 comparisons	74
2.12	Design II(b) – Size and power of Delta and Fieller-type tests for GE_2 comparisons	75

2.13 Design II(c) – Size and power of Delta and Fieller-type tests for GE_1 comparisons	76
2.14 Design II(c) – Size and power of Delta and Fieller-type tests for GE_2 comparisons	77
2.15 Design III(a) – Size and power of delta and Fieller-type tests for GE_1 comparisons	78
2.16 Design III(a) – Size and power of delta and Fieller-type tests for GE_2 comparisons	79
2.17 Design III(b) – Size and power of delta and Fieller-type tests for GE_1 comparisons	80
2.18 Design III(b) – Size and power of delta and Fieller-type tests for GE_2 comparisons	81
2.19 Design III(c) – Size and power of delta and Fieller-type tests for GE_1 comparisons	82
2.20 Design III(c) – Size and power of delta and Fieller-type tests for GE_2 comparisons	83
2.21 Size and Power of two-sample tests	84
2.22 Size and Power of two-sample tests	85
2.23 Rejection frequencies of the tests inverted to derive the Delta method and Fieller’s confidence sets over the sensitivity parameter γ	86
2.24 Rejection frequencies of the tests inverted to derive the Delta method and Fieller’s confidence sets over the tail index ξ_y	87

Chapter 1

Confidence sets for inequality

measures: Fieller-type methods

1.1 Introduction

Asymptotic inference methods for inequality indices are for the most part unreliable due to the complex empirical features of the underlying distributions, particularly in the case of income. Typically, the presence of heavy tails invalidates standard parametric and nonparametric inference methods based on central limit theory (CLT), leading to spurious conclusions with samples of realistic size. This problem persists even with very large samples. Moreover, for some parameter values, the moments of widely used distributions in this literature, such as the Singh-Maddala and Pareto distributions, do not exist. Early references can be traced back to [Maasoumi \(1997\)](#) or [Mills and Zandvakili \(1997\)](#); for a survey, see [Cowell and Flachaire \(2015\)](#).

Bootstrap inference methods emerge as an appealing alternative, since observa-

tions can often be viewed as independent random draws from the population. The first study to use and recommend bootstrap methods for inequality indices is the one of [Mills and Zandvakili \(1997\)](#). [Biewen \(2002\)](#) studied the performance of standard bootstrap methods in the context of inequality measures assuming a lognormal distribution as the Data Generating Process (DGP). Although his results suggest that the bootstrap performs well in finite samples, the lognormal distribution he used does not capture the thick tails typically observed in empirical work ([Davidson and Flachaire, 2007](#)). Other simulation studies based on heavy-tailed distributions, such as the Singh-Maddala distribution, confirm that bootstrapping fails – often by far – to control coverage rates, despite the fact that they lead to higher-order refinements relative to asymptotic methods ([Davidson and Flachaire, 2007](#); [Cowell and Flachaire, 2007](#)).

Non-standard inference methods have recently been suggested in an attempt to improve the quality of inference for inequality measures. Two notable approaches are permutation tests ([Dufour et al., 2019](#)) and semi-parametric methods ([Davidson and Flachaire, 2007](#); [Cowell and Flachaire, 2007](#)). The permutational approach focuses on testing the equality of two indices, and the authors show that it performs very well when the two indices come from similar distributions. The semi-parametric bootstrap approach assumes a parametric distribution for the right tail and a nonparametric empirical distribution function (EDF) for the rest. This method leads to considerable refinement over their asymptotic and bootstrap counterparts, provided the probability of the tail (p) and the ordered statistics defining the upper tail (k) are well chosen, which is usually not an easy task. Thus except for very specific cases, accurate

inference methods on inequality measures are not available.

In this chapter, we introduce the Fieller method for the Theil Index, and we assess its finite-sample properties through a Monte Carlo simulation study. Fieller's method was originally introduced for inference on the ratio of two means of normal variates. It is based on inverting a t -test of a linear restriction associated with the ratio, and allows one to get exact confidence sets for this ratio. This holds promise relative to the standard Delta method especially when the denominator of the ratio approaches zero, since the implicit linear reformulation addresses the underlying weak identification. Most inequality indices can be written as a ratio of functions of moments; so a Fieller-type method may plausibly lead to more reliable inference on these indices. However, given the non-linear dependence between the numerator and the denominator of the indices along with the typically positive support of the underlying distributions, the advantages from employing a Fieller-type method should not be taken for granted. This motivates the present work.

The method first introduced by [Fieller \(1940, 1954\)](#) was extended to independent samples of different sizes ([Bennett, 1962](#)), multivariate models ([Bennett, 1959](#)), general exponential regression models ([Cox, 1967](#)), general linear regression models ([Zerbe, 1978](#); [Dufour, 1997](#)), and dynamic models with possibly persistent covariates ([Bernard et al., 2007, 2019](#)). [Bolduc et al. \(2010\)](#) used several variants of Fieller's approach to build simultaneous confidence sets for multiple (possibly weakly identified) ratios and they showed in a simulation study that a Fieller-type method outperforms the Delta method and controls level globally. Empirically, Fieller's approach has been routinely applied in medical research and to a lesser extent in economics ([Srivastava,](#)

1986; Willan and O'Brien, 1996; Johannesson et al., 1996; Laska et al., 1997).

Fieller-type confidence sets may be perceived as counter-intuitive, because they can produce unbounded regions including the whole real line.¹ This perhaps gives reason for their unpopularity in applied work relative to Delta method-based confidence sets (DCS), despite their solid theoretical foundation. However, the geometric interpretation of Fieller's method is quite intuitive (see Von Luxburg and Franz (2004)). More to the point here, in the presence of identification problems, valid coverage requires possibly unbounded outcomes (Gleser and Hwang, 1987; Dufour, 1997), which is allowed by a Fieller-type solution as opposed to the Delta method.

Generally speaking, identification problem pertains to the situation where the parameters of interest can not be uniquely derived from the model and the data. In the context of the Theil index, it is possible that different models yield the same value of the index. This particularly true for DGPs for which the denominator in the definition of the index is arbitrarily close to zero which results in an index of an infinite value. Although weak identification problems might not invalidate asymptotic behaviour, they can render asymptotic inference unreliable as (i) estimators can be heavily biased with distributions far from the limiting normal distribution even with large samples and/or (ii) standard Wald-type tests and confidence intervals become invalid (Dufour, 1997, 2003).

An infinite value of the index and thus can as denominator gets arbitrarily close to zero, the index approaches infinity Identification Dufour and Hsiao (2008) refers

¹See Scheffé et al. (1970) for a modified version of Fieller's method that avoid the confidence set \mathbb{R} .

to our ability to recover objects of interest from available models and data.

Our simulation results provide evidence on the superiority of a Fieller-type method in terms of reducing size distortions in many useful cases. In particular, a Fieller-type method improves coverage over the Delta method when the distribution under the null allows for bunching of low observations (close to zero) in addition to a thick right tail. For such cases, the denominator of the Theil index is small relative to the numerator and inequality is high. Methodologically, our findings suggest that studies focusing only on the upper tail may misrepresent finite-sample distortions with positive support distributions. In contrast, our design does allow us to assess further irregularities arising from low observations. As illustrated by [Cowell and Victoria-Feser \(1993\)](#), both boundaries may matter for general entropy class of indices, although not necessarily for the Theil index.

The chapter is organized as follows. Section [1.2](#) presents the Fieller-type method for the Theil index. In section [1.3](#), Monte Carlo results are provided. Section [1.4](#) concludes.

1.2 Fieller-type inference for inequality measures

Most income inequality indices depend solely on the underlying distribution of income. Technically speaking, they can be typically written as a functional which maps the space of cumulative distribution functions (CDFs) of income to the positive real line.

1.2.1 General functional ratios

Denote by Y the random variable representing income, and by $F_Y(y)$ its CDF. The class of indices considered in this chapter can be written as the ratio of functions of two moments, namely the mean μ and another moment $\nu = \mathbb{E}(\phi(Y))$, where $\phi(\cdot)$ is a given function. In particular, for the Theil index $\phi(Y) = Y \log(Y)$. In general, most inequality indices can be written as

$$I = \psi(\mu; \nu) = \frac{\psi_1(\mu; \nu)}{\psi_2(\mu; \nu)}. \quad (1.2.1)$$

The index I can be estimated using sample moments:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \hat{\nu} = \frac{1}{n} \sum_{i=1}^n \phi(Y_i), \quad (1.2.2)$$

where Y_1, \dots, Y_n is a sample of observations on Y , and $\phi(Y_i)$ is a function that takes different forms for different inequality indices. If we assume that the estimator is asymptotically normal, then the asymptotic variance can be estimated by

$$V(\hat{I}) = \frac{1}{n} \begin{bmatrix} \frac{\partial \psi}{\partial \mu} & \frac{\partial \psi}{\partial \nu} \end{bmatrix} \begin{bmatrix} \hat{\sigma}_\mu^2 & \hat{\sigma}_{\mu\nu} \\ \hat{\sigma}_{\mu\nu} & \hat{\sigma}_\nu^2 \end{bmatrix} \begin{bmatrix} \frac{\partial \psi}{\partial \mu} \\ \frac{\partial \psi}{\partial \nu} \end{bmatrix} \quad (1.2.3)$$

where $\hat{V}(\hat{I}) \equiv V(\hat{I})|_{\mu=\hat{\mu};\nu=\hat{\nu}}$. Here, $\hat{\sigma}_\mu^2$, $\hat{\sigma}_\nu^2$ and $\hat{\sigma}_{\mu\nu}$ are, respectively, estimates of the variance of Y , the variance of $\phi(Y)$, and the covariance of Y and $\phi(Y)$.²

²Note that the variance of $\hat{\mu}$, the variance of $\hat{\nu}$ and covariance of the $(\hat{\mu}, \hat{\nu})$ are equal to $\hat{\sigma}_\mu^2/n$, $\hat{\sigma}_\nu^2/n$ and $\hat{\sigma}_{\mu\nu}/n$.

In this chapter, we consider the problem of building Fieller-type confidence sets (FCS) and Delta method confidence sets (DCS) for an index of the form I in (1.2.1). In general, this can be viewed as equivalent to finding the values of I_0 which are not rejected when one tests null hypotheses of the form

$$H_0(I_0) : \frac{\psi_1(\mu, \nu)}{\psi_2(\mu, \nu)} = I_0 \quad (1.2.4)$$

where I_0 is any admissible value of I . Here, this can be achieved by inverting the absolute value or the square of the relevant t -type statistic. To invert a t -test with respect to the parameter tested, we collect all the values of this parameter for which the test is not significant at a given level.

Following the Delta method, we invert the test statistic

$$t(I_0)^2 = \frac{(\hat{I} - I_0)^2}{\hat{V}(\hat{I})} \quad (1.2.5)$$

which leads to the confidence set

$$\text{DCS}(I; 1 - \alpha) = \left[\hat{I} - z_{\alpha/2}[\hat{V}(\hat{I})]^{1/2}; \hat{I} + z_{\alpha/2}[\hat{V}(\hat{I})]^{1/2} \right] \quad (1.2.6)$$

where $z_{\alpha/2}$ is the usual α critical point based on the normal distribution (*i.e.*, $\mathbb{P}[Z \geq z_{\alpha/2}] = \alpha/2$ for $Z \sim N[0, 1]$).

By contrast, the Fieller approach can be applied as follows. For each possible value

I_0 , the Fieller-type approach consists in considering the equivalent linear hypothesis

$$H_L(I_0) : \theta(I_0) = 0, \quad \text{where } \theta(I_0) = \psi_1(\mu, \nu) - I_0 \psi_2(\mu, \nu) \quad (1.2.7)$$

where the subscript L is added to differentiate the original null hypothesis from its linear reformulation. Through this exact linearization, the Fieller-type method avoids possible (weak) identification problems when the denominator $\psi_2(\mu, \nu)$ is close to zero. To construct the FCS, we consider the square of the t -statistic associated with $H_L(I_0)$ in (1.2.7):

$$t(I_0)^2 = \frac{\hat{\theta}(I_0)^2}{\hat{V}[\hat{\theta}(I_0)]} \quad (1.2.8)$$

where $\hat{V}[\hat{\theta}(I_0)]$ is an estimate of the variance of $\hat{\theta}(I_0)$. If the statistic follows asymptotically a standard normal distribution, then a confidence set with level $1 - \alpha$ for the index I can be built by noting that

$$t(I_0)^2 \leq z_{\alpha/2}^2 \quad \Leftrightarrow \quad \frac{\hat{\theta}(I_0)^2}{\hat{V}[\hat{\theta}(I_0)]} \leq z_{\alpha/2}^2 \quad \Leftrightarrow \quad \hat{\theta}(I_0)^2 - z_{\alpha/2}^2 \hat{V}[\hat{\theta}(I_0)] \leq 0. \quad (1.2.9)$$

This yields the confidence set

$$\text{FCS}(I; 1 - \alpha) = \left\{ I_0 : \hat{\theta}(I_0)^2 - z_{\alpha/2}^2 \hat{V}[\hat{\theta}(I_0)] \leq 0 \right\}. \quad (1.2.10)$$

Since $\hat{\theta}(I_0)$ is linear in I_0 , $\hat{\theta}(I_0)^2$ and $\hat{V}[\hat{\theta}(I_0)]$ are quadratic functions of I_0 :

$$\hat{\theta}(I_0)^2 = A_1 I_0^2 + B_1 I_0 + C_1, \quad V[\hat{\theta}(I_0)] = A_2 I_0^2 + B_2 I_0 + C_2, \quad (1.2.11)$$

where the coefficients (defined below) depend on the data and the Gaussian critical point. On substituting (2.2.29) into (2.2.14), we get the quadratic inequality

$$AI_0^2 + BI_0 + C \leq 0 \quad (1.2.12)$$

where

$$A = A_1 - z_{\alpha/2}^2 A_2, \quad B = B_1 - z_{\alpha/2}^2 B_2, \quad C = C_1 - z_{\alpha/2}^2 C_2. \quad (1.2.13)$$

The coefficients, $A_1, B_1, C_1, A_2, B_2, C_2$ are functions of the sample moments and their variance estimates. The FCS solve the second degree polynomial inequality in (1.2.12) for I_0 . Let $\Delta = B^2 - 4AC$, then the $(1 - \alpha)$ -level Fieller-type confidence set is characterized as follows:

1. if $\Delta > 0$ and $A > 0$, then $\text{FC}(I; 1 - \alpha) = \left[\frac{-B - \sqrt{\Delta}}{2A}, \frac{-B + \sqrt{\Delta}}{2A} \right]$,
2. if $\Delta > 0$ and $A < 0$, then $\text{FC}(I; 1 - \alpha) = \left] -\infty, \frac{-B + \sqrt{\Delta}}{2A} \right] \cup \left[\frac{-B - \sqrt{\Delta}}{2A}, +\infty \right[$,
3. if $\Delta < 0$, then $A < 0$ and $\text{FC}(I; 1 - \alpha) = \mathbb{R}$.

For more details, see [Bolduc et al. \(2010\)](#) and the references therein.

1.2.2 Fieller-type inference for the Theil Index

The Theil index belongs to the family of GE indices and can be written as a function of two moments $\mu = \mathbb{E}(Y)$ and $\nu = \mathbb{E}[Y \log(Y)]$, where μ and ν can be estimated using their sample counterparts. In this chapter, we will use the following expression

for the Theil index:

$$I_T = \frac{\nu}{\mu} - \log(\mu). \quad (1.2.14)$$

For the Theil index (I_T), the null hypothesis defined in (1.2.4) can be written as:

$$H_0(I_{T0}) : \frac{\nu}{\mu} - \log(\mu) = I_{T0}. \quad (1.2.15)$$

The variance of the estimated Theil index can be derived using the Delta method and it is defined by (1.2.3) where the expressions of the derivatives in this context are:

$$\frac{\partial \psi}{\partial \mu} = -\frac{(\nu + \mu)}{\mu^2}, \quad \frac{\partial \psi}{\partial \nu} = \frac{1}{\mu}. \quad (1.2.16)$$

The Fieller-type method for the Theil index starts by considering the equivalent linear hypothesis as shown in (1.2.7):

$$H_0(I_{T0}) : \nu - \mu \log(\mu) - \mu I_{T0} = 0, \quad (1.2.17)$$

along with the corresponding t -statistics (squared). The confidence set for I_T is then obtained by solving the quadratic inequality described by (2.2.14)-(1.2.12). For this, we derive the parameters A_1 , B_1 , C_1 , A_2 , B_2 and C_2 in equation (2.2.29) for the Theil index:

$$A_1 = \hat{\mu}^2, \quad B_1 = -2\hat{\mu} [\hat{\nu} - \hat{\mu} \log(\hat{\mu})], \quad C_1 = [\hat{\nu} - \hat{\mu} \log(\hat{\mu})]^2. \quad (1.2.18)$$

To get the variance of $\hat{\theta}(I_0)$, we apply the Delta method to $\theta(I_0)$ in (1.2.7):

$$A_2 = \hat{\sigma}_\mu^2/n, \quad B_2 = (2\hat{\sigma}_\mu^2[\log(\hat{\mu}) + 1] - 2\hat{\sigma}_{\mu\nu})/n, \quad (1.2.19)$$

$$C_2 = (\hat{\sigma}_\mu^2[\log(\hat{\mu}) + 1]^2 - 2\hat{\sigma}_{\mu\nu}[\log(\hat{\mu}) + 1] + \hat{\sigma}_\nu^2)/n, \quad (1.2.20)$$

where $\hat{\sigma}_\mu^2$, $\hat{\sigma}_\nu^2$ and $\hat{\sigma}_{\mu\nu}$ are defined in (1.2.3).

1.3 Simulation results

In this section, we provide Monte Carlo evidence on the finite-sample properties of the Fieller-type method for the Theil index. We conduct several simulation studies focusing on the behaviour of the Fieller method when the hypothesized income distribution under the null is characterized by thick tails. To this end, we simulate data sets from the Singh-Maddala distribution ($Y_i \sim \text{SM}(a, b, q)$), the Gamma distribution ($\text{Gamma}(k, \theta)$), and finite mixtures of the latter. These distributions have been used in the literature in the context of income inequality measures (Brachmann et al., 1995; McDonald, 2008; Kleiber and Kotz, 2003; Cowell and Victoria-Feser, 1993).

The CDF of the Singh-Maddala distribution can be written as

$$F(y) = 1 - \left[1 + \left(\frac{y}{b}\right)^a\right]^{-q}, \quad (1.3.1)$$

where a is a shape parameter which affects both tails, q is another shape parameter which affects only the right tail, and b is a scale parameter which has no impact on our analysis (for the indices in question are scale invariant. For this distribution, the

k -th moment exists for $-a < k < aq$.

The expectation of the Singh-Maddala distribution can be expressed as

$$\mu = \frac{qb\Gamma(a^{-1} + 1)\Gamma(q - a^{-1})}{\Gamma(q + 1)}. \quad (1.3.2)$$

In this case, a closed-form expression for $\nu = \mathbb{E}(Y \log(Y))$ is also available:

$$\nu = \mu a^{-1}[\psi(a^{-1} + 1) - \psi(q - a^{-1}) + a \log(b)] \quad (1.3.3)$$

where $\Gamma(\cdot)$ is the Gamma function and $\psi(\cdot) \equiv \Gamma'(\cdot)/\Gamma(\cdot)$ is the digamma function.

The other distribution we consider in the simulations is the Gamma distribution with density function

$$f(y) = \frac{y^{k-1}e^{-(k/\theta)}}{\theta^k\Gamma(k)}, \quad y > 0, \quad (1.3.4)$$

where k is a shape parameter and θ is a scale parameter. The expectation of this distribution (μ) is the scale multiplied by the shape parameter ($\mu = k\theta$). The value of ν for the Gamma distribution was computed by numerical methods.

The number of replications was set to $N = 10000$. For each sample, we compute the Theil inequality, and the underlying estimated variance, and the t -type statistics associated with the Delta and Fieller-type methods. Because of the duality between tests and confidence sets, the coverage rate of the confidence sets can be evaluated by computing the rejection probabilities of these tests. The coverage error rate (or equivalently the rejection probability) is computed as the proportion of times the relevant t -statistic rejects the null hypothesis. For a significance level α , we say the

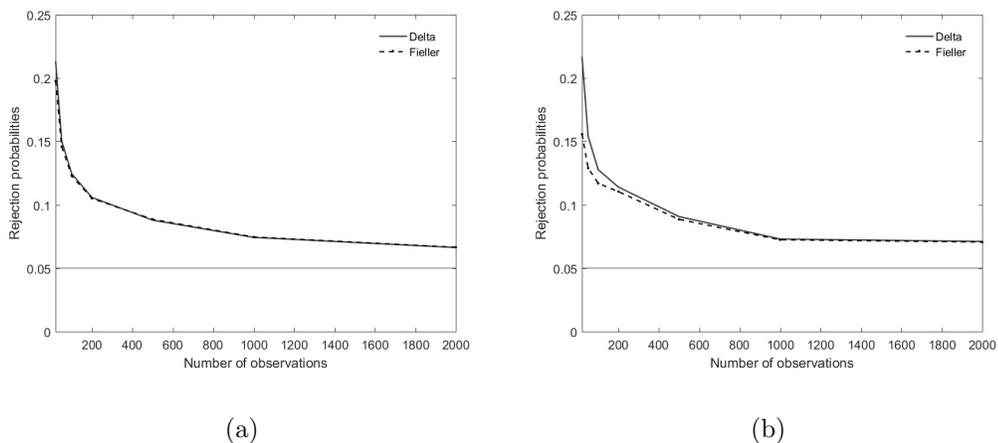


Figure 1.1: Rejection probabilities for Delta and Fieller methods

Null hypotheses: Singh-Maddala distributions

Left panel: $SM(a = 2.8, q = 1.7)$; Right panel: $SM(a = 1.1, q = 5)$

Note - The Delta method and Fieller method statistics are defined by (1.2.5) and (1.2.8) respectively. The null hypothesis tested is $H(I_{T0}) : I = I_0$ where I_0 is computed analytically.

test approaches the nominal level when the rejection rate approaches α .

The main results of the simulation experiments are presented in the form of plots where the numbers of observations are on the x -axis and the coverage error rates on the y -axis. The 5% nominal level is maintained for all tests. The horizontal solid lines in the graphs represent the nominal level 0.05.

Our simulation results show that the Fieller-type method has better coverage than the Delta method in several cases, especially when the underlying distribution involves heavy lower and upper tails.

Figure 1.1 plots the rejection probabilities of the Fieller-type and Delta methods under Singh-Maddala distributions. In the left panel, the distribution is Singh-Maddala with parameters $a = 2.8$ and $q = 1.7$. The Fieller-type and Delta methods have similar coverage. However, the other designs considered reveal important

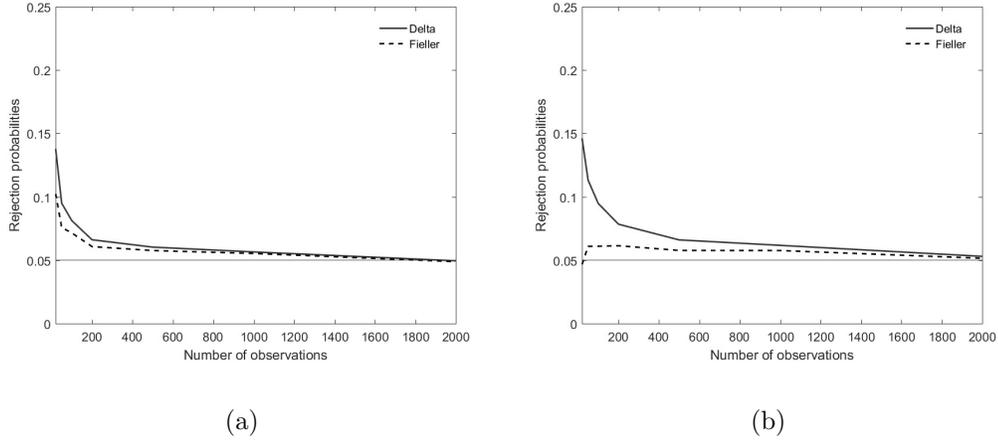


Figure 1.2: Rejection probabilities for Delta and Fieller methods
Null hypotheses: Gamma distributions

Left panel: Gamma($k = 1, \theta = 1$) ; Right panel: Gamma($k = 0.3, \theta = 1$)

Note - The Delta method and Fieller method statistics are defined by (1.2.5) and (1.2.8) respectively. The null hypothesis tested is $H(I_{T_0}) : I = I_0$ where I_0 is computed analytically.

improvement with the Fieller method. In the right panel, the distribution is Singh-Maddala with parameters $a = 1.1$ and $q = 5$. For this choice of parameters, the distribution exhibits bunching of low observations. In this context, the Fieller-type method outperforms the Delta method for relatively small samples up to 400 observations.

The same conclusion can be drawn from the case where we assume a Gamma distribution under the null. The size improvements we find with the Fieller-type method increase as the left tail of the distribution gets thicker. In the left panel of Figure 1.2, we plot the rejection probabilities under both methods under a Gamma($k = 1, \theta = 1$) distribution. The differences in the rejection probabilities for samples of size 20 is around 4%, and around 2% with 100 observations. As we increase the proportion of low observations, the Fieller-type method provides remarkable size improvements,

and in some cases it approaches the 5% nominal level for sample sizes as small as 200. In the right panel of this figure, the Fieller-type method coverage error is less than that of Delta method by around 9% going down from almost 15% to 6%.

The shape of the distributions underlying the aforementioned results represent populations where most of the individuals are poor and few are rich. The choice of these distributions was made to study the performance of the two methods when tails are fat both near the zero boundary and to the right. As we will discuss shortly our findings conform with the theoretical work of [Cowell and Victoria-Feser \(1993\)](#).

In Figures [1.3](#) and [1.4](#), we consider mixed designs with bimodal distributions under the null. These distributions, their parameters and the associated mixture weights are chosen to capture tail thickness at both ends of the distribution. Since the analytical expression for the Theil under the null of mixtures does not exist, we used an estimate of the Theil index based on a very large sample ($n = 1000000000$). This approach of computing the true Theil index under the null is justified by the consistency of the Theil Index.

Figure [1.3](#) plots the rejection probabilities for mixtures of two Singh-Maddala distributions. In the left panel, the mixture combines a $SM(1.1, 5)$ with probability 0.7 weight and $SM(2.8, 1.7)$ with probability 0.3: *i.e.* on average, we draw 70% of the sample from a distribution with a peak near low incomes, and 30% from a distribution characterized by a thick right tail. For this design, the Fieller-type method approaches the nominal significance level for samples as small as 50 observations.

In the right panel, we increase the weight for the first distribution from 0.7 to 0.9. Thus we are giving more weight to the distribution with irregularities on the left tail

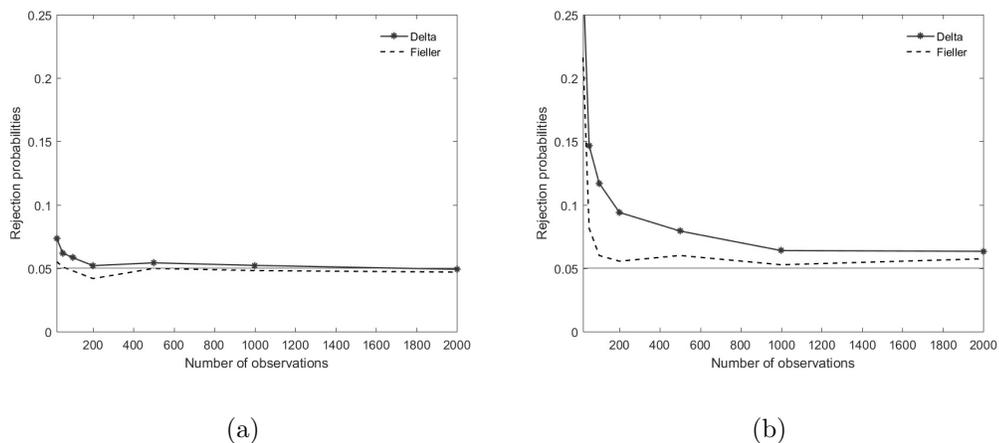


Figure 1.3: Rejection probabilities for Delta and Fieller methods
Null hypotheses: Mixtures of Singh-Maddala distributions
Left panel: $0.7*SM(1.1, 5) + 0.3*SM(2.8, 1.7)$
Right panel: $0.9*SM(1.1, 5) + 0.1*SM(2.8, 1.7)$

Note - The Delta method and Fieller method statistics are defined by (1.2.5) and (1.2.8) respectively. The null hypothesis tested is $H(I_{T_0}) : I = I_0$ where I_0 is calibrated via a separate simulation.

rather than the one characterized with right tail thickness. The Fieller-type method dominates the Delta method by wide margins. This confirms our previous conclusion that the Fieller-type method is superior to the Delta method, especially when the distribution exhibits bunching of low observations.

Further evidence appears in Figure 1.4 where we consider mixtures of Gamma(0.3, 1) and SM(2.8, 1.7) distributions. Again the left panel gives to the first distribution (the Gamma distribution) a weight of 0.7, while the right one increases this weight to 0.9. Again, the Fieller-type method improves coverage, especially for small samples.

The designs we considered can be interpreted through the work of Cowell and Victoria-Feser (1996). This chapter views the underlying distribution as a mixture of a finite number of other distributions where bunching and tail behaviour can be formally

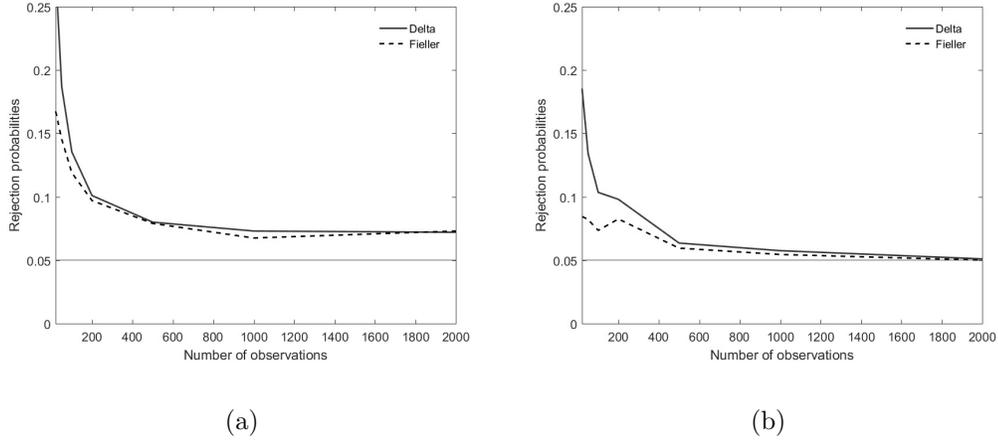


Figure 1.4: Rejection probabilities for Delta and Fieller methods
Null hypotheses: Mixtures of Gamma and Singh-Maddala distributions
Left panel: $0.7 * \text{Gamma}(0.3, 1) + 0.3 * \text{SM}(2.8, 1.7)$
Right panel: $0.9 * \text{Gamma}(0.3, 1) + 0.1 * \text{SM}(2.8, 1.7)$

Note - The Delta method and Fieller method statistics are defined by (1.2.5) and (1.2.8) respectively. The null hypothesis tested is $H(I_{T_0}) : I = I_0$ where I_0 is calibrated via a separate simulation.

modelled. Results for cases of “extreme” behaviour are pointed out, including the zero and infinite boundaries. In particular, the Theil index can have an unbounded influence function when some of the data approaches ∞ , although the zero boundary may matter for other inequality indices.

The influence function measures the change in the estimator for a small perturbation of the data. It is related to the bias of the estimator in the sense that when the IF is unbounded the bias can be infinite. Cowell and Victoria-Feser (1993) show that any decomposable scale invariant index for which the mean is estimated from the sample has an unbounded IF . We find that the Fieller-type method improves coverage, especially when small (near zero) or large observations are highly probable.

1.4 Conclusion

This chapter proposes Fieller-type procedures for inference on the Theil inequality index and illustrates its superiority relative to its standard Delta method counterpart, using various empirically relevant simulation designs. Our results confirm that, in contrast with the Delta method, the proposed procedures can capture some of the distributional irregularities arising from the concentration of low observations and the thickness of the right tail. More broadly, our findings suggest that the Fieller-type approach holds concrete promise for many other inequality measures, as well as for inference on differences between measures.

Chapter 2

Identification-robust inequality analysis

2.1 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. Typically, popular measures such as generalized entropy (GE) and Gini indices involve nonlinear transformations of parameters ([Cowell and Flachaire, 2015](#)).

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; for recent perspectives and references, see [Dufour et al. \(2019\)](#).

An important additional difficulty is that common inequality measures – such as GE and Gini indices – take the form of moment ratios, which may easily be ill-conditioned (or poorly identified). Such nonlinear forms have non-trivial implications on the properties of the associated estimators and test statistics; see [Dufour \(1997\)](#). The *first* objective of this paper is to underscore and address such weak identification in the context of inequality measures: (i) the one-sample problem of analyzing a single index, and (ii) the two-sample problem of assessing differences between two indices.

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 in this context.¹

¹See *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 \(1987\)](#).

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 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 policymakers 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 in this literature.

Indeed, the literature on statistical inference for inequality measures is relatively recent and can be traced back to [Mills and Zandvakili \(1997\)](#), who considered and recommended the use of bootstrap methods. They estimated bootstrap confidence sets for the Gini and Theil indices as well as for the “within” and “between” components of the decomposed Theil index and compared them with standard Wald-type asymptotic confidence intervals.

[Biewen \(2002\)](#) was the first to formally compare the finite-sample properties of bootstrap and asymptotic methods for inequality measures, showing that bootstrap methods lead to considerable refinement. However, as noted by [Davidson and Flachaire \(2007\)](#), Biewen’s simulation experiments assumed a log-normal distribution under the null hypothesis, which does not mimic the heavy-tailed income distributions observed in applied work. When more realistic null distributions – such as Singh-Maddala distribution – are considered, bootstrapping fails; see [Davidson and Flachaire \(2007\)](#) and [Cowell and Flachaire \(2007\)](#).

In this context, alternative methods are needed, but remain scarce for both one-sample ([Davidson and Flachaire, 2007](#); [Dufour et al., 2018](#)) and two-sample problems ([Dufour et al., 2019](#)). [Davidson and Flachaire \(2007\)](#) suggest a semiparametric bootstrap where the upper tail is modeled parametrically and the main body of the distribution is estimated in a nonparametric way. To do this, the authors use the upper-tail stability index, which is sensitive to the choice of the order statistic which defines the upper tail (h). The choice of h raises a trade-off between variance and bias, and any suboptimal choice might erode the refinements achieved. Using an approach

inspired by the weak identification literature, [Dufour et al. \(2018\)](#) take into account the possibility of ill-conditioning by proposing a Fieller-type method ([Fieller, 1944, 1954](#)). However, these results are restricted to the very special case of the Theil index (which is a member of the GE family) and only deal with the one-sample problem.

It is important to note that the two-sample problem (comparing the inequality indices of two different populations) is much more challenging than the one-sample problem (testing the value of a single index for a given population). 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. In specific cases, this approach allows one to make exact inference. In particular, a rescaling procedure validates permutations for inequality measures such as the Gini coefficient and the Theil index. The size improvements delivered through this method are most prominent when the underlying distributions are equal or similar under the null hypothesis. As the null distributions diverge, the performance of this method deteriorates which is expected as the exchangeability assumption underlying the exactness of permutation-based approaches will no longer hold. However, these results are limited to testing the equality 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 Generalized Entropy (GE thereafter) family of inequality indices, for both single and two-sample problems, with a focus on the latter. We study the general comparison problem of testing any (possibly non-zero) difference between inequality 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 this method is 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 \(2007\)](#).

The crucial difference between Fieller-type method and the standard Delta method is that the latter 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 index (or the true value of the difference of the two indices for the two-sample problem), 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).

On the GE class of inequality indices, this paper makes the following contributions. *First*, we provide analytical solutions for the Fieller one and two-sample problems, when samples can be dependent or of different sizes. *Second*, we show in a simulation study that the proposed solutions are more reliable than Delta counterparts. *Third*, for the two-sample testing problem, we show that our approach outperforms most simulation-based alternatives including the permutation test recently proposed by 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. The solutions we propose are analytically tractable. *Fifth*, we provide useful empirical evidence supporting the seemingly counter-intuitive bounds that Fieller-type methods can produce.

The simulations presented cover many cases with possibly dependent samples when the distributions under the null hypothesis can be identical or different. These results can be summarized as follows. On the one-sample problem, Fieller-type confidence sets outperform the standard Delta method under various parameter specifications for the GE_1 and GE_2 indices. For the two-sample problem, we show the superiority of Fieller-type methods across the board: (1) the improved level control (over the 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 pre-

serve 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. Note also a permutational approach is not available (at this point) to deal with the general problem we consider here. Overall, 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 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 robust confidence sets to document the economic and policy significance of statistical decisions. The empirical literature on growth relies mostly on the variance of log incomes as a measure of dispersion in per-capita income distributions [see *e.g.* [Blundell et al. \(2008\)](#)]. But this measure does not satisfy 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. Interestingly, with the GE_2 index, the Fieller-type method and the Delta method lead to contradictory conclusions: in contrast to the former, the latter suggests that inequality declines are insignificant at usual levels.

Further results with non-OECD countries stress the severe consequences of ignoring identification problems: again, 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 2.2 derives Fieller-type confidence sets for the one-sample problem and the two-sample problem. Section 2.3 reports the results of the simulation study for the one-sample and the two-sample problem. Section 2.4 contains the inter-state convergence application, and Section 2.5 concludes. The figures and tables are presented in Appendix.

2.2 Fieller-type confidence sets for Generalized Entropy inequality measures

An inequality measure is an estimator 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.

For the purpose of this paper, it is important to differentiate between two distinct inference problems encountered in the inequality literature. First, the one-sample problem refers to testing the equality of a single inequality measure with a specified (possibly zero) value, or alternatively, to building a confidence interval around a

single inequality measure. Second, the two-sample problem involves testing whether the difference between two inequality measures is equal to a specific value. The two-sample problem is more relevant in practice, as it helps evaluate the distributional impact of an economic policy or shock on the evolution of inequality over time. Examples include comparing pre-tax and post-tax income inequality, comparing inequality between economic regions, or comparing inequality over the different phases of the business cycle.

We consider the popular Generalized Entropy class of indices.² This family of indices satisfies a set of key axiomatic principles: scale invariance, the Pigou-Dalton transfer principle, the symmetry principle, and the Dalton population principle. It is also additively decomposable.

Many inequality measures, including the Generalized Entropy measures, depend solely on the underlying income distribution and they 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 . Denote by X the random variable with a typical realization representing the income of a randomly chosen individual in the population. If we denote by F_X the CDF of X , then we can express the general entropy class of indices as a ratio of functions of the moments of F_X , mainly as a function of two particular moments: the mean $\mu_X = \mathbb{E}_F(X)$ and another moment $\nu_X(\gamma) = \mathbb{E}_F(X^\gamma)$ where γ reflects different perceptions of inequality. The parameter γ characterizes the sensitivity to changes over different parts of the income distribution. The inequality measure is more sensitive to differences in the top (bottom) tail with more positive

²See Cowell (2000) for a detailed discussion on the properties of the GE indices.

(negative) γ .

The General Entropy family nests several inequality indices including two well-known inequality measures 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](#)). It is also worth noting that the Atkinson index can be obtained from the $GE_\gamma(X)$ index using an appropriate transformation. The $GE_\gamma(X)$ can be expressed as in [Shorrocks \(1980\)](#):

$$\begin{aligned} GE_\gamma(X) &= \frac{1}{\gamma(\gamma-1)} \left[\frac{\mathbb{E}_F(X^\gamma)}{[\mathbb{E}_F(X)]^\gamma} - 1 \right] \quad \text{for } \gamma \neq 0, 1, \\ GE_0(X) &= \mathbb{E}_F[\log(X)] - \log[\mathbb{E}_F(X)], \\ GE_1(X) &= \frac{\mathbb{E}_F[X \log(X)]}{\mathbb{E}_F(X)} - \log[\mathbb{E}_F(X)]. \end{aligned} \tag{2.2.1}$$

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

$$\hat{F}_X(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x) \tag{2.2.2}$$

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

$$\widehat{GE}_\gamma(X) = \frac{1}{\gamma(\gamma-1)} \left[\frac{\hat{\nu}_X(\gamma)}{\hat{\mu}_X^\gamma} - 1 \right] \tag{2.2.3}$$

where

$$\hat{\mu}_X = \int x d\hat{F}_X = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\nu}_X(\gamma) = \int x^\gamma d\hat{F}_X = \frac{1}{n} \sum_{i=1}^n X_i^\gamma. \quad (2.2.4)$$

Our aim is to make inference on the inequality measures defined by (2.2.1). In particular, we wish to build an asymptotic Fieller-type confidence set (FCS) for the difference between two $GE_\gamma(X)$ inequality measures. We call this problem the *two-sample problem*, as opposed to the *one-sample problem* where the objective simply consists in testing and building a confidence interval for a single $GE_\gamma(X)$.

We start by extending earlier work on the one-sample problem [Dufour et al. (2018) for the Theil index $GE_1(X)$] in order to cover $GE_\gamma(X)$ – for every $\gamma \in (0, 2)$. Then, we move to the two-sample problem, where we derive FCS for three cases as we shall see shortly.

2.2.1 One-sample problem

We are interested in building a confidence interval associated with the null hypothesis

$$H_D(\delta_0) : GE_\gamma(X) = \delta_0 \quad (2.2.5)$$

where δ_0 is some admissible value of the index. For the sake of comparing the performance of the FCS approach with that of the standard asymptotic procedure based on the Delta method (DCS), we now derive confidence interval formulas for both the delta method and the Fieller method. The Delta method confidence set (DCS) is

obtained by inverting the square (or the absolute value) of the standard Wald-type t-test. 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 . This can be mathematically carried out by solving the following inequality for δ

$$\hat{T}_D(\delta_0)^2 = \left[\frac{\widehat{GE}_\gamma(X) - \delta_0}{(\hat{V}_D[\widehat{GE}_\gamma(X)])^{1/2}} \right]^2 \leq z_{\alpha/2}^2 \quad (2.2.6)$$

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.

Assuming that the estimator is asymptotically normal, we can use the Delta method to estimate $V_D[\widehat{GE}_\gamma(X)]$:

$$V_D[GE_\gamma(X)] = \frac{1}{n} \begin{bmatrix} \frac{\partial GE_\gamma(x)}{\partial \mu_x} & \frac{\partial GE_\gamma(x)}{\partial \nu_x} \end{bmatrix} \begin{bmatrix} \sigma_x^2 & \sigma_{x,x^\gamma} \\ \sigma_{x^\gamma,x} & \sigma_{x^\gamma}^2 \end{bmatrix} \begin{bmatrix} \frac{\partial GE_\gamma(x)}{\partial \mu_x} & \frac{\partial GE_\gamma(x)}{\partial \nu_x} \end{bmatrix}' \quad (2.2.7)$$

where $\sigma_X^2, \sigma_{X^\gamma}^2$ and σ_{X, X^γ} represent the variance of X , the variance of X^γ and the covariance between X and X^γ respectively. Note that the estimated variance of $\hat{\mu}_X$, variance of $\hat{\nu}_X$ and covariance between $\hat{\mu}_X$ and $\hat{\nu}_X$ are equal to $\hat{\sigma}_X^2/n$, $\hat{\sigma}_{X^\gamma}^2/n$ and $\hat{\sigma}_{X, X^\gamma}/n$ respectively. In our estimation, we use the sample counterparts for these population moments estimated using the EDF of the two samples. Solving (2.2.6) for δ_0 and plugging in the estimate $\hat{V}_D := \hat{V}_D[\widehat{GE}_\gamma(X)]$ of the variance in (2.2.7) we get

the DCS:

$$\text{DCS}[GE_\gamma(X); 1 - \alpha] = \left[\widehat{GE}_\gamma(X) - z_{\alpha/2} \widehat{V}_D^{1/2}, \widehat{GE}_\gamma(X) + z_{\alpha/2} \widehat{V}_D^{1/2} \right]. \quad (2.2.8)$$

Fieller-type confidence sets are typically applied when testing parameter ratios. The main difference between the Delta method approach and that of Fieller is that the latter writes the null hypothesis in the linear form. Thus, the Fieller-type method rewrites $H_D(\delta_0)$ defined by (2.2.5) as

$$H_F(\delta_0) : \theta(\delta_0) = 0, \quad \text{where} \quad \theta(\delta_0) = \nu_X(\gamma) - \mu_X^\gamma - \gamma(1 - \gamma)\mu_X^\gamma \delta_0. \quad (2.2.9)$$

The method then proceeds by inverting the square of the t-test associated with the linear null hypothesis $H_F(\delta_0)$. Technically, it solves the quadratic inequality

$$\hat{T}_F(\delta_0)^2 = \left[\frac{\hat{\theta}(\delta_0)}{(\hat{V}[\hat{\theta}(\delta_0)])^{1/2}} \right]^2 \leq z_{\alpha/2}^2 \quad (2.2.10)$$

where $\hat{V}[\hat{\theta}(\delta_0)]$ is an estimate of the variance of $\hat{\theta}(\delta_0)$. Assuming that the statistic follows asymptotically a standard normal distribution, a Fieller-type confidence set with level $1 - \alpha$ can be constructed by noting that

$$\hat{T}_F(\delta_0)^2 \leq z_{\alpha/2}^2 \quad \Leftrightarrow \quad \frac{\hat{\theta}(\delta_0)^2}{\hat{V}[\hat{\theta}(\delta_0)]} \leq z_{\alpha/2}^2 \quad \Leftrightarrow \quad \hat{\theta}(\delta_0)^2 - z_{\alpha/2}^2 \hat{V}[\hat{\theta}(\delta_0)] \leq 0. \quad (2.2.11)$$

Since $\hat{\theta}(\delta_0)$ is linear in δ_0 , $\hat{\theta}(\delta_0)^2$ and $\hat{V}[\hat{\theta}(\delta_0)]$ are quadratic functions of δ_0 :

$$\hat{\theta}(\delta_0)^2 = \hat{A}_1\delta_0^2 + \hat{B}_1\delta_0 + \hat{C}_1, \quad \hat{V}[\hat{\theta}(\delta_0)] = \hat{A}_2\delta_0^2 + \hat{B}_2\delta_0 + \hat{C}_2. \quad (2.2.12)$$

On substituting (2.2.12) into (2.2.11), we get the quadratic inequality

$$\hat{A}\delta_0^2 + \hat{B}\delta_0 + \hat{C} \leq 0 \quad (2.2.13)$$

which yields the following (Fieller-type) confidence set for δ_0 :

$$\text{FCS}[GE_\gamma(X); 1 - \alpha] = \left\{ \delta_0 : \hat{A}\delta_0^2 + \hat{B}\delta_0 + \hat{C} \leq 0 \right\} \quad (2.2.14)$$

with

$$A = A_1 - z_{\alpha/2}^2 A_2, \quad B = B_1 - z_{\alpha/2}^2 B_2, \quad C = C_1 - z_{\alpha/2}^2 C_2. \quad (2.2.15)$$

The coefficients A_1 , B_1 , C_1 , A_2 , B_2 , and C_2 are functions of the sample moments and their variance estimates:

$$\begin{aligned} A_1 &= \hat{\mu}_X^{2\gamma} [\gamma^2 - \gamma]^2 \\ B_1 &= -2\hat{\mu}_X^\gamma [\gamma^2 - \gamma] [\hat{\nu}_X(\gamma) - \hat{\mu}_X^\gamma] \\ C_1 &= (\hat{\nu}_X(\gamma) - \hat{\mu}_X^\gamma)^2. \end{aligned} \quad (2.2.16)$$

$$\begin{aligned}
A_2 &= [\sigma_X^2 \gamma^2 [\gamma^2 - \gamma]^2 \mu_X^{2(\gamma-1)}] / n \\
B_2 &= [2\sigma_X^2 \gamma^2 [\gamma^2 - \gamma] \mu_X^{2(\gamma-1)} - 2\gamma [\gamma^2 - \gamma] \sigma_{X, X^\gamma} \mu_X^{\gamma-1}] / n \\
C_2 &= [\sigma_X^2 \gamma^2 \mu_X^{2(\gamma-1)} - 2\gamma \sigma_{X, X^\gamma} \mu_X^{\gamma-1} + \sigma_{X^\gamma}^2] / n.
\end{aligned} \tag{2.2.17}$$

On setting $D = B^2 - 4AC$, the $(1 - \alpha)$ -level Fieller-type confidence set can be described as follows:

1. if $D > 0$ and $A > 0$, $\text{FC}(I; 1 - \alpha) = \left[\frac{-B - \sqrt{D}}{2A}, \frac{-B + \sqrt{D}}{2A} \right]$;
2. if $D > 0$ and $A < 0$, $\text{FC}(I; 1 - \alpha) = \left] -\infty, \frac{-B + \sqrt{D}}{2A} \right] \cup \left[\frac{-B - \sqrt{D}}{2A}, +\infty \right[$;
3. if $D < 0$, $A < 0$ and $\text{FC}(I; 1 - \alpha) = \mathbb{R}$.

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](#)).

2.2.2 Two-sample problem

Although the construction of Fieller-type confidence sets in the two-sample problem follows the same steps as in the one-sample problem, the fact that we have two samples gives rise to three cases defined by sample size and the dependence between the two samples: (1) samples are of equal sizes and independent; (2) samples have different sample size and are independent; (3) samples are dependent and are of equal sizes.

In what follows, we derive the DCS and FCS for each of the three cases. 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.

Let us first introduce some notation. We denote by X the random variable representing the incomes of individuals from the first population with cumulative distribution functions (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, \dots, X_n and Y_1, \dots, Y_m from each population. EDF's are obtained as usual:

$$\hat{F}_X(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i < x), \quad \hat{F}_Y(y) = \frac{1}{m} \sum_{i=1}^m \mathbf{1}(Y_i < y). \quad (2.2.18)$$

Using \hat{F}_X and \hat{F}_Y , we can estimate the inequality measures $GE_\gamma(X)$ and $GE_\gamma(Y)$ for the samples. We consider null hypotheses of the form

$$H_D(\Delta_0) : \Delta GE_\gamma = \Delta_0 \quad (2.2.19)$$

where $\Delta GE_\gamma := GE_\gamma(X) - GE_\gamma(Y)$ and Δ_0 is any known admissible value of ΔGE_γ (including possibly $\Delta_0 = 0$, for equality). The square of the asymptotic t-type statistic for the hypothesis $H_D(\Delta_0)$ is

$$\hat{W}_D(\Delta_0)^2 = \frac{[\Delta \widehat{GE}_\gamma - \Delta_0]^2}{\hat{V}_D[\Delta \widehat{GE}_\gamma]} \quad (2.2.20)$$

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

$$\text{DCS}(\Delta\widehat{GE}_\gamma; 1 - \alpha) = \left[\widehat{GE}_\gamma(X) - z_{\alpha/2} [\hat{V}_D(\Delta\widehat{GE}_\gamma)]^{1/2}, \widehat{GE}_\gamma(X) + z_{\alpha/2} [\hat{V}_D(\Delta\widehat{GE}_\gamma)]^{1/2} \right]. \quad (2.2.21)$$

The estimation of the variance $V_D(\Delta\widehat{GE}_\gamma)$ in (2.2.21) 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:

$$V(\Delta\widehat{GE}_\gamma) = \begin{bmatrix} \frac{\partial\Delta\widehat{GE}_\gamma}{\partial\mu_x} \\ \frac{\partial\Delta\widehat{GE}_\gamma}{\partial\nu_x} \\ \frac{\partial\Delta\widehat{GE}_\gamma}{\partial\mu_y} \\ \frac{\partial\Delta\widehat{GE}_\gamma}{\partial\nu_y} \end{bmatrix}' \begin{bmatrix} \Sigma_{xx}/n & \Sigma_{xy}/n \\ \Sigma_{yx}/n & \Sigma_{yy}/n \end{bmatrix} \begin{bmatrix} \frac{\partial\Delta\widehat{GE}_\gamma}{\partial\mu_x} \\ \frac{\partial\Delta\widehat{GE}_\gamma}{\partial\nu_x} \\ \frac{\partial\Delta\widehat{GE}_\gamma}{\partial\mu_y} \\ \frac{\partial\Delta\widehat{GE}_\gamma}{\partial\nu_y} \end{bmatrix} = R \begin{bmatrix} \Sigma_{xx}/n & \Sigma_{xy}/n \\ \Sigma_{yx}/n & \Sigma_{yy}/n \end{bmatrix} R' \quad (2.2.22)$$

where

$$\Sigma_{XX} = \begin{bmatrix} \sigma_x^2 & \sigma_{x,x^\gamma} \\ \sigma_{x^\gamma,x} & \sigma_{x^\gamma}^2 \end{bmatrix}, \quad \Sigma_{YY} = \begin{bmatrix} \sigma_y^2 & \sigma_{y,y^\gamma} \\ \sigma_{y^\gamma,y} & \sigma_{y^\gamma}^2 \end{bmatrix}, \quad \Sigma_{XY} = \begin{bmatrix} \sigma_{x,y} & \sigma_{x,y^\gamma} \\ \sigma_{x^\gamma,y} & \sigma_{x^\gamma,y^\gamma} \end{bmatrix}, \quad \Sigma_{YX} = \Sigma'_{XY}. \quad (2.2.23)$$

The variance under the first case can be determined simply by setting Σ_{XY} in (2.2.22) equal to zero since the samples are assumed to be independent: in this case,

$$V_1(\Delta\widehat{GE}_\gamma) = R \begin{bmatrix} \Sigma_{xx}/n & \mathbf{0} \\ \mathbf{0} & \Sigma_{yy}/n \end{bmatrix} R'. \quad (2.2.24)$$

In the second case, where the samples are independent with unequal sizes, the variance

is determined by setting Σ_{XY} in (2.2.22) equal to zero and by dividing Σ_{YY} by m instead of n :

$$V_2(\widehat{\Delta GE}_\gamma) = R \begin{bmatrix} \Sigma_{xx}/n & \mathbf{0} \\ \mathbf{0} & \Sigma_{yy}/m \end{bmatrix} R'. \quad (2.2.25)$$

The third case requires taking into account the dependence between the two samples. Since the samples are of equal size, the computation of the variance is straightforward. In this case, the covariance matrix Σ_{XY} is not equal to zero. The variance is actually nothing but the expression defined by (2.2.22) itself without imposing earlier restrictions [*i.e.*, $V_3(\widehat{\Delta GE}_\gamma) = V(\widehat{\Delta GE}_\gamma)$].

Turning to the Fieller-type method, it proceeds, as in the one-sample problem, by reformulating the null hypothesis in a linear form (without the ratio transformation). Such a reformulation can be obtained through multiplication of both sides of (2.2.19) by the common denominator $\gamma(\gamma - 1)\mu_X^\gamma\mu_Y^\gamma$:

$$H_F(\Delta_0) : \Theta(\Delta_0) = 0 \quad \text{where} \quad \Theta(\Delta_0) = \nu_X(\gamma)\mu_Y^\gamma - \nu_Y(\gamma)\mu_X^\gamma - \gamma(\gamma - 1)\mu_X^\gamma\mu_Y^\gamma\Delta_0. \quad (2.2.26)$$

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

$$\hat{W}_F(\Delta_0)^2 = \left[\frac{\hat{\Theta}(\Delta_0)}{(\hat{V}[\hat{\Theta}(\Delta_0)])^{1/2}} \right]^2 \leq z_{\alpha/2}^2 \quad (2.2.27)$$

where $\hat{V}[\hat{\Theta}(\Delta_0)]$ is an estimate of the variance of $\hat{\Theta}(\Delta_0)$. Again, we will consider three empirically relevant scenarios where for every one of them we will have a different expression for the variance; more on the estimation of the variance below.

To obtain a Fieller-type confidence set, we solve the inequality in (2.2.27) for Δ_0 . This can be done by reformulating the inequality in (2.2.27) as a quadratic inequality:

$$\hat{W}_F(\Delta_0)^2 \leq z_{\alpha/2}^2 \Leftrightarrow \frac{\hat{\Theta}(\Delta_0)^2}{\hat{V}[\hat{\Theta}(\Delta_0)]} \leq z_{\alpha/2}^2 \Leftrightarrow \hat{\Theta}(\Delta_0)^2 - z_{\alpha/2}^2 \hat{V}[\hat{\Theta}(\Delta_0)] \leq 0. \quad (2.2.28)$$

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

$$\hat{\Theta}(\Delta_0)^2 = \hat{A}_1 \Delta_0^2 + \hat{B}_1 \Delta_0 + \hat{C}_1, \quad \hat{V}[\hat{\Theta}(\Delta_0)] = \hat{A}_2 \Delta_0^2 + \hat{B}_2 \Delta_0 + \hat{C}_2, \quad (2.2.29)$$

where the parameters A_1, B_1, C_1, A_2, B_2 and C_2 are functions of the moments of the EDFs and are defined by equations (2.2.32) to (2.2.35). Since $\hat{\Theta}(\Delta_0)^2$ and $\hat{V}[\hat{\Theta}(\Delta_0)]$ are both quadratic functions of Δ_0 , then $\hat{\Theta}(\Delta_0)^2 - z_{\alpha/2}^2 \hat{V}[\hat{\Theta}(\Delta_0)] \leq 0$ is a quadratic inequality which can be expressed as $A\Delta_0^2 + B\Delta_0 + C \leq 0$. The Fieller-type confidence set is then:

$$\text{FCS}(\Delta GE_\gamma; 1 - \alpha) = \left\{ \Delta_0 : \hat{A}\Delta_0^2 + \hat{B}\Delta_0 + \hat{C} \leq 0 \right\} \quad (2.2.30)$$

where

$$A = A_1 - z_{\alpha/2}^2 A_2, \quad B = B_1 - z_{\alpha/2}^2 B_2, \quad C = C_1 - z_{\alpha/2}^2 C_2, \quad (2.2.31)$$

$$\begin{aligned} A_1 &= \mu_X^{2\gamma} \mu_Y^{2\gamma} [\gamma^2 - \gamma]^2, \\ B_1 &= -2\mu_X^\gamma \mu_Y^\gamma [\gamma^2 - \gamma] [\nu_X(\gamma) \mu_Y^\gamma - \nu_Y(\gamma) \mu_X^\gamma], \\ C_1 &= [\nu_X(\gamma) \mu_Y^\gamma - \nu_Y(\gamma) \mu_X^\gamma]^2. \end{aligned} \quad (2.2.32)$$

The parameters A_2, B_2 and C_2 are obtained from the formula of $\hat{V}[\hat{\Theta}(\Delta_0)]$, so they differ depending on the case considered. Below is the most general form of the parameters A_2, B_2 and C_2 . For each of the three cases, we will impose some restrictions on the parameters defined below to get the parameters associated with the relevant variance:

$$\begin{aligned}
A_2 = & \left[\gamma[\gamma^2 - \gamma] \mu_X^{(\gamma-1)} \mu_Y^\gamma \right]^2 \frac{\sigma_X^2}{n} + \left[\gamma[\gamma^2 - \gamma] \mu_Y^{(\gamma-1)} \mu_X^\gamma \right]^2 \frac{\sigma_Y^2}{m} \\
& + 2\gamma^2[\gamma^2 - \gamma] \mu_X^{2\gamma-1} \mu_Y^{2\gamma-1} \frac{\sigma_{X,Y}}{n}, \tag{2.2.33}
\end{aligned}$$

$$\begin{aligned}
B_2 = & 2 \left[\gamma^2[\gamma^2 - \gamma] \mu_X^{2(\gamma-1)} \mu_Y^\gamma \nu_Y(\gamma) \frac{\sigma_X^2}{n} - \gamma[\gamma^2 - \gamma] \mu_X^{\gamma-1} \mu_Y^{2\gamma} \frac{\sigma_{X,X^\gamma}}{n} \right. \\
& - \gamma^2[\gamma^2 - \gamma] \mu_Y^{2(\gamma-1)} \mu_X^\gamma \nu_X(\gamma) \sigma_Y^2 + \gamma[\gamma^2 - \gamma] \mu_Y^{\gamma-1} \mu_X^{2\gamma} \frac{\sigma_{Y,Y^\gamma}}{m} \\
& + \gamma^2[\gamma^2 - \gamma] \mu_X^{\gamma-1} \mu_Y^{\gamma-1} (\mu_X^\gamma \nu_Y(\gamma) - \mu_Y^\gamma \nu_X(\gamma)) \frac{\sigma_{X,Y}}{n} \\
& \left. + \gamma[\gamma^2 - \gamma] \mu_X^{2\gamma-1} \mu_Y^\gamma \frac{\sigma_{X,Y^\gamma}}{n} - \gamma[\gamma^2 - \gamma] \mu_Y^{2\gamma-1} \mu_X^\gamma \frac{\sigma_{Y,X^\gamma}}{n} \right], \tag{2.2.34}
\end{aligned}$$

$$\begin{aligned}
C_2 = & \left[\gamma \nu_Y(\gamma) \mu_X^{\gamma-1} \right]^2 \frac{\sigma_X^2}{n} + \mu_Y^{2\gamma} \frac{\sigma_{X^\gamma}}{n} - 2\gamma \nu_Y(\gamma) \mu_X^{\gamma-1} \mu_Y^\gamma \frac{\sigma_{X,X^\gamma}}{n} \\
& + \left[\gamma \nu_X(\gamma) \mu_Y^{\gamma-1} \right]^2 \frac{\sigma_Y^2}{m} + \mu_X^{2\gamma} \frac{\sigma_{Y^\gamma}}{m} - 2\gamma \nu_X(\gamma) \mu_Y^{\gamma-1} \mu_X^\gamma \frac{\sigma_{Y,Y^\gamma}}{m} \\
& - 2\gamma^2 \mu_X^{\gamma-1} \mu_Y^{\gamma-1} \nu_X(\gamma) \nu_Y(\gamma) \frac{\sigma_{X,Y}}{n} + 2\gamma \mu_X^{2\gamma-1} \nu_Y(\gamma) \frac{\sigma_{X,Y^\gamma}}{n} \\
& + 2\gamma \mu_Y^{2\gamma-1} \nu_X(\gamma) \frac{\sigma_{Y,X^\gamma}}{n} - 2\mu_X^\gamma \mu_Y^\gamma \frac{\sigma_{X^\gamma,Y^\gamma}}{n}. \tag{2.2.35}
\end{aligned}$$

The parameters A_2, B_2 and C_2 in each one of the three cases can be obtained as follows:

- case one: $\sigma_{X,Y}$, $\sigma_{X^\gamma,Y}$, σ_{X,Y^γ} and $\sigma_{X^\gamma,Y^\gamma}$ equal to zero and setting $n = m$;
- case two: $\sigma_{X,Y}$, $\sigma_{X^\gamma,Y}$, σ_{X,Y^γ} and $\sigma_{X^\gamma,Y^\gamma}$ equal to zero;
- case three : $n = m$.

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 (2.2.6), (2.2.20), (2.2.10) and (2.2.27) 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.

2.3 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

$$F_X(x) = 1 - \left[1 + \left(\frac{x}{b_X} \right)^{a_X} \right]^{-q_X} \quad (2.3.1)$$

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:

$$\nu_X(\gamma) := \mathbb{E}(X^\gamma) = \frac{b_X^\gamma \Gamma(\gamma a_X^{-1} + 1) \Gamma(q_X - \gamma a_X^{-1})}{\Gamma(q_X)} \quad (2.3.2)$$

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% for both size and power analysis.

2.3.1 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 $\hat{T}_D(\delta_0)^2 > z_{\alpha/2}^2$, where $\hat{T}_D(\delta_0)^2$ is defined in (2.2.6), or check whether the confidence set $\text{DCS}[GE_\gamma(X); 1 - \alpha]$ defined in (2.2.8) contains the tested value δ_0 . Similarly, for the Fieller method, we can either use the critical region $\hat{T}_F(\delta_0)^2 > z_{\alpha/2}^2$, where $\hat{T}_F(\delta_0)^2$ is defined in (2.2.10), or check whether the confidence set $\text{FCS}[GE_\gamma(X); 1 - \alpha]$ defined in (2.2.14) contains the tested value δ_0 . Both approaches are numerically equivalent and yield the same results. If $\mathbb{P}[\hat{T}_F(\delta_0)^2 > z_{\alpha/2}^2] = p(\delta_0)$ for a distribution which satisfies $GE_\gamma = \delta_0$, then $\mathbb{P}[\delta_0 \in \text{FCS}[GE_\gamma(X); 1 - \alpha]] = 1 - p(\delta_0)$ is the coverage probability for δ_0 in this case, and similarly for $\hat{T}_D(\delta_0)^2$.

The left panel of Figure 2.1 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 2.2]. 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 robust-

ness 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 2.1 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 2.1 and 2.2 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.

2.3.2 Simulation results: two-sample problem

To assess the reliability of the procedures proposed above, 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 ($\Delta_0 \neq 0$). The two-sample problem differs from the one-sample problem in more than one aspect. First, the underlying two samples may not have the same size; and second, the two samples could be dependent. So, 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 $F_X \neq F_Y$;
- (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 $F_X \neq F_Y$;
- (c) $\Delta_0 \neq 0$.

3. Experiment III – Dependent samples of equal sizes ($m = n$):

- (a) $\Delta_0 = 0$ with $F_X = F_Y$;
- (b) $\Delta_0 = 0$ with $F_X \neq F_Y$;
- (c) $\Delta_0 \neq 0$.

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 $\hat{W}_D(\Delta_0)^2 >$

$z_{\alpha/2}^2$, where $\hat{W}_D(\Delta_0)^2$ is defined in (2.2.20), or check whether the confidence set $\text{DCS}(\Delta GE_\gamma; 1 - \alpha)$ defined in (2.2.21) contains the tested value Δ_0 . Similarly, for the Fieller method, we can either use the critical region $\hat{W}_F(\Delta_0)^2 > z_{\alpha/2}^2$, where $\hat{W}_F(\Delta_0)^2$ is defined in (2.2.27), or check whether the confidence set $\text{FCS}(\Delta GE_\gamma; 1 - \alpha)$ defined in (2.2.30) contains the tested value Δ_0 . Both approaches are numerically equivalent and yield the same results (a feature we did check). If $\mathbb{P}[\hat{W}_F(\Delta_0)^2 > z_{\alpha/2}^2] = p(\Delta_0)$ for a pair of distribution which satisfy $GE_\gamma(X) - GE_\gamma(Y) = \Delta_0$, then $\mathbb{P}[\Delta_0 \in \text{FCS}[GE_\gamma(X); 1 - \alpha]] = 1 - p(\Delta_0)$ 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 remainder 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 2.3 and 2.4 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 [$\Delta_0 = 0$ with $F_X \neq F_Y$]. The left panel of Figure 2.5 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 2.7 and 2.8, 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 2.3 to 2.8 illustrate the power of FCS and DCS for both GE_1 and GE_2 under the three scenarios considered: $[\Delta_0 = 0$ with $F_X = F_Y]$, $[\Delta_0 = 0$ 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 $n = 2m$.³

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 con-

³The results presented here are not sensitive to choice of the ratio between n and m

trol 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 illustrated graphically in Figures 2.9 to 2.14.

Figures 2.9 and 2.10 depict the rejection frequencies of FCS and DCS for GE_1 and GE_2 in the left panel and the right panel respectively where the true indices are assumed equal with identical null distributions [$\Delta_0 = 0$ with $F_X = F_Y$]. For instance, the left panel shows that FCS reduces the size distortions by about 6 percentage points when the two sample sizes are 100 and 50 respectively. In Figures 2.11 and 2.12 the indices are still assumed to be equal under the null hypothesis, however the underlying null distributions are not identical [$\Delta_0 = 0$ with $F_X \neq F_Y$]. Under this scenario, FCS has almost exact level for both GE_1 and GE_2 whereas the DCS suffers from overrejection when sample sizes are relatively small. Similar conclusions can be made regarding the last scenario which assumes unequal indices under the null hypothesis, see Figures 2.13 and 2.14.

The right panels of Figures 2.9 to 2.14 illustrate the power of FCS and DCS for both GE_1 and GE_2 . Results reveal that the Fieller-type method is as powerful as Delta method when compared at sample sizes where both FCS and DCS have similar rejection frequencies.

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 simu-

lation 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, the results on the Fieller-type method are in line with the independent cases in terms of reducing size distortions especially in small samples and when larger γ is used. The extent of the size gains is however lower than their counterparts under the independent experiments. This is mainly due to the fact that dependence reduces the information content conveyed by the data. In terms of power, both the FCS and DCS methods perform similarly when compared at sample sizes for which the rejection frequencies of both methods are similar and close to the nominal level. The left panels of Figures 2.15 to 2.20 show the size plots, whereas the right panels show those of power.

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. These distortions are expected since the exchangeability assumption underlying the exactness of the permutation method no longer holds.

Figures 2.21 and 2.22 plot size and power of the permutation method and the Fieller-type method 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 focus on comparing the methods when the exactness result 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, 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 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 2.2 and 2.3.

The results in these tables rely on samples of 50 observations. Table 2.2 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 $q_X = 1.7$ and the parameter a_Y and q_Y are varied in a way such that the left tail becomes thicker and the right tail is left unchanged. This is done by decreasing the value of the parameter a_Y , and increasing the parameter q_Y enough to keep the tail index fixed ($\xi_X = \xi_Y = 4.76$). The last column of Table 2.3 shows the percentage difference of the rejection frequencies between the Fieller method over the Delta method. As the left tail becomes thicker, the performance of the Delta method deteriorates gets substantially inferior to the more than 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 2.23. The left panel plots rejection frequencies of DCS and FCS for $\gamma \in [0, 3.5]$ assuming the two samples are independent. 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 not affected by the independence assumption as shown in the right panel where the rejection frequencies are plotted against γ assuming 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 2.24, 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 – As pointed out above, 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 2.4 and Table 2.5 corroborate this fact, as over-rejections remain even with samples as large as 200000, particularly with the GE_2 which 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 2.5 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 2.4 illustrate.

2.4 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 [drawn from [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, we find 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 2.6](#), 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 [2.7](#) presents estimates and confidence sets for the difference of inequality measures between the two periods. The first column reports the results associated with the Theil index and the second with the GE_2 index. The main result that we would like to highlight from this table 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, uninformative.

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 from insignificant no-change test outcomes. Far more attention needs to be 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.

2.5 Conclusion

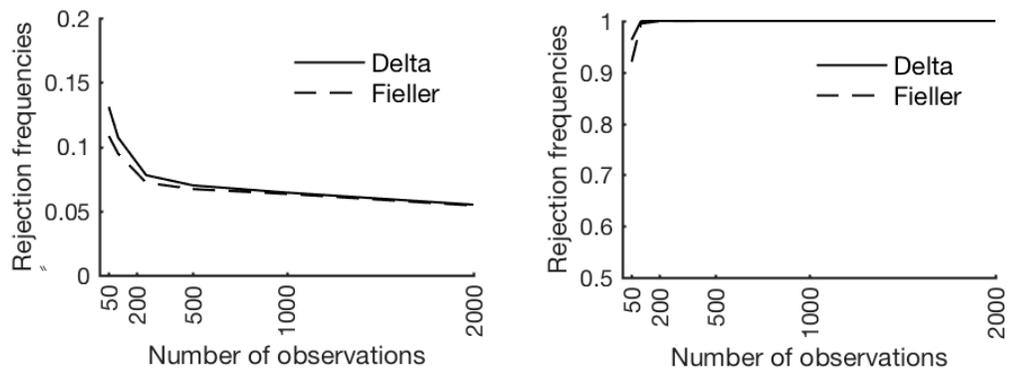
This paper introduces a Fieller-type method for inference on the GE class of inequality indices, in the one and two-sample problem with a focus on the latter. Simulation results confirm that the Fieller-type method outperforms standard counterparts including the permutation test, over all experiments 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.

Appendix: Figures and Tables

One-sample problem



(a) GE_1 : Size

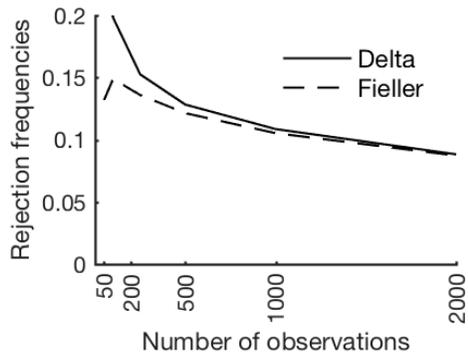
(b) GE_1 : Power

Figure 2.1: Size and power of Delta(-method) and Fieller-type tests for GE_1 (Theil) index

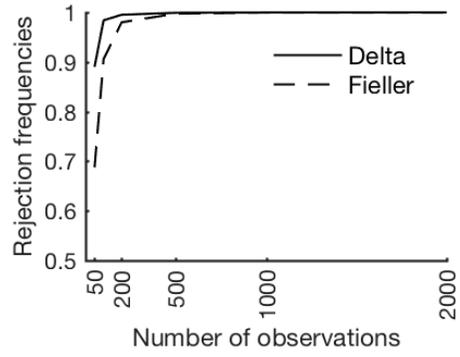
$$H_0: GE_1(X) = 0.4929$$

Left panel DGP: $SM_X(a_X = 1.1, q_X = 4.327273)$. $GE_1(X) = 0.4929$

Right panel DGP: $SM_X(a_X = 1.7, q_X = 2.8)$. $GE_1(X) = 0.27137$



(a) GE_2 : Size



(b) GE_2 : Power

Figure 2.2: Size and power of Delta(-method) and Fieller-type tests for GE_2 index

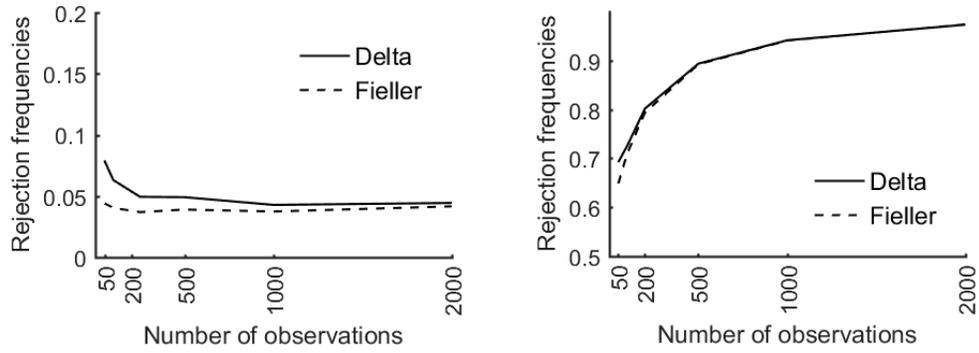
$$H_0: GE_2(X) = 0.71578$$

Left panel DGP: $SM_X(a_X = 1.1, q_X = 4.327273)$. $GE_2(X) = 0.71578$

Right panel DGP: $SM_X(a_X = 1.7, q_X = 2.8)$. $GE_2(X) = 0.33503$

Two-sample problem

Design I(a) – Independent samples: $n = m$, $F_X = F_Y$, $\Delta_0 = 0$



(a) GE_1 : Size

(b) GE_1 : Power

Figure 2.3: Design I(a) – Size and power of Delta and Fieller-type tests for GE_1 comparisons (used to derive confidence sets)

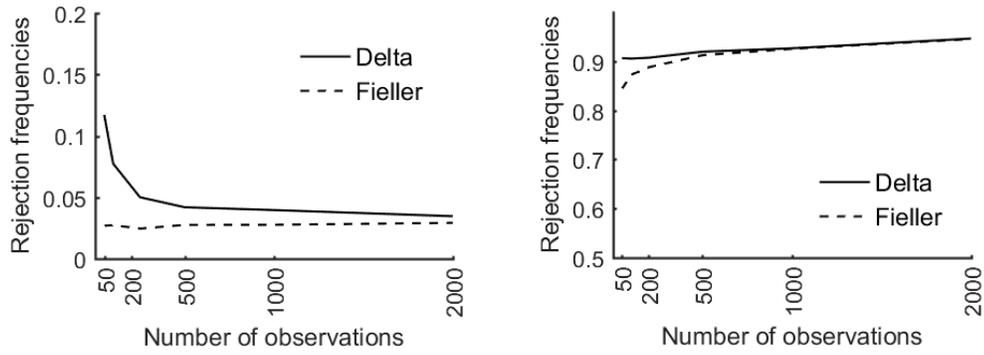
$$H_0: GE_1(X) = GE_1(Y)$$

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: $SM_X(a_X = 4.8, q_X = 0.499616)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$.

$$GE_1(X) = 0.22857, GE_1(Y) = 0.09514$$



(a) GE_2 : Size

(b) GE_2 : Power

Figure 2.4: Design I(a) – Size and power of Delta and Fieller-type tests for GE_2 comparisons (used to derive confidence sets)

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

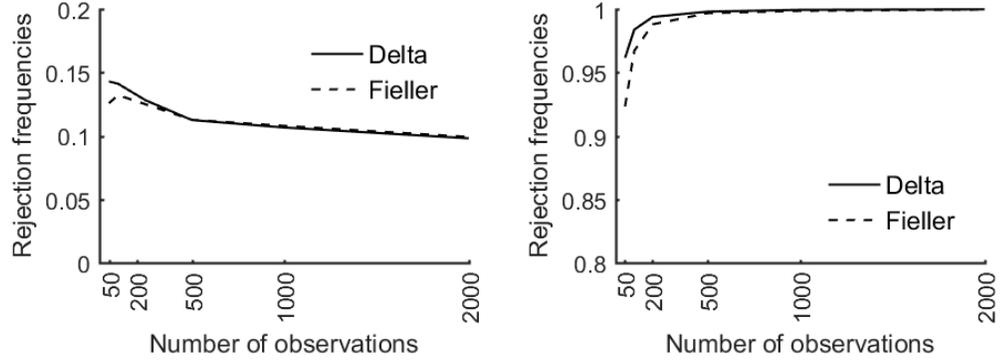
Left panel: $SM_X(a_X = 5.8, q_X = 0.499616)$, $SM_Y(a_Y = 5.8, q_Y = 0.499616)$.

$$GE_2(X) = GE_2(Y) = 0.24396$$

Right panel: $SM_X(a_X = 4.8, q_X = 0.499616)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$.

$$GE_2(X) = 0.63705, GE_2(Y) = 0.13806$$

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



(a) GE_1 : Size

(b) GE_1 : Power

Figure 2.5: Design I(b) – Size and power of Delta and Fieller-type tests for GE_1 comparisons

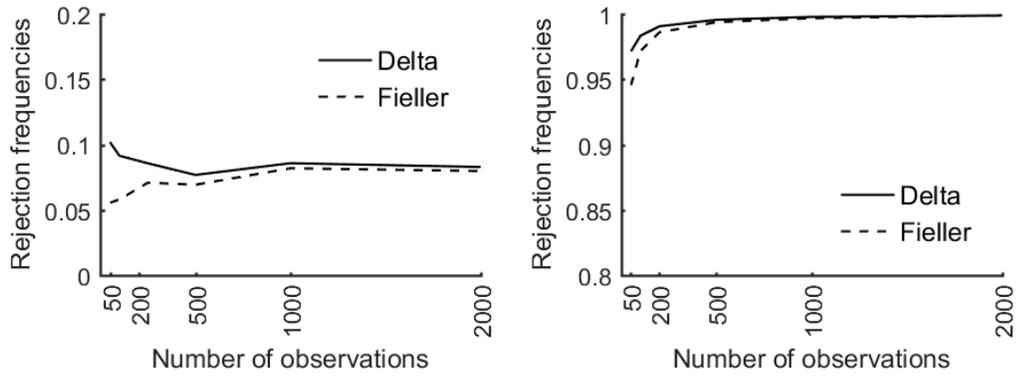
$$H_0: GE_1(X) = GE_1(Y)$$

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: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$.

$$GE_1(X) = 0.33830, GE_1(Y) = 0.09514$$



(a) GE_2 : Size

(b) GE_2 : Power

Figure 2.6: Design I(b) – Size and power of Delta and Fieller-type tests for GE_2 comparisons

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

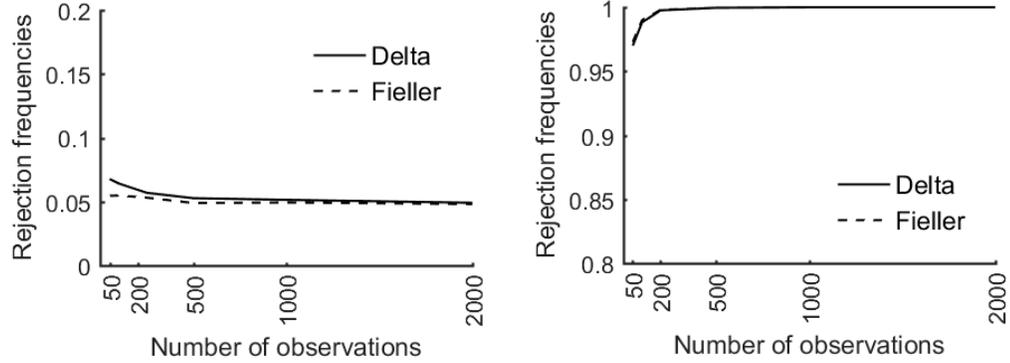
Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 3.8, q_Y = 0.9831)$.

$$GE_2(X) = GE_2(Y) = 0.16204$$

Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4.8, q_Y = 0.9831)$.

$$GE_2(X) = 0.5479, GE_2(Y) = 0.08835$$

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



(a) GE_1 : Size

(b) GE_1 : Power

Figure 2.7: Design I(c) – Size and power of Delta and Fieller-type tests for GE_1 comparisons

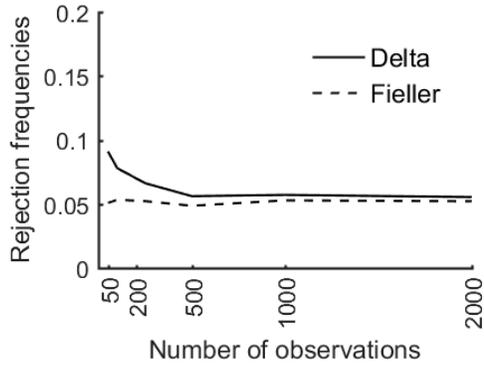
$$H_0: GE_1(X) - GE_1(Y) = 0.04670$$

Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 3.8, q_Y = 1.3061)$.

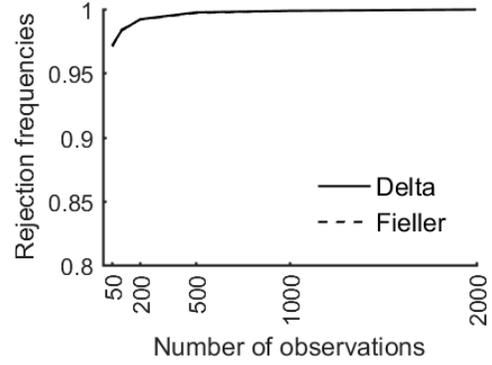
$$GE_1(X) = 0.14011, GE_1(Y) = 0.09340$$

Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4.8, q_Y = 1.3061)$.

$$GE_1(X) = 0.33829, GE_1(Y) = 0.05839$$



(a) GE_2 : Size



(b) GE_2 : Power

Figure 2.8: Design I(c) – Size and power of Delta and Fieller-type tests for GE_2 comparisons

$$H_0: GE_2(X) - GE_2(Y) = 0.05401$$

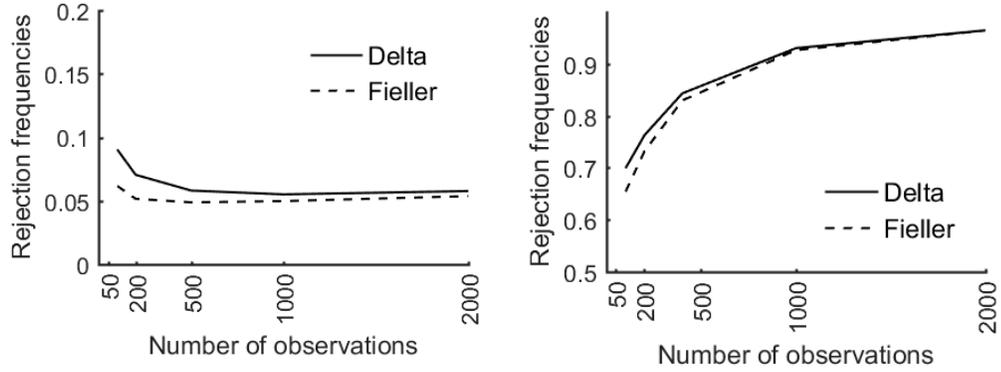
Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 3.8, q_Y = 1.2855)$.

$$GE_2(X) = 0.16203, GE_2(Y) = 0.10802$$

Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4.8, q_Y = 1.2855)$.

$$GE_2(X) = 0.54790, GE_2(Y) = 0.06367$$

Design II(a) – Independent samples: $n = 2m$, $F_X = F_Y$, $\Delta_0 = 0$



(a) GE_1 : Size

(b) GE_1 : Power

Figure 2.9: Design II(a) – Size and power of delta and Fieller-type tests for GE_1 comparisons

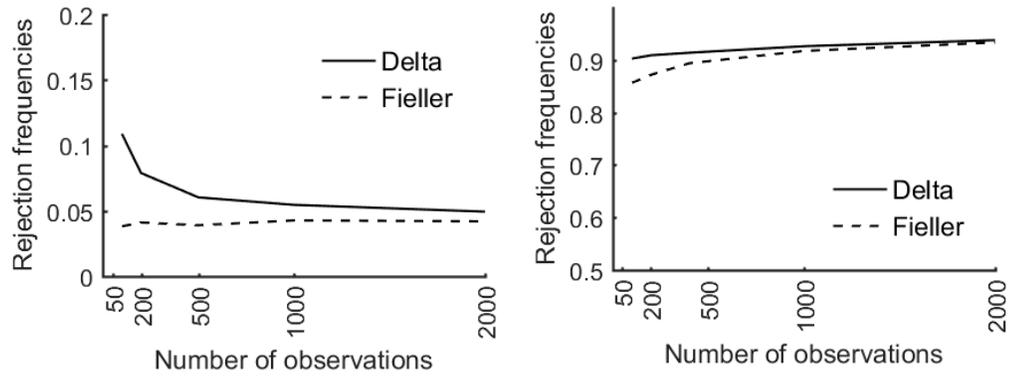
$$H_0: GE_1(X) = GE_1(Y)$$

Left panel DGP: $SM_X(a_X = 5.8, q_X = 0.499616)$, $SM_Y(a_Y = 5.8, q_Y = 0.499616)$.

$$GE_1(X) = GE_1(Y) = 0.14011$$

Right panel: $SM_X(a_X = 4.8, q_X = 0.499616)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$.

$$GE_1(X) = 0.22857, GE_1(Y) = 0.09514$$



(a) GE_2 : Size

(b) GE_2 : Power

Figure 2.10: Design II(a) – Size and power of delta and Fieller-type tests for GE_2 comparisons

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

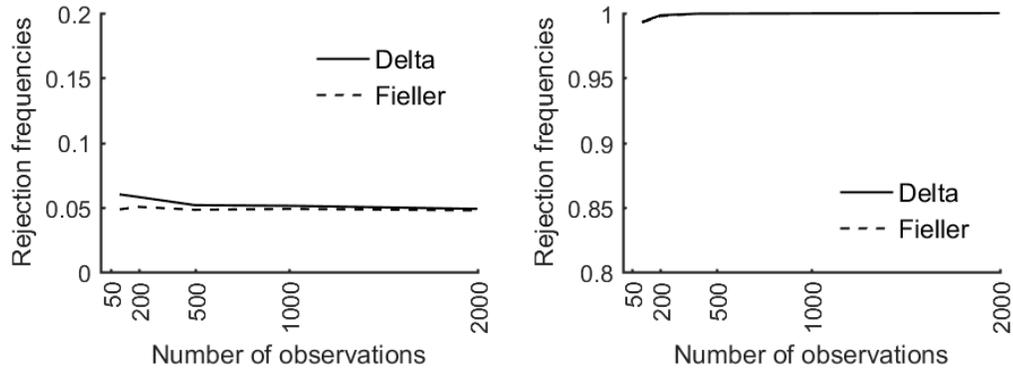
Left panel: $SM_X(a_X = 5.8, q_X = 0.499616)$, $SM_Y(a_Y = 5.8, q_Y = 0.499616)$.

$$GE_2(X) = GE_2(Y) = 0.24396$$

Right panel: $SM_X(a_X = 4.8, q_X = 0.499616)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$.

$$GE_2(X) = 0.63704, GE_2(Y) = 0.13805$$

Design II(b) – Independent samples: $n = 2m$, $F_X \neq F_Y$, $\Delta_0 = 0$



(a) GE_1 : Size

(b) GE_1 : Power

Figure 2.11: Design II(b) – Size and power of delta and Fieller-type tests for GE_1 comparisons

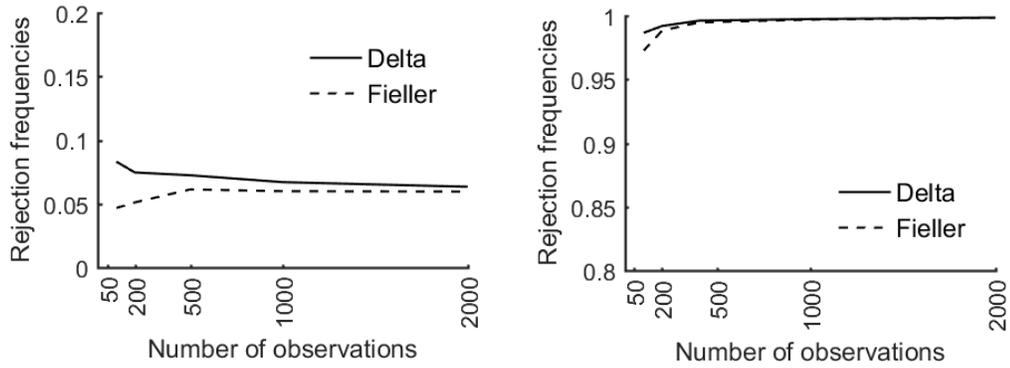
$$H_0: GE_1(X) = GE_1(Y)$$

Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 2.7, q_Y = 1.894309)$. $GE_1(X) =$

$$GE_1(Y) = 0.14011$$

Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 3.7, q_Y = 1.894309)$.

$$GE_1(X) = 0.33829, GE_1(Y) = 0.07637$$



(a) GE_2 : Size

(b) GE_2 : Power

Figure 2.12: Design II(b) – Size and power of Delta and Fieller-type tests for GE_2 comparisons

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

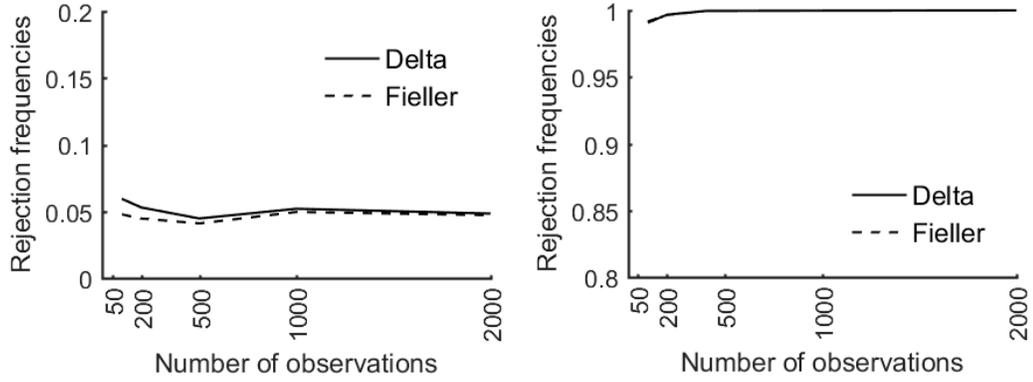
Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 3, q_Y = 1.4684)$.

$$GE_2(X) = GE_2(Y) = 0.16203$$

Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4, q_Y = 1.4684)$.

$$GE_2(X) = 0.54790, GE_2(Y) = 0.08343$$

Design II(c) – Independent samples: $n = 2m$, $F_X \neq F_Y$, $\Delta_0 \neq 0$



(a) GE_1 : Size

(b) GE_1 : Power

Figure 2.13: Design II(c) – Size and power of Delta and Fieller-type tests for GE_1 comparisons

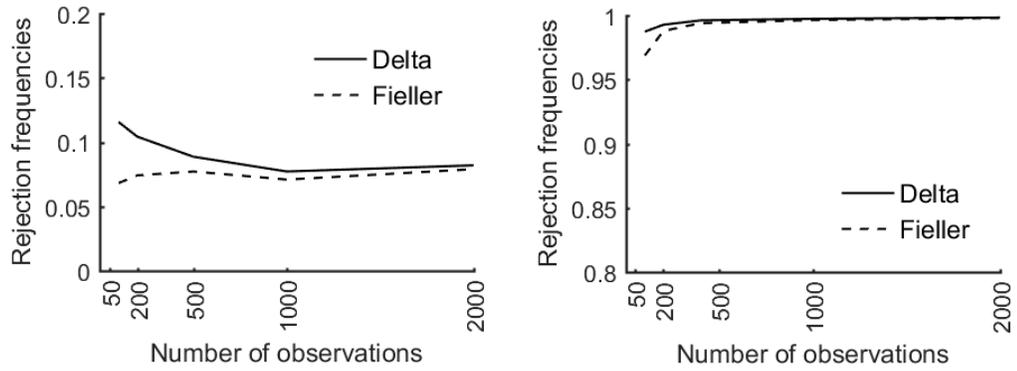
$$H_0: GE_1(X) - GE_1(Y) = 0.04671$$

Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 3.8, q_Y = 1.3061)$.

$$GE_1(X) = 0.14011, GE_1(Y) = 0.0934$$

Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4.8, q_Y = 1.3061)$.

$$GE_1(X) = 0.33829, GE_1(Y) = 0.05839$$



(a) GE_2 : Size

(b) GE_2 : Power

Figure 2.14: Design II(c) – Size and power of Delta and Fieller-type tests for GE_2 comparisons

$$H_0: GE_2(X) - GE_2(Y) = -0.0405$$

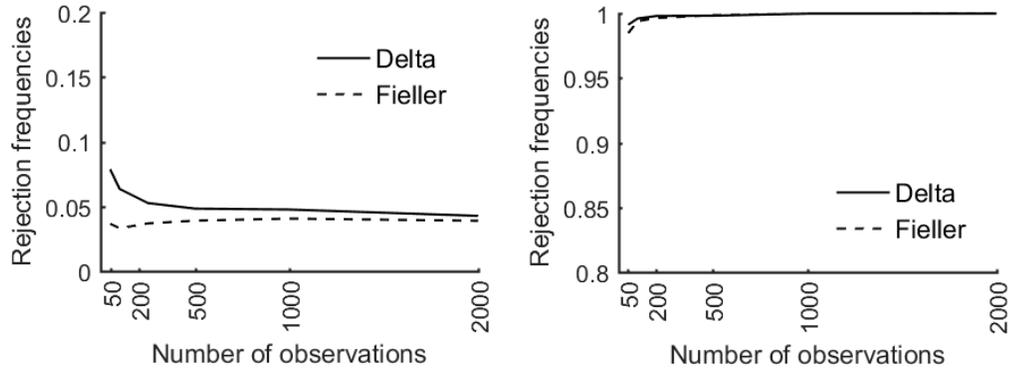
Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 2.5, q_Y = 1.778)$.

$$GE_2(X) = 0.16203, GE_2(Y) = 0.20253$$

Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 3.5, q_Y = 1.778)$.

$$GE_2(X) = 0.54790, GE_2(Y) = 0.09448$$

Design III(a) – Dependent samples: $n = m$, $F_X = F_Y$, $\Delta_0 = 0$



(a) GE_1 : Size

(b) GE_1 : Power

Figure 2.15: Design III(a) – Size and power of delta and Fieller-type tests for GE_1 comparisons

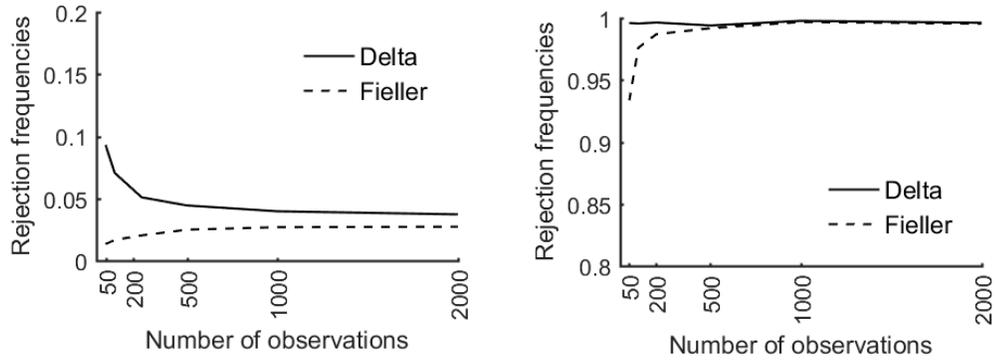
$$H_0: GE_1(X) = GE_1(Y)$$

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

$$GE_1(X) = GE_1(Y) = 0.14011$$

Right panel: $SM_X(a_X = 4.8, q_X = 0.499616)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$.

$$GE_1(X) = 0.22857, GE_1(Y) = 0.09514$$



(a) GE_2 : Size

(b) GE_2 : Power

Figure 2.16: Design III(a) – Size and power of delta and Fieller-type tests for GE_2 comparisons

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

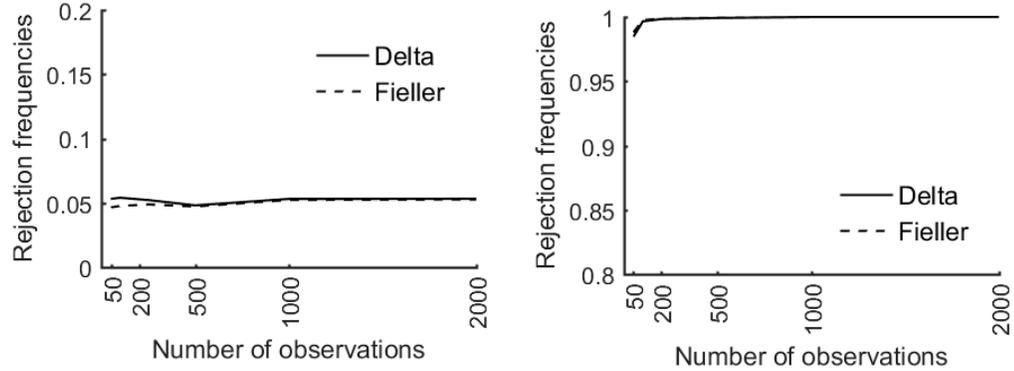
Left panel: $SM_X(a_X = 5.8, q_X = 0.499616)$, $SM_Y(a_Y = 5.8, q_Y = 0.499616)$.

$$GE_2(X) = GE_2(Y) = 0.24396$$

Right panel: $SM_X(a_X = 4.8, q_X = 0.499616)$, $SM_Y(a_Y = 6.8, q_Y = 0.499616)$.

$$GE_2(X) = 0.63704, GE_2(Y) = 0.13805$$

Design III(b) – Dependent samples: $n = m$, $F_X \neq F_Y$, $\Delta_0 = 0$



(a) GE_1 : Size

(b) GE_1 : Power

Figure 2.17: Design III(b) – Size and power of delta and Fieller-type tests for GE_1 comparisons

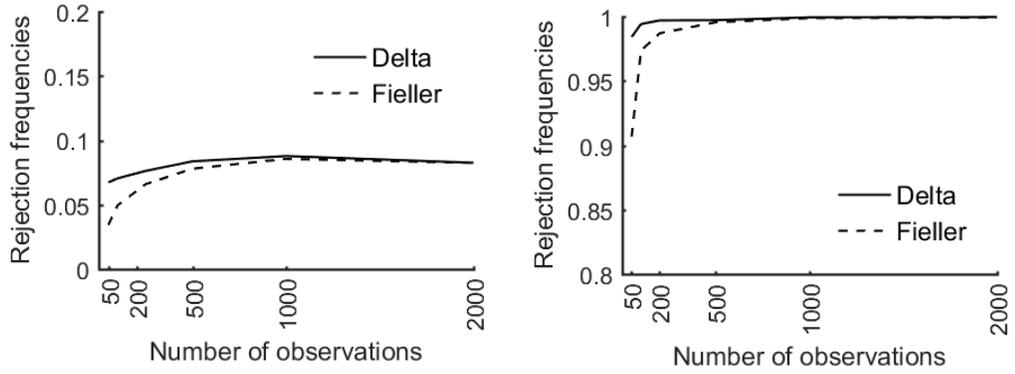
$$H_0: GE_1(X) = GE_1(Y)$$

Left panel: $SM_X(a_X = 2.8, q_X = 1.7)$, $SM_Y(a_Y = 2.7, q_Y = 1.894309)$.

$$GE_1(X) = GE_1(Y) = 0.14011$$

Right panel: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 2.7, q_Y = 1.894309)$.

$$GE_1(X) = 0.33829, GE_1(Y) = 0.14012$$



(a) GE_2 : Size

(b) GE_2 : Power

Figure 2.18: Design III(b) – Size and power of delta and Fieller-type tests for GE_2 comparisons

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

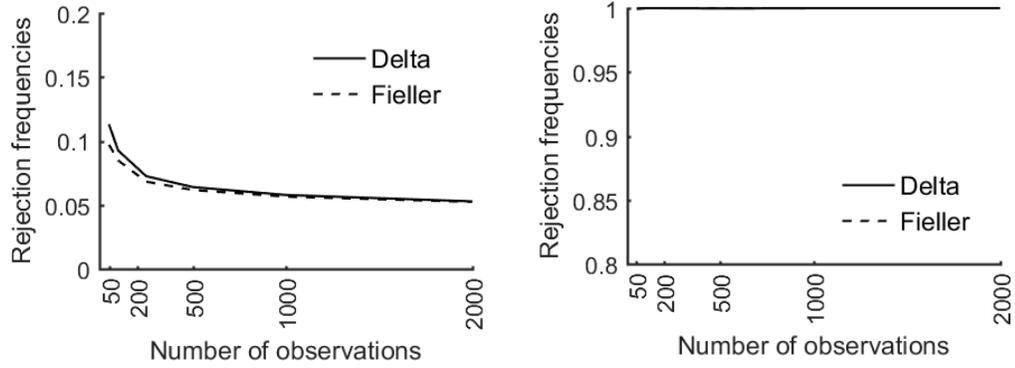
Left panel: $SM_X(a_X = 2.8, q_X = 1.7), SM_Y(a_Y = 3, q_Y = 1.4684)$.

$$GE_2(X) = GE_2(Y) = 0.16203$$

Right panel: $SM_X(a_X = 1.8, q_X = 1.7), SM_Y(a_Y = 3, q_Y = 1.4684)$.

$$GE_2(X) = 0.54790, GE_2(Y) = 0.16204$$

Design III(c) – Dependent samples: $n = m$, $F_X \neq F_Y$, $\Delta_0 \neq 0$



(a) GE_1 : Size

(b) GE_1 : Power

Figure 2.19: Design III(c) – Size and power of delta and Fieller-type tests for GE_1 comparisons

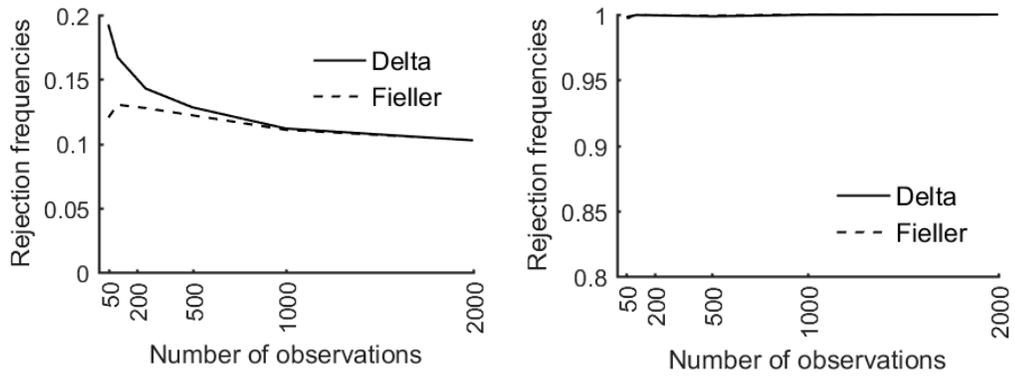
$$H_0: GE_1(X) - GE_1(Y) = 0.046715$$

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: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4.8, q_Y = 1.3061)$.

$$GE_1(X) = 0.33829, GE_1(Y) = 0.05839$$



(a) GE_2 : Size

(b) GE_2 : Power

Figure 2.20: Design III(c) – Size and power of delta and Fieller-type tests for GE_2 comparisons

$$H_0: GE_2(X) - GE_2(Y) = 0.054001$$

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: $SM_X(a_X = 1.8, q_X = 1.7)$, $SM_Y(a_Y = 4.8, q_Y = 1.2855)$.

$$GE_2(X) = 0.54790, GE_2(Y) = 0.06367$$

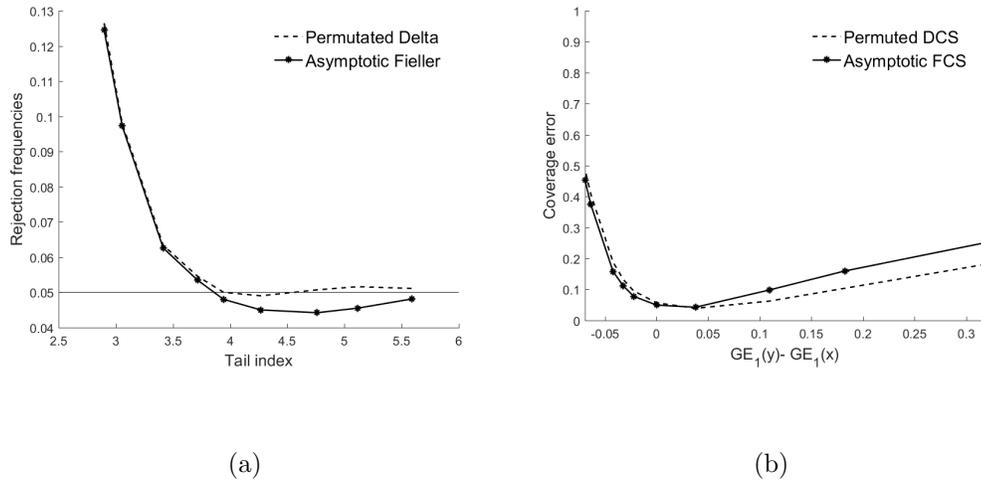


Figure 2.21: Size and Power of two-sample tests

Rejection frequencies of asymptotic Fieller-type and permutated 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 Permutated 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 $q_Y = 10$.

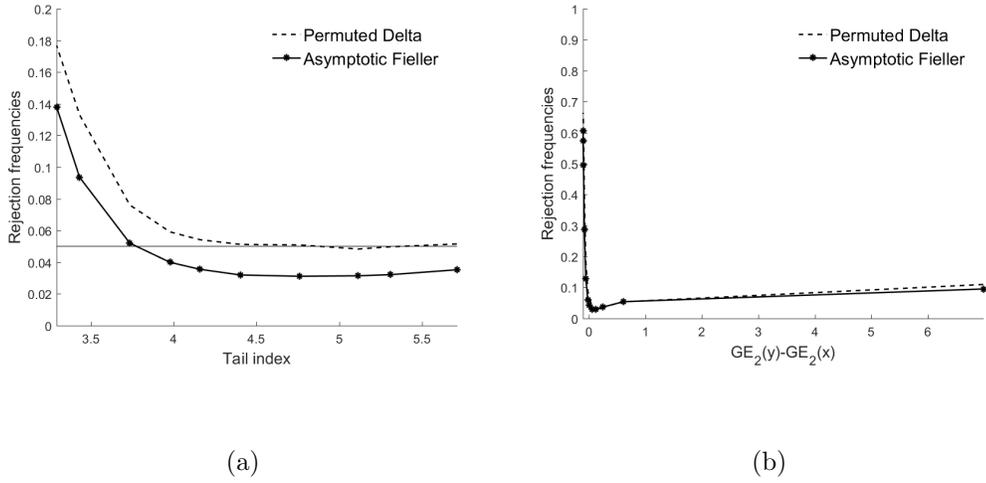


Figure 2.22: Size and Power of two-sample tests

Rejection frequencies of asymptotic Fieller-type and permuted delta method

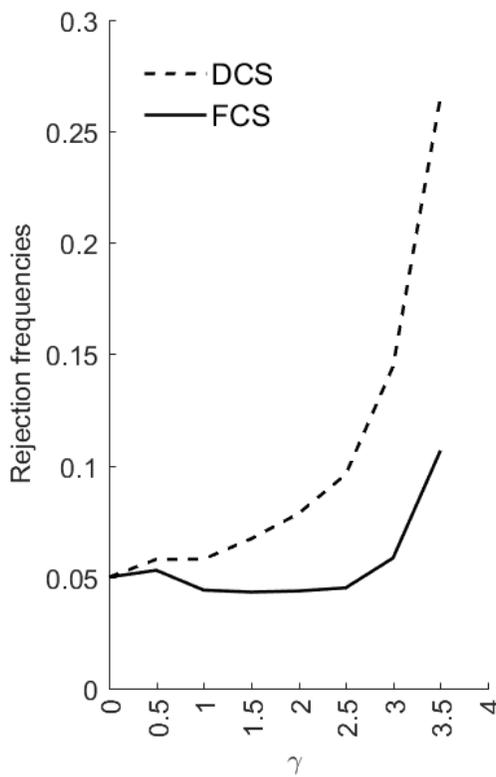
Note – Samples are independent and $n = m$. $F_X = F_Y$ and $GE_\gamma(X) = GE_\gamma(Y)$. The left

panel pertains to the size analysis and it plots the Rejection frequencies of asymptotic

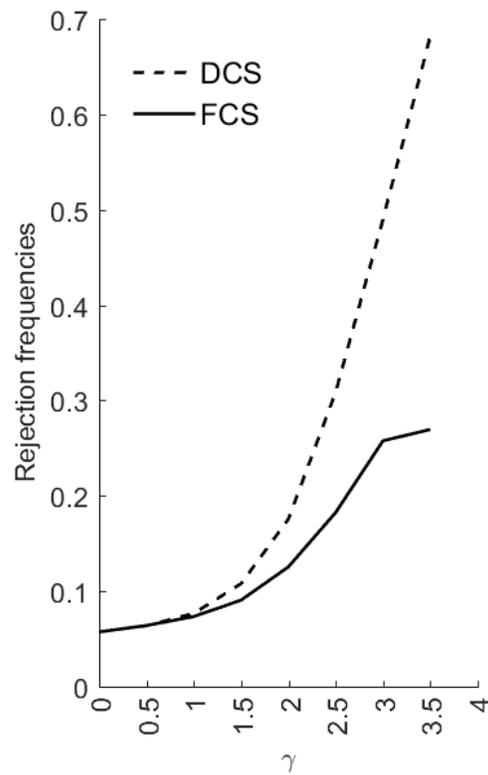
Fieller-type and Permuted delta methods against the tail index: $\xi = [2.897, 6.256]$. Power

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 $q_2 = 10$.



(a) Independent Samples

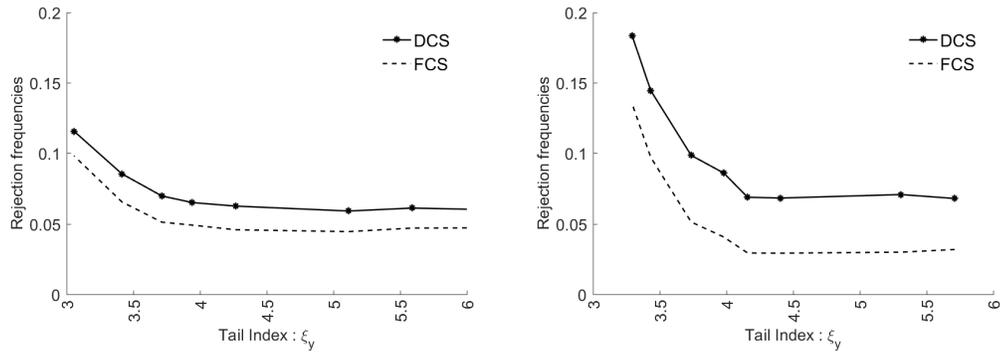


(b) Dependent Samples

Figure 2.23: 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:

$$SM_X(a_X = 2.8, q_X = 1.7) \text{ and } SM_Y(a_Y = 2.8, q_Y = 2). \quad n = m = 50$$



(a) GE_1

(b) GE_2

Figure 2.24: 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

$$\xi_Y = [3.293, 5.7107]. \quad n = m = 50$$

Tables

Table 2.1: 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 2.2: 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 2.3: Rejection frequencies of Delta and Fieller methods: effect of left-tail thickness in the two sample problem; $n = 50$.

α_X	q_X	α_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 α_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 $\Delta_{0,2}$ are given in columns 10 and 11 respectively.

Table 2.4: 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 $1 - (\text{Rejection probability})$. The results in this table pertains to the same case in the left panel of Figure 2.2: GE_2 index with $SM_X(a_X = 1.1, q_X = 4.327273)$. $H_0: GE_2 = 0.71577$

Table 2.5: 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 $1 - (\text{Rejection probability})$. The results in this table pertain to the same case in the left panel of Figure 2.20: 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 2.6: 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 p - Value	0.014 Inequality decreases	0.014 Inequality decreases
Number of states	48	48

Table 2.7: 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 p – value	0.886	0.992
Number of countries	72	72

Chapter 3

Reliable non-parametric inference for quantile ratios

3.1 Introduction

The first two chapters of the thesis focus on moment-based inequality measures. In this chapter we focus on alternative class of measures based on quantiles, in particular, the quantile ratio index which is defined as the ratio of two quantiles of the same distribution such as the top 10 percentile to the bottom 10 percentile of the underlying income distribution (reads as 90/10 percentile ratio or simply 90/10 ratio). More intuitively, the 90/10 ratio for instance, compares the lowest income received by the richest 10 percent of the population to the highest income received by the poorest 10 percent of the same population.

Reliable methods for quantifying precision and uncertainty for the quantile ratio index are lacking in the literature, which motivates this work. In particular, the

chapter proposes three inference methods on the quantile ratio index and assesses their finite sample properties (level and power) via a set of simulation experiments. The methods include the standard Fieller confidence sets (FCS) and bootstrap based alternative Wald and Fieller procedures (WCS-B and FTT-B respectively) that circumvent the main limitation of quantile methods, namely the fact that variance estimates depend on the unknown density. Special attention is given to simulation designs involving ratios with extreme quantiles and heavy-tailed null distributions.

As a matter of fact, thick tails invalidate asymptotic and standard bootstrap inference methods. To circumvent these problems, a considerable literature is available that proposes improvements mainly on moment-based measures. See e.g., [Biewen \(2002\)](#) on the bootstrap, [Davidson and Flachaire \(2007\)](#) on semi-parametric bootstraps methods, [Dufour et al. \(2019\)](#) on permutation methods, and [Dufour et al. \(2018\)](#) on weak identification and testability. Broadly, improvements remain restricted to special cases and where parametric methods are employed ([Cowell and Victoria-Feser, 1993](#)).

In addition to asymptotic complications, thick tails cause serious non-testability problems when moments are involved. In particular, the so called Bahadur-Savage impossibility ([Bahadur and Savage, 1956](#); [Dufour, 2003](#); [Dufour et al., 2008](#); [Coudin and Dufour, 2009, 2017](#); [Dufour, 1997](#); [Gleser and Hwang, 1987](#); [Bertanha and Moreira, 2016](#)) implies that the power of a testing procedure cannot exceed its level. The Bahadur-Savage critique refers to hypothesis tests for which rejection probabilities when the null hypothesis does not hold cannot exceed the null rejection probabilities. This arises when (any) alternative is very hard to distinguish from the null hypothe-

sis, and/or under weak identification. Bahadur and Savage (1956) raised this problem for inference on means non-parametrically, that is, without any further bound on the parameter space nor restriction on the considered family of distributions. In the absence of any such constraints, it is not difficult to find a model within the alternative set that is practically observationally equivalent to the null set. Until the more recent literature on weak-identification (reviewed in e.g. Dufour (1997), this problem has received limited practical attention in econometrics, except perhaps among proponents of quantile-based methods (examples of which that are relevant for this Chapter are reviewed below).

Most popular inequality measures such as the Gini coefficient, the Theil index, the Atkinson family of inequality indices are non-linear functions of moments and are thus vulnerable to this critique. Unlike parameters based on moments, quantiles are testable even in a fully non-parametric framework (Coudin and Dufour, 2009, 2017). Yet the main difficulty with quantiles for inference purpose rests in their studentization. Available formulas for the variance of quantile estimates depend on the underlying (unknown) density, typical estimates of which require smoothing (Hall and Sheather, 1988; Falk, 1986; Jones, 1992). Resulting complications arise from the choice of kernel and the selection of (optimal) bandwidths (Jones et al., 1996). These problem are well documented and motivate the proposed solution in this chapter that eschews estimation of densities.

In this context, this chapter has three contributions. First, it shows that standard methods for correcting inference on ratios fail to deliver expected improvements when using quantiles. In particular, the standard Fieller correction using density-based es-

estimates of variances and covariances is as distorted as its Wald-type counterpart. This stands in sharp contrast with the ratio of moments case, where important improvements were shown over the Wald-type method in the first two chapters of this thesis. This suggests that estimation of variances and covariances of quantiles is a more compelling problem than the discontinuity of the ratio. Second, I propose alternative Wald-type statistics that avoid density estimation by bootstrapping the ratio transformation, so that its variance can be directly obtained by simulation. The problematic variance formulas for quantiles are thus avoided altogether. The Wald-type statistics in question are: (i) a standard t-statistic studentized by simulation; (ii) a Fieller-type alternative studentized under the null hypothesis, again by simulation. Inverting the former leads to confidence intervals of the usual form, whereas inverting the latter needs to be conducted numerically, since its denominator varies with the tested value of the ratio.

The third contribution of this chapter is to show that both statistics work equally well despite the presence of heavy tails and with extreme quantiles. This reinforces the conclusion drawn from the first exercise, namely on the importance of avoiding estimating densities relative to the discontinuity of the ratio. The former problem matters much more with quantiles whereas the latter severely impedes working with moments. In fact, the proposed studentization for quantiles fails completely when used with moments (an exercise I carefully verified). Robustness alleviates the discontinuity problem, in which case scale-invariance also improves the performance of resampling based studentization. Estimating the variance of each quantile (and their covariance) by bootstrapping is known to fail (which is also duly verified in this con-

text). In contrast, results show that directly bootstrapping the ratio instead seems a tractable and reliable solution. Existing work on quantile ratios is scarce. A few broadly related results are available, for example on comparing two quantiles of two distinct distributions (Mood et al., 1954; Wilcox and Charlin, 1986; Hettmansperger, 1973; Hettmansperger and Malin, 1975; Cheng and Wu, 2010; Dominici and Zeger, 2005). Originally, work in this area focused on comparing the medians as an alternative to comparing the means (Wilcox and Charlin (1986)). To the best of my knowledge, the only contribution on the quantile ratio index is by Prendergast and Staudte (2017). The authors apply the Delta method to compute the variance of the ratio for which they suggest various kernel estimators. However, their method is sensitive to bandwidth choices in important cases, and their optimal suggestions are restricted to parametric families which is counter-intuitive in non-parametric perspective. Their study covers only quantiles that add up to one (such as 95/5 or 80/20 ratios) and seems to exclude ratios with extreme quantiles, which is restrictive for inequality work.

In this chapter, the proposed methods are fully non-parametric and analysis covers extreme quantiles, quantiles that do not add up to one, and extremely heavy-tailed null distributions. Overall, results can be summarized as follows. For nearly all designs considered, the standard Wald-type confidence sets as well as their Fieller-based counterparts have levels that deviate arbitrarily away from their nominal ones, as well as very low power. Instead, replacing the asymptotic variance which requires a density estimate with a bootstrap alternative applied directly to the ratio seems to solve the problem. Despite existing works that discourage reliance on bootstrap

variance estimates of quantiles, applying the bootstrap directly to the ratio seems to perform well. The same approach does not work when ratios of moments are considered. This suggests that robustness in conjunction with scale-invariance jointly justify the success of the proposed methodology. The rest of the chapter is organized as follows: section two presents the standard Wald-type confidence interval and the three proposed methods. In section three, the simulation results are reported and the last section concludes.

3.2 Methodology

Given the complex nature of the underlying income distributions and the functional forms of inequality measures, standard inference methods might fail leading to flawed conclusions. Three related issues cause such distortions. First, underlying distributions are heavy tailed which causes slow convergence. Second, many popular measures are based on moments thus subject to the Bahadur-Savage problem. Third, most inequality measures are non-linear transformations of parameters (in particular, ratios) raising identification concerns. These problems combined cause severe identification complications.

A consequence of the above defined Bahadur-Savage problem, testability concerns are more prevalent with moments-based inequality indices. Moreover, non-parametric confidence set estimation methods should permit unbounded outcomes. Asymptotic confidence sets based on the Delta method have bounded limits and thus they are unreliable.

Unlike parameters based on moments, quantiles are testable even in a fully non-parametric framework. In fact, one of the most appealing features of the quantile-based measures is that they are insensitive to the tails of the underlying distributions. Thus, these measures are not affected by the presence of extreme values and outliers or any sort of contamination taking place at either ends of the distribution. This suggests that quantile ratios are less susceptible to non-testability problems than their moments-based counterparts. Yet robustness does not shield these measures from other complications. The main difficulty is that the variance of quantile estimates used to build the standard asymptotic Wald-type confidence sets, depends on the underlying density as described below.

Denote by X the random variable representing income with a set of independent observations x_1, \dots, x_n from the distribution function $F_X(x)$. Denote by $\{x_{(1)}, \dots, x_{(n)}\}$ the order statistic of $\{x_1, \dots, x_n\}$. The quantile function is defined as $Q_w = F_X^{-1}(w) = \inf\{x : F_X(x) \geq w\}$ with $0 < w < 1$. Numerous definitions have been proposed for estimating the sample quantile \hat{Q}_w and are conveniently summarized and compared by [Langford \(2006\)](#). In what follows, \hat{Q}_w is obtained using the fourth definition in [Langford \(2006\)](#).

Denote by $\hat{\rho}$ and ρ_0 the estimated and the true value of the quantile ratio ρ defined by:

$$\rho = \frac{Q_u}{Q_l}, \quad 0 < l < u < 1. \quad (3.2.1)$$

The null hypothesis is defined by:

$$H_0^W : \rho = \rho_0. \quad (3.2.2)$$

3.2.1 Method 1 - Wald-type Confidence Sets (WSC)

The standard Wald-type asymptotic confidence set based on the Delta method (WCS) can be constructed by inverting the square or the absolute value of the standard Wald-type test statistic. Test inversion refers to the collection of the parameter's values for which we fail to reject the null hypothesis, and is typically carried out by solving the following inequality:

$$T_W^2 = \frac{(\hat{\rho} - \rho_0)^2}{\hat{V}_W(\hat{\rho})} \leq z_{\alpha/2}^2, \quad (3.2.3)$$

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 $\hat{V}_W(\hat{\rho})$ is the estimate of the quantile ratio variance denoted by $V_W(\rho)$ and which can be derived using the Delta method:

$$V_W(\rho) = \left(\frac{1}{Q_l}\right)^2 \sigma_u^2 + \left(\frac{Q_u}{Q_l}\right)^2 \sigma_l^2 - 2 \left(\frac{1}{Q_l}\right) \left(\frac{Q_u}{Q_l^2}\right) \sigma_{u,l}, \quad (3.2.4)$$

with σ_u^2 , σ_l^2 , and $\sigma_{u,l}$ are the variance of Q_u , the variance of Q_l and the covariance between Q_U and Q_l respectively. The standard definitions of these statistics are given by:

$$\begin{cases} \sigma_u^2 = [u(1-u)q_u^2]/n \\ \sigma_l^2 = [l(1-l)q_l^2]/n \\ \sigma_{u,l}^2 = [u(1-l)q_u q_l]/n, \end{cases} \quad (3.2.5)$$

where q_u and q_l are the quantile density functions which are equal to the reciprocal of the density function evaluated at their respective quantiles: $q_u = 1/f(Q_u)$ and $q_l = 1/f(Q_l)$. Thus, the estimation of the variance in (3.2.4) requires the estimation of the density function involving kernel methods and bandwidth selection. For information on estimation methods for sample quantile and their variances see [Hyndman and Fan \(1996\)](#); [Langford \(2006\)](#); [Hall and Sheather \(1988\)](#); [Falk \(1986\)](#); [Jones \(1992\)](#); [Dominici et al. \(2005\)](#).

Typical density estimates require smoothing which involves the choice of the optimal kernel function and the bandwidth. Sensitivity of choice of kernel is an enduring issue. In addition, optimal bandwidth selection methods are validated via subsumed distributional restrictions. When these lack fit, serious problems may arise with density estimates and the estimates of variances that rely on them. [Prendergast and Staudte \(2017\)](#) suggest estimating the quantile density function using kernels as follows:

$$\widehat{q}_w = \sum_{i=1}^n x(i) \left\{ \kappa_b \left(w - \frac{(i-1)}{n} \right) - \kappa_b \left(w - \frac{i}{n} \right) \right\} \quad (3.2.6)$$

where $\kappa_b(\cdot) = \kappa(\cdot/b)/b$ with $\kappa(\cdot)$ is the Epanechnikov function and b the bandwidth. The authors rely on the Quantile Optimality Ratio (QOR) to choose the optimal bandwidth for the underlying kernel estimator. The QOR is defined as the ratio of the

quantile density function to its second derivative: $QOR(w) = q_w/q_w''$. [Prendergast and Staudte \(2017\)](#) consider three distributions (log-normal, Pareto of the second type, and Chi-square) and derive their quantile optimality ratio. However, the empirical literature on income inequality suggests that heavier-tailed distributions better fit observed income distributions. Popular distributions in this category include the generalized Beta of the second type (GB2) and two of the distributions it nests: the Dagum distribution and the Sing-Maddala distribution. For the derivation of the QOR for the log-normal, Pareto II, and Chi-square distributions see the appendix of [Prendergast and Staudte \(2016\)](#). For the derivation of the QOR for the GB2 distribution see the appendix below.

Once the quantile densities are estimated, estimation of (3.2.4) follows straightforwardly leading to the $1 - \alpha$ confidence interval:

$$WCS(\rho; 1 - \alpha) = \left[\hat{\rho} \pm z_{\alpha/2} [\hat{V}_W(\hat{\rho})]^{1/2} \right]. \quad (3.2.7)$$

The problems surrounding the inference on ratios and the quantile variance estimation are well documented in the literature ([Dufour, 1997](#); [Von Luxburg and Franz, 2009](#); [Hall et al., 1989](#); [Hall and Sheather, 1988](#); [Prendergast and Staudte, 2016](#)) and motivate the proposed solutions in this chapter that eschew estimation of densities and/or are robust to ratio-induced identification issues.

3.2.2 Method 2- Density-free Bootstrap Wald Confidence Sets (WCS-B)

An appealing alternative for the standard estimate of quantile variance defined by (3.2.4) is to estimate the variance using resampling techniques such as bootstrap techniques and the leave-one-out Jackknife estimator. However, the Jackknife variance for quantiles was shown to be inconsistent Efron (1992) and the studentized t-statistic using the Jackknife variance is not asymptotically normal and has an infinite mean (Martin, 1990).

In the same vein, bootstrap variance estimates for quantile were also shown to suffer from slow convergence (Hall and Sheather, 1988; Hall et al., 1989).

Instead, I propose estimating the quantile variance by bootstrapping the ratio directly which avoids the problematic variance formula in (3.2.4). The associated Wald-type statistic is thus studentized by simulation which upon inversion leads to the usual confidence sets. The motivation behind bootstrapping the ratio directly rather than bootstrapping the individual quantiles is that the ratio is scale invariant whereas quantiles are not. It is well known that scale raises important concerns for inference Dagenais and Dufour (1991, 1992); Dufour and Dagenais (1992). Thus a scale invariant transformation will be immune to this problem. The following algorithm describes the proposed procedure:

- i. Draw B bootstrap samples of size n with replacement $\{(x_1^{(1)}, \dots, x_n^{(1)}), \dots, ((x_1^{(B)}, \dots, x_n^{(B)}))\}$

- ii. For each sample, estimate the quantile ratio. This will yield a sequence of B bootstrap estimates of the quantile ratio $(\hat{\rho}^{(1)}, \dots, \hat{\rho}^{(B)})$
- iii. Use the standard sample variance formula to estimate the variance of the sequence produced in (ii):

$$\hat{V}_{W-B}(\hat{\rho}) = \frac{\sum_{i=1}^B (\hat{\rho}^{(i)} - \bar{\hat{\rho}})^2}{B-1}, \quad \text{with} \quad \bar{\hat{\rho}} = \frac{\sum_{i=1}^B \hat{\rho}^{(i)}}{B}. \quad (3.2.8)$$

- iv. Replace the variance in (3.2.3) with bootstrap variance estimate defined in (3.2.8). This leads to the usual Wald-type statistic but standardized with the bootstrap variance rather than the asymptotic one:

$$T_{W-B}^2 = \frac{(\hat{\rho} - \rho_o)^2}{\hat{V}_{W-B}(\hat{\rho})} \leq z_{\alpha/2}^2, \quad (3.2.9)$$

- v. Inverting T_{W-B}^2 gives the WCS-B confidence interval:

$$\text{WCS-B}(\rho; 1 - \alpha) = \left[\hat{\rho} \pm z_{\alpha/2} [\hat{V}_{W-B}(\hat{\rho})]^{1/2} \right]. \quad (3.2.10)$$

To the best of my knowledge, this method has not been considered in the literature.

3.2.3 Method 3 - Fieller's Confidence Sets (FCS)

Fieller's method is an asymptotic inference method proposed for ratios and is most advantageous over the standard Wald-type method when the ratio suffers from identification problem (Fieller, 1940, 1944, 1954; Bolduc et al., 2010; Von Luxburg and

Franz, 2004). Related literature was reviewed in the previous two chapters, where its usefulness for the general entropy family of indices was demonstrated.

Nevertheless, Fieller's method requires that each of the numerator and the denominator of the ratio as well as their variances and the covariance to be well-estimated. This leads us to question the usefulness of this method in the context of quantile ratios. In view of its success in the context of moments, it is still worth a formal investigation.

The method rewrites the null hypothesis defined in (3.2.2) in the linear form:

$$H_0^F : \theta(\rho_0) = Q_u - Q_l \rho_0 = 0 \quad (3.2.11)$$

and then it proceeds by inverting the square of the associated Wald-type statistic based on this reformulated linear hypothesis:

$$T_F^2 = \frac{\hat{\theta}(\rho_0)^2}{V_F(\hat{\theta}(\rho_0))} \leq z_{\alpha/2}^2, \quad (3.2.12)$$

with $\hat{\theta}(\rho_0) = \hat{Q}_u - \hat{Q}_l \rho_0$ and $V_F(\hat{\theta}(\rho_0))$:

$$V_F(\hat{\theta}(\rho_0)) = \sigma_u^2 + (-\rho_0)^2 \sigma_l^2 - 2\rho_0 \sigma_{u,l} \quad (3.2.13)$$

where σ_u^2 , σ_l^2 , and $\sigma_{u,l}$ are estimated using the definitions in (3.2.5). Note that since $\hat{\theta}^2$ and $V_F(\hat{\theta}(\rho_0))$ are both quadratic in ρ_0 , the inversion problem is a problem of solving a quadratic inequality of the form $A\rho_0^2 + B\rho_0 + C \leq 0$. After some algebra, the following forms can be derived for the parameters A , B , and C :

$$\begin{cases} A = Q_l^2 - Z_{\alpha/2}^2 \sigma_l^2 \\ B = 2(Z_{\alpha/2}^2 \sigma_{u,l} - Q_u Q_l) \\ C = Q_u^2 - Z_{\alpha/2}^2 \sigma_u^2, \end{cases} \quad (3.2.14)$$

On setting $D = B^2 - 4AC$, the $(1 - \alpha)$ -level Fieller-type confidence set is characterized as follows:

1. if $D > 0$ and $A > 0$, $\text{FC}(I; 1 - \alpha) = \left[\frac{-B - \sqrt{D}}{2A}, \frac{-B + \sqrt{D}}{2A} \right]$;
2. if $D > 0$ and $A < 0$, $\text{FC}(I; 1 - \alpha) = \left] -\infty, \frac{-B + \sqrt{D}}{2A} \right] \cup \left[\frac{-B - \sqrt{D}}{2A}, +\infty \right[$;
3. if $D < 0$, $A < 0$ and $\text{FC}(I; 1 - \alpha) = \mathbb{R}$.

3.2.4 Method 4 - Density-free Bootstrap Fieller-type method

FTT-B

As mentioned above, Fieller's method requires that each of the numerator and the denominator of the ratio as well as their variances and the covariance to be well-estimated. I thus propose to combine the Fieller principle with the resampling studentization as in section (3.2.2). Formally, this entails replacing the usual variance in (3.2.12) by a bootstrap-based estimator of the linearized transform $\theta(\rho_0)$ under the null. This method is expected to simultaneously circumvent problems associated with density estimation and ratios. The following algorithm outlines the steps of estimating FTT-B:

- i. Draw B bootstrap samples of size n with replacement $\{(x_1^{(1)}, \dots, x_n^{(1)}), \dots, (x_1^{(B)}, \dots,$

$x_n^{(B)}\}$

- ii. For each sample, estimate Q_u and Q_l and plug them in the definition of $\theta = Q_u - Q_l\rho_0$. Note that the true quantile ratio ρ_0 is used here and not an estimate of it. In this sense, this is a bootstrap under the null. Repeating the estimation of θ for every bootstrap sample yields a sequence of B bootstrap estimates $(\hat{\theta}^{(1)}(\rho_0), \dots, \hat{\theta}^{(B)}(\rho_0))$.
- iii. Use the standard sample variance formula to estimate the variance of the sequence produced in (ii):

$$\hat{V}_{F-B}(\hat{\theta}(\rho_0)) = \frac{\sum_{i=1}^B (\hat{\theta}^{(i)}(\rho_0) - \bar{\hat{\theta}}^2(\rho_0))^2}{B-1}, \quad \text{with} \quad \bar{\hat{\theta}}(\rho_0) = \frac{\sum_{i=1}^B \hat{\theta}^{(i)}(\rho_0)}{B}. \quad (3.2.15)$$

- iv. Replace the variance in (3.2.12) with bootstrap variance estimate defined in (3.2.8). This leads to the usual Wald-type statistic but standardized with the bootstrap variance rather than the asymptotic one:

$$T_{F-B}^2 = \frac{\hat{\theta}^2}{V_{F-B}(\hat{\theta}(\rho_0))}. \quad (3.2.16)$$

Inversion needs to be performed numerically, since the studentization depends on ρ_0 .

3.3 Simulation Results

To assess the performance of the proposed methods, several simulation experiments were conducted for both the level and the power analysis. The simulation designs considers the 99/20, the 99/50 and the 80/20 ratios for the level analysis and the 99/20 ratio for the power analysis. The power analysis of the other two ratios yield similar results and is available upon request. The null distributions are the generalized beta of the second type (GB2) and its two special cases: the Singh-Madalla distribution (SM) and the Dagum distribution. These distributions have been used in the empirical work on inequality analysis as they mimic actual income distributions of the developed countries ([McDonald, 2008](#); [Jenkins, 2007](#); [Brachmann et al., 1995](#); [Cowell and Flachaire, 2007](#)).

The simulation experiments are based on 10,000 replications with 199 bootstrap repetitions each. The number of observations varies between 100 and 2000 for the 80/20 ratio and between 500 and 2000 for the other two ratios (99/20 and 99/50). The reason for considering samples of 500 observations or more for the latter two ratios is to ensure that there is enough variation at the top of the distribution when extreme quantiles like the 99th percentile are involved.

Unlike [Prendergast and Staudte \(2017\)](#), the simulation study here considers extreme quantiles such as the 99th percentile. [Prendergast and Staudte \(2017\)](#) reports that the performance of the inference methods worsens with extreme quantiles and thus it is imperative to check how the proposed methods behave in such cases es-

pecially that ratios involving extreme quantiles such as the 99/20 or the 99/50 are empirically relevant.

Another difference between the simulation study in this chapter and that of [Pren-dergast and Staudte \(2017\)](#) is that I assumes more heavy tailed distributions. The main challenges facing inference on inequality analysis arise from the presence of thick tails in the underlying income distributions and a sound simulation design should take that into account by assessing the behavior of the proposed statistics in this context.

Tables (3.2) to (3.3) report the coverage error rate for the 80/20, 99/50 and 99/20 quantile ratios respectively. For every ratio, the results of the standard Wald-type asymptotic confidence interval (DCS) and the three proposed methods: asymptotic Fieller's confidence sets (FCS), asymptotic Wald-type confidence interval using the bootstrap variance (WCS-B), and the asymptotic Fieller-type t-test with bootstrap variance (FTT-B) are reported. Each table reports the coverage error rates/level of these inference methods for different sets of parameters characterizing the three null distributions: Dagum, SM and the GB2.

The GB2 distribution has three shape parameters (a, p, q) and one scale parameter (b) . The parameter a affects both tails while the parameter p affects the left tail and the parameter q affects the right tail. In all simulation experiments, the parameter a is held fixed and the parameters p and q are allowed to vary so as to discern between the performance of the proposed methods when the underlying distribution is characterized by (i) a thick left tail, (ii) a thick right tail, and (iii) a thick left tail and a thick right tail at the same time. The motive for differentiating between the impact of the right tail and the left tail separately comes from [Dufour](#)

et al. (2018) where the authors show that left tail problems might negatively affect the performance of the inference methods as much as the right tail.

A salient results regarding the level analysis are reported in tables (3.2) to (3.3) and can be summarized as follows: First, the standard DCSs deviate arbitrarily from the nominal level as they suffers from overcoverage and undercoverage. Over or under coverage are not induced by a specific choice or shape of the underlying distribution nor by the size of the sample. They persist regardless of the sample size and regardless of the thickness level of the left and/or the right tails. The ill-estimated variances and covariance of the upper and the lower quantiles forming the ratio have a severe negative arbitrary impact on the reliability of the resulting asymptotic confidence intervals.

Consider table 3.2 which reports results for the 80/20 ratio. For sample size of 100, assuming a Dagum(2.8,0.9) distribution, WCS coverage error rate is around 0.035 indicating that the confidence interval suffers from overcoverage, whereas when much thicker left tail Dagum(2.8,0.5) is assumed, the method tends to undercover the true ratio with coverage error rate greater than the 5% level. Same conclusions can be drawn from the results pertaining to a sample size of 500 or 2000.

In the same vein, simulation results relevant to the Singh-Madalla distribution and the GB2 distribution yield similar behavior. Even with large sample sizes and relatively thinner left and right tails the coverage error rate of the WCS is still far from the nominal: the coverage error rate for $n=2000$ assuming a GB2(2.8, 1.7, 0.9) is 0.02, i.e WCS suffers from overcoverage since the coverage probability ($1-0.02=0.98$) is greater than the nominal level of the confidence interval (0.95).

Second, standard asymptotic confidence intervals using the bootstrap variance (WCS-B) perform exceptionally well as they exhibit coverage error rates very close to the nominal level across the board. The results are very stable and they are insensitive to the choice of the underlying distribution and to tails thickness, and perform very well even in samples of small sizes. For instance, for the 80/20 ratio, the WCS-B coverage error rates for $n = 100$ with distributions characterized by very heavy left tail (Dagum(2.8,0.5)) and very heavy right tail is approximately 0.05. Even when both tails are extremely thick as in the case of GB2(2.8,0.75,0.5), the method's coverage error (0.046) is very close to the 5% level.

Third, the asymptotic Fieller's method (FCS) performs poorly across various scenarios with an unpredictable manner in the sense that the undercoverage and the overcoverage patterns cannot be explained by the shape or the choice of the underlying distribution nor by the sample sizes. The other observation about the FCS results is that the level (and the power) results are almost identical to those of the DCS which implies that quantile-based measures are less susceptible than the general entropy measures to identification problems in the designs covered in this chapter.

Fourth, for extreme quantiles, such as the 99/50 or the 99/20 ratios, WCS and FCS performance worsens compared to that of the 80/20 ratio. On the other hand, methods based on bootstrapping the variance (WCS-BV and FTT-B) perform very well where they deviate only marginally from the nominal level for small samples. The small deviation from the nominal level diminishes as the sample size increases and the coverage errors of WCS-B and FTT-B converge back to the nominal level.

Fifth, the discussion so far has focused on size. Yet another important finding on

power also seems to be new to this literature: power is very low in all considered cases unless the proposed estimation for the variance is used. Size distortions for all other methods come along with low power. In contrast, the extent of power improvements (in addition to correct size) with the proposed method is noteworthy. The power of the methods using the asymptotic variance (WCS and FCS) for the 99/20 ratio, as table (3.4) shows, is extremely low and in many cases is even less than the size especially with small sample sizes.

Both Wald-type and Fieller methods require the numerator and denominator to be well estimated, as well as their variances and covariance. In addition, the underlying t-statistics (although different) should converge reasonably fast to the considered limiting distribution (or its bootstrap counterpart when - possibly double - bootstraps are considered). In contrast to the Wald-type method, the Fieller method would, then, correct for the discontinuity entailed by the ratio transformation. Here we see that the expected correction fails, because the variance/covariance estimates of numerator and denominator are unreliable. In contrast, bootstrapping the ratio directly to compute its variance avoids this problem altogether and thus yields a simple and successful correction.

One reason that may explain this finding is that the ratio is scale invariant. In fact, it is well known that if a parameter is poorly identified and/or poorly estimated, some transformation of this parameter could be identifiable. This includes, in particular, transformations that eliminate pernicious nuisance parameters. It is well known that scale raises important concerns for inference. Thus a scale invariant transformation will restore the validity of the bootstrap, and addresses both size and power problems

for non-parametric inference on the ratio.

Bootstrap failures when estimating quantile variances, which I suspect have deterred their application for quantile based related inequality analysis, can be circumvented with scale-invariant transformations such as ratios. This is one of the major findings of this chapter which is new to the literature. Because the need for kernels of any form is eliminated and the Wald analytic form is preserved, the proposed solution remains analytically tractable. The computational simplicity of the proposed solution relative to the seriousness of the problem and the fact that it has escaped notice to date, is noteworthy.

This being said, one nevertheless observes from the results that the Fieller method based on any variance estimate adopted here, provides no improvement over the Wald-type method. This stands in sharp contrast with what is observed when working with moments. To further understand this result, the above bootstrap correction is repeated for ratios of moments. The findings indicate that in contrast to quantile ratios, the correction does not work at all with the GE class of measures, despite the fact that all these measures are also scale invariant. Yet with moments, the Fieller method provides tangible improvement over the Wald-type method (although all over-rejections were not eliminated altogether because slow convergence remains an issue).

The main difference between quantiles and moments is their relative sensitivity to extreme values, which in the case of ratios, directly impacts the discontinuity region and thus identification. In particular, the denominator will be more biased towards zero when means are considered, whereas quantiles can be to some extent bounded

away from zero unless the distribution has a point mass at zero. This may occur e.g. with wage distributions. In the latter case, or when very extreme quantiles are used in the denominator, one would expect the Fieller method to make a difference. Both Fieller and Wald-type method have similar asymptotic properties when the ratio is identified; see Bolduc, Khalaf and Yelou (2010) and the references therein.

3.4 Conclusion

The present chapter proposes three analytically tractable methods for inference on the quantile ratio index and investigates their level and power properties via a set of simulation experiments.

The methods include the standard Fieller and bootstrap-based alternative studentized Wald and Fieller procedures that circumvent complications such as the dependence of the quantile variance on the density function and the potential identification issues induced by the ratio transformation. Results show that the standard asymptotic approach of estimating the quantile variance has detrimental consequences on the performance of both

Wald- and Fieller-type inference procedures. In contrast, the studentization proposed in the chapter which relies on bootstrapping the ratio directly restores coverage and improves power remarkably even with samples of small sizes drawn from extremely heavy-tailed distributions and with ratios involving extreme quantiles. Overall, the Fieller method based on any variance estimate adopted in this chapter, provides no improvement over the Wald-type method, which stands in sharp contrast

with previous works with moments. While the literature on quantiles discourages bootstrapping their variance, the striking successes obtained here nevertheless result from the conjunction of robustness with scale invariance of the ratio.

Appendix A: Simulation tables

Table 3.1: Coverage error of the proposed inference methods for the 99/50 ratio

	a	p	q	WCS	FCS	WCS-B	FTT-B	Width WCS	Width WCS- B
Dagum									
n=500	2.8	0.9	1	0.010	0.009	0.070	0.069	142.922	4.017
	2.8	0.7	1	0.009	0.009	0.075	0.074	157.658	4.219
	2.8	0.5	1	0.006	0.006	0.072	0.069	183.916	4.611
n=2000	2.8	0.9	1	0.031	0.031	0.060	0.060	17.352	1.731
	2.8	0.7	1	0.032	0.032	0.063	0.061	19.097	1.815
	2.8	0.5	1	0.027	0.026	0.065	0.064	21.815	1.985
Singh-Madalla									
n=500	2.8	1	1.7	0.006	0.006	0.070	0.068	48.179	1.451
	2.8	1	1.2	0.008	0.008	0.071	0.069	91.982	2.666
	2.8	1	0.75	0.009	0.009	0.075	0.073	295.365	8.595
n=2000	2.8	1	1.7	0.030	0.029	0.061	0.061	5.785	0.672
	2.8	1	1.2	0.032	0.032	0.067	0.066	10.961	1.181
	2.8	1	0.75	0.034	0.034	0.065	0.064	33.903	3.406
GB2									
n=500	2.8	1.7	0.9	0.007	0.007	0.070	0.068	48.580	1.484
	2.8	1.7	0.7	0.006	0.006	0.068	0.067	54.610	1.604
	2.8	1.7	0.5	0.007	0.007	0.064	0.059	63.577	1.853
	2.8	0.75	0.9	0.007	0.007	0.072	0.071	309.572	8.698
	2.8	0.75	0.7	0.008	0.008	0.068	0.066	345.775	9.062
	2.8	0.75	0.5	0.008	0.008	0.075	0.072	407.316	9.697
n=2000	2.8	1.7	0.9	0.029	0.028	0.064	0.063	6.151	0.692
	2.8	1.7	0.7	0.030	0.030	0.060	0.059	6.552	0.749
	2.8	1.7	0.5	0.028	0.028	0.057	0.057	7.701	0.863
	2.8	0.75	0.9	0.035	0.035	0.069	0.069	36.552	3.436
	2.8	0.75	0.7	0.028	0.027	0.066	0.065	41.102	3.566
	2.8	0.75	0.5	0.027	0.027	0.059	0.058	48.115	3.845

Table 3.2: Coverage error of the proposed inference methods for the 80/20 ratio

	a	p	q	WCS	FCS	WCS-B	FTT-B	Width	Width
								WCS	WCS-B
Dagum									
n=100	2.8	0.9	1	0.035	0.029	0.046	0.044	1.464	1.360
	2.8	0.7	1	0.061	0.057	0.048	0.048	1.662	1.721
	2.8	0.5	1	0.289	0.289	0.052	0.051	1.565	2.634
n=500	2.8	0.9	1	0.031	0.030	0.050	0.050	0.620	0.575
	2.8	0.7	1	0.059	0.059	0.055	0.053	0.694	0.720
	2.8	0.5	1	0.231	0.231	0.048	0.050	0.669	1.092
n=2000	2.8	0.9	1	0.032	0.031	0.051	0.050	0.307	0.283
	2.8	0.7	1	0.052	0.053	0.047	0.047	0.342	0.353
	2.8	0.5	1	0.218	0.218	0.053	0.053	0.334	0.534
Singh-Madalla									
n=100	2.8	1	1.7	0.029	0.025	0.046	0.047	1.050	0.951
	2.8	1	1.2	0.024	0.019	0.046	0.049	1.286	1.116
	2.8	1	0.75	0.184	0.184	0.048	0.052	1.305	1.599
n=500	2.8	1	1.7	0.026	0.024	0.050	0.049	0.453	0.410
	2.8	1	1.2	0.023	0.021	0.050	0.050	0.558	0.479
	2.8	1	0.75	0.117	0.117	0.047	0.047	0.557	0.672
n=2000	2.8	1	1.7	0.026	0.025	0.050	0.050	0.225	0.201
	2.8	1	1.2	0.021	0.022	0.053	0.054	0.276	0.235
	2.8	1	0.75	0.102	0.102	0.052	0.054	0.276	0.330
GB2									
n=100	2.8	1.7	0.9	0.024	0.020	0.047	0.048	1.204	1.056
	2.8	1.7	0.7	0.041	0.035	0.050	0.050	1.412	1.399
	2.8	1.7	0.5	0.340	0.340	0.050	0.051	1.154	2.304
	2.8	0.75	0.9	0.038	0.031	0.049	0.047	2.001	1.704
	2.8	0.75	0.7	0.046	0.036	0.044	0.041	2.224	2.058
	2.8	0.75	0.5	0.251	0.251	0.045	0.046	2.033	3.039
n=500	2.8	1.7	0.9	0.021	0.020	0.050	0.052	0.520	0.450
	2.8	1.7	0.7	0.040	0.038	0.050	0.051	0.607	0.592
	2.8	1.7	0.5	0.292	0.292	0.054	0.054	0.513	0.955
	2.8	0.75	0.9	0.025	0.023	0.050	0.050	0.825	0.715
	2.8	0.75	0.7	0.038	0.036	0.051	0.049	0.914	0.863
	2.8	0.75	0.5	0.195	0.195	0.053	0.052	0.842	1.245
n=2000	2.8	1.7	0.9	0.020	0.020	0.052	0.050	0.258	0.221
	2.8	1.7	0.7	0.046	0.045	0.056	0.055	0.301	0.291
	2.8	1.7	0.5	0.283	0.283	0.056	0.056	0.256	0.468
	2.8	0.75	0.9	0.022	0.022	0.050	0.049	0.407	0.350
	2.8	0.75	0.7	0.035	0.035	0.049	0.050	0.450	0.422
	2.8	0.75	0.5	0.173	0.173	0.047	0.048	0.420	0.609

Table 3.3: Coverage error of the proposed inference methods for the 99/20 ratio

	a	p	q	WCS	FCS	WCS-B	FTT-B	Width	Width
								WCS	WCS-B
Dagum									
n=500	2.8	0.9	1	0.007	0.006	0.070	0.068	244.942	6.863
	2.8	0.7	1	0.008	0.008	0.070	0.065	289.542	7.854
	2.8	0.5	1	0.009	0.009	0.062	0.056	382.030	10.262
n=2000	2.8	0.9	1	0.027	0.027	0.063	0.061	29.877	2.995
	2.8	0.7	1	0.027	0.027	0.065	0.064	34.885	3.392
	2.8	0.5	1	0.037	0.037	0.061	0.058	45.368	4.427
Singh-Madalla									
n=500	2.8	1	1.7	0.007	0.007	0.063	0.062	75.751	2.395
	2.8	1	1.2	0.006	0.006	0.067	0.064	147.814	4.366
	2.8	1	0.75	0.011	0.011	0.074	0.071	504.892	14.910
n=2000	2.8	1	1.7	0.026	0.025	0.050	0.050	0.225	0.201
	2.8	1	1.2	0.021	0.022	0.053	0.054	0.276	0.235
	2.8	1	0.75	0.102	0.102	0.052	0.054	0.276	0.330
GB2									
n=500	2.8	1.7	0.9	0.006	0.006	0.068	0.065	80.805	2.514
	2.8	1.7	0.7	0.005	0.005	0.060	0.055	94.788	3.036
	2.8	1.7	0.5	0.014	0.014	0.061	0.056	129.655	4.257
	2.8	0.75	0.9	0.009	0.008	0.069	0.068	535.895	15.118
	2.8	0.75	0.7	0.009	0.010	0.070	0.066	638.909	16.929
	2.8	0.75	0.5	0.010	0.010	0.064	0.060	861.432	21.273
n=2000	2.8	1.7	0.9	0.020	0.020	0.060	0.059	9.740	1.184
	2.8	1.7	0.7	0.020	0.020	0.054	0.052	11.863	1.416
	2.8	1.7	0.5	0.047	0.047	0.053	0.051	15.887	1.986
	2.8	0.75	0.9	0.026	0.025	0.064	0.062	64.444	6.043
	2.8	0.75	0.7	0.026	0.026	0.061	0.059	76.857	6.799
	2.8	0.75	0.5	0.032	0.032	0.060	0.058	101.711	8.440

Table 3.4: Empirical Power of the covered inference methods for the 99/20 ratio

	a	p	q	WCS	FCS	WCS-B	FTT-B	Width	Width
								WCS	WCS-B
Dagum									
n=500	2.8	0.9	1	0.037	0.037	0.832	0.837	245.850	6.874
	2.8	0.7	1	0.040	0.040	0.847	0.854	290.288	7.858
	2.8	0.5	1	0.037	0.037	0.871	0.881	386.770	10.198
n=2000	2.8	0.9	1	0.262	0.262	1.000	1.000	29.570	2.995
	2.8	0.7	1	0.268	0.268	1.000	1.000	34.667	3.390
	2.8	0.5	1	0.264	0.264	1.000	1.000	47.255	4.431
Singh-Madalla									
n=500	2.8	1	1.7	0.056	0.057	0.964	0.967	75.527	2.372
	2.8	1	1.2	0.044	0.044	0.894	0.897	150.269	4.376
	2.8	1	0.75	0.031	0.031	0.689	0.693	499.093	14.738
n=2000	2.8	1	1.7	0.380	0.380	1.000	1.000	9.262	1.106
	2.8	1	1.2	0.300	0.300	1.000	1.000	17.930	1.957
	2.8	1	0.75	0.225	0.225	1.000	1.000	59.395	5.847
GB2									
n=500	2.8	0.9	1.7	0.0524	0.0534	0.9666	0.9689	80.805	2.514
	2.8	0.7	1.7	0.0585	0.059	0.9707	0.9744	94.788	3.036
	2.8	0.5	1.7	0.0605	0.0605	0.9824	0.9848	129.655	4.257
	2.8	0.9	0.75	0.027	0.0273	0.6954	0.6989	535.895	15.118
	2.8	0.7	0.75	0.0291	0.0292	0.7113	0.7196	638.909	16.929
	2.8	0.5	0.75	0.0301	0.0301	0.7374	0.7521	861.432	21.273
n=2000	2.8	0.9	1.7	0.3772	0.3778	1	1	9.740	1.184
	2.8	0.7	1.7	0.3921	0.3931	1	1	11.863	1.416
	2.8	0.5	1.7	0.4092	0.4093	1	1	15.887	1.986
	2.8	0.9	0.75	0.214	0.2145	0.9996	0.9996	64.444	6.043
	2.8	0.7	0.75	0.203	0.2033	0.9999	0.9999	76.857	6.799
	2.8	0.5	0.75	0.1983	0.1983	1	1	101.711	8.440

Note – The DGP under the null assumes a shape parameter $a = 2.8$ while the DGP under

the alternative assumes a shape parameter $a = 4.2$.

Appendix B: Quantile optimality ratio (QOR) for the GB2 distribution

QOR is given by $QOR(u) = q_u/q_u''$ where q_u is the quantile density function. The pdf of the GB2 distribution is given by:

$$f(x) = \frac{|a| x^{ap-1}}{b^{ap} \beta(p, q) \left[1 + \left(\frac{x}{b}\right)^a\right]^{p+q}} \quad (3.4.1)$$

Let $N = |a| x^{ap-1}$ and $D = b^{ap} \beta(p, q) \left[1 + \left(\frac{x}{b}\right)^a\right]^{p+q}$, then $f(x) = N/D$. Then the first derivative of the $f(x)$ with respect to x is:

$$f'(x) = \frac{N' D - N D'}{D^2} \quad (3.4.2)$$

where $N' = d_1 N$ and $D' = d_2 D$ with $d_1 = (ap - 1)x^{-1}$ and d_2 :

$$d_2 = (p + q) \left(\frac{a}{b}\right) \left(\frac{x}{b}\right)^{a-1} \left[1 + \left(\frac{x}{b}\right)^a\right]^{-1} \quad (3.4.3)$$

Plugging the definitions of N' and D' into (3.4.2) and rearranging gives:

$$f'(x) = \frac{d_1 N D - N d_2 D}{D^2} = \frac{N(d_1 - d_2)}{D} = f(x)(d_1 - d_2). \quad (3.4.4)$$

The second derivative of $f(x)$ is thus:

$$f''(x) = f'(x)(d_1 - d_2) + f(x)(d_1' - d_2'), \quad (3.4.5)$$

where $d'_1 = -x^{-1}d_1$ and,

$$d'_2 = d_2 \left[(a-1) \left(\frac{x}{b}\right)^{-1} + \left[\left(\frac{a}{b}\right) \left(\frac{x}{b}\right)^{a-1} \right] \left[1 + \left(\frac{x}{b}\right)^a \right]^{-1} \right] \quad (3.4.6)$$

Now, the quantile density is defined as: $q_u = \frac{1}{f_Q}$ with $f_Q = f(x) \Big|_{x=Q_u}$. The first derivative of the quantile density can be derived using the (3.4.1), (3.4.4) and (3.4.5) evaluated at the quantile as follows:

$$q'_u = -\frac{f'_Q}{f_Q^3} \quad (3.4.7)$$

$$q''_u = \frac{-[f''_Q f_Q - 3(f'_Q)^2]}{f_Q^5} \quad (3.4.8)$$

where $f'_Q = f'(x) \Big|_{x=Q(u)}$ and $f''_Q = f''(x) \Big|_{x=Q(u)}$. Dividing q_u by q''_u gives the $QOR(u)$.

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.
- 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)*, pages 59–62.
- Bennett, B. (1962). On combining estimates of relative potency in bioassay. *Epidemiology & Infection*, 60(3):379–385.
- 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 Yéλου, 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*.
- Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics*, 108(2):317–342.
- 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. *Allgemeines Statistisches Archiv*, 80:285–298.
- Cheng, C. and Wu, J. (2010). Interval estimation of quantile ratios applied to anti-cancer drug screening by xenograft experiments. *Statistics in medicine*, 29(26):2669–2678.
- Coudin, E. and Dufour, J.-M. (2009). Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form. *The Econometrics Journal*, 12:S19–S49.
- Coudin, É. and Dufour, J.-M. (2017). Finite-sample generalized confidence distributions and sign-based robust estimators in median regressions with heterogeneous dependent errors.
- Cowell, F. and Victoria-Feser, M.-P. (1993). Robustness properties of inequality measures: The influence function and the principle of transfers.
- Cowell, F. A. (2000). Measurement of inequality. *Handbook of income distribution*, 1:87–166.
- 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.
- Cox, D. R. (1967). Fieller’s theorem and a generalization. *Biometrika*, 54(3-4):567–572.
- Dagenais, M. G. and Dufour, J.-M. (1991). Invariance, nonlinear models, and asymptotic tests. *Econometrica*, 59(6):1601–1615.
- Dagenais, M. G. and Dufour, J.-M. (1992). On the lack of invariance of some asymptotic tests to rescaling. *Economics Letters*, 38(3):251–257.
- Davidson, R. and Flachaire, E. (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics*, 141(1):141–166.
- Dominici, F., Cope, L., Naiman, D. Q., and Zeger, S. L. (2005). Smooth quantile ratio estimation. *Biometrika*, 92(3):543–557.
- Dominici, F. and Zeger, S. L. (2005). Smooth quantile ratio estimation with regression: estimating medical expenditures for smoking-attributable diseases. *Biostatistics*, 6(4):505–519.
- Dufour, J., Jouneau, F., and Torrès, O. (2008). On (non)-testability: Applications to linear and nonlinear semiparametric and nonparametric regression models. Technical report, Technical report, Department of Economics and CIREQ, McGill University, and

- Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65(6):1365–1387.
- Dufour, J.-M. (2003). Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics/Revue canadienne d'économique*, 36(4):767–808.
- Dufour, J.-M. and Dagenais, M. G. (1992). Nonlinear models, rescaling and test invariance. *Journal of Statistical Planning and Inference*, 32(1):111–135.
- 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 Zalgout, 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 lin-

- ear 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.
- Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, pages 569–593. Springer.
- Falk, M. (1986). On the estimation of the quantile density function. *Statistics & Probability Letters*, 4(2):69–73.
- Fieller, E. C. (1940). The biological standardization of insulin. *Supplement to the Journal of the Royal Statistical Society*, 7(1):1–64.
- Fieller, E. C. (1944). A fundamental formula in the statistics of biological assay, and some applications. *Quart. J. Pharm.*, 17:117–123.
- Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 16(2):175–185.
- 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 $(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variables and related models. *The Annals of Statistics*, 15(4):1351–1362.
- Hall, P., DiCiccio, T. J., Romano, J. P., et al. (1989). On smoothing and the bootstrap. *The Annals of Statistics*, 17(2):692–704.
- Hall, P. and Sheather, S. J. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(3):381–391.
- Hettmansperger, T. P. (1973). A large sample conservative test for location with unknown scale parameters. *Journal of the American Statistical Association*, 68(342):466–468.
- Hettmansperger, T. P. and Malin, J. S. (1975). A modified mood's test for location with no shape assumptions on the underlying distributions. *Biometrika*, 62(2):527–529.
- 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.
- Hyndman, R. J. and Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, 50(4):361–365.

- Jenkins, S. P. (2007). Inequality and the gb2 income distribution.
- Johannesson, M., Jönsson, B., and Karlsson, G. (1996). Outcome measurement in economic evaluation. *Health economics*, 5(4):279–296.
- Jones, M. C. (1992). Estimating densities, quantiles, quantile densities and density quantiles. *Annals of the Institute of Statistical Mathematics*, 44(4):721–727.
- Jones, M. C., Marron, J. S., and Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American statistical association*, 91(433):401–407.
- Kleiber, C. and Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*, volume 470. John Wiley & Sons.
- 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.
- Langford, E. (2006). Quartiles in elementary statistics. *Journal of Statistics Education*, 14(3).
- 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.

- Martin, M. A. (1990). On using the jackknife to estimate quantile variance. *Canadian Journal of Statistics*, 18(2):149–153.
- McDonald, J. B. (2008). Some generalized functions for the size distribution of income. In *Modeling Income Distributions and Lorenz Curves*, pages 37–55. Springer.
- Mills, J. A. and Zandvakili, S. (1997). Statistical inference via bootstrapping for measures of inequality. *Journal of Applied econometrics*, 12(2):133–150.
- Mood, A. M. et al. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *The Annals of Mathematical Statistics*, 25(3):514–522.
- Prendergast, L. A. and Staudte, R. G. (2016). Exploiting the quantile optimality ratio in finding confidence intervals for quantiles. *Stat*, 5(1):70–81.
- Prendergast, L. A. and Staudte, R. G. (2017). When large n is not enough—distribution-free interval estimators for ratios of quantiles. *The Journal of Economic Inequality*, 15(3):277–293.
- 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*, pages 1019–1036.
- Scheffé, H. et al. (1970). Multiple testing versus multiple estimation. improper confidence sets. estimation of directions and ratios. *The Annals of Mathematical Statistics*, 41(1):1–29.
- 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: Journal of the Econometric Society*, pages 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.
- Von Luxburg, U. and Franz, V. H. (2004). Confidence sets for ratios: A purely geometric approach to fieller's theorem.
- Von Luxburg, U. and Franz, V. H. (2009). A geometric approach to confidence sets for ratios: Fieller's theorem, generalizations and bootstrap. *Statistica Sinica*, pages 1095–1117.
- Wilcox, R. R. and Charlin, V. L. (1986). Comparing medians: A monte carlo study. *Journal of Educational Statistics*, 11(4):263–274.

- 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.