



**HAL**  
open science

## Goodness of Fit: an axiomatic approach

Frank A. Cowell, Russell Davidson, Emmanuel Flachaire

► **To cite this version:**

Frank A. Cowell, Russell Davidson, Emmanuel Flachaire. Goodness of Fit: an axiomatic approach. 2011. halshs-00639075

**HAL Id: halshs-00639075**

**<https://shs.hal.science/halshs-00639075>**

Preprint submitted on 8 Nov 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Goodness of Fit: an axiomatic approach**

**Frank A. Cowell**  
**Russell Davidson**  
**Emmanuel Flachaire**

**November 2011**

DT-GREQAM

# Goodness of Fit: an axiomatic approach

by

**Frank A. Cowell**

STICERD  
London School of Economics  
Houghton Street  
London, WC2A 2AE, UK

**email: [f.cowell@lse.ac.uk](mailto:f.cowell@lse.ac.uk)**

**Russell Davidson**

GREQAM  
Centre de la Vieille Charité  
2, rue de la Charité  
13236 Marseille cedex 02, France

Department of Economics and CIREQ  
McGill University  
Montreal, Quebec, Canada  
H3A 2T7

**email: [russell.davidson@mcgill.ca](mailto:russell.davidson@mcgill.ca)**

and

**Emmanuel Flachaire**

GREQAM, Aix-Marseille Université  
Centre de la Vieille Charité  
2, rue de la Charité  
13236 Marseille cedex 02, France

**email: [emmanuel.flachaire@univ-cezanne.fr](mailto:emmanuel.flachaire@univ-cezanne.fr)**

July 2011

## Abstract

An axiomatic approach is used to develop a one-parameter family of measures of divergence between distributions. These measures can be used to perform goodness-of-fit tests with good statistical properties. Asymptotic theory shows that the test statistics have well-defined limiting distributions which are however analytically intractable. A parametric bootstrap procedure is proposed for implementation of the tests. The procedure is shown to work very well in a set of simulation experiments, and to compare favourably with other commonly used goodness-of-fit tests. By varying the parameter of the statistic, one can obtain information on how the distribution that generated a sample diverges from the target family of distributions when the true distribution does not belong to that family. An empirical application analyses a UK income data set.

**Keywords:** Goodness of fit, axiomatic approach, measures of divergence, parametric bootstrap

**JEL codes:** D63, C10

# 1 Introduction

In this paper, we propose a one-parameter family of statistics that can be used to test whether an IID sample was drawn from a member of a parametric family of distributions. In this sense, the statistics can be used for a goodness-of-fit test. By varying the parameter of the family, denoted by  $\alpha$ , a range of statistics is obtained and, when the null hypothesis that the observed data were indeed generated by a member of the family of distributions is false, the different statistics can provide valuable information about the nature of the divergence between the unknown true data-generating process (DGP) and the target family.

We do not seek just to add to the collection of convenient goodness-of-fit statistics based on the empirical distribution function (EDF) of the sample. Our approach here is to motivate the goodness-of-fit criterion in the same sort of way as is commonly done with other measurement problems in economics and econometrics.<sup>1</sup> The role of axiomatisation is central. We invoke a relatively small number of axioms to capture the idea of divergence of one distribution from another using an informational structure that is common in studies of income mobility. From this divergence concept one immediately obtains a class of goodness-of-fit measures that inherit the principles embodied in the axioms. As it happens, the measures in this class also have a natural and attractive interpretation.

In order to be used for testing purposes, the goodness-of-fit statistics should have a distribution under the null that is known or can be simulated. Asymptotic theory shows that the null distribution of the members of the family of statistics is independent of the parameter  $\alpha$ , although that is certainly not true in finite samples. We show that the asymptotic distribution (as the sample size tends to infinity) exists, although it is not analytically tractable. However, its existence serves as an asymptotic justification for the use of a parametric bootstrap procedure for inference.

In addition to a set of simulation experiments designed to uncover the size and power properties of bootstrap tests based on our proposed family of statistics, we analyse a UK data set on households with below-average incomes, and show that we can derive a stronger conclusion by use of our tests than with most commonly used goodness-of-fit tests.

The paper is organised as follows. Section 2 sets out the formal framework and establishes a series of results that characterise the required class of measures. Section 3 derives the distribution of the members of this new

---

<sup>1</sup> As examples of the axiomatic method see Sen (1976a) on national income, Sen (1976b) on poverty and Ebert (1988) on inequality.

class. Section 4 examines the performance of the goodness-of-fit criteria in practice, and section 5 concludes.

## 2 Axiomatic foundation

Our approach is to characterise a measure of aggregate discrepancy between two empirical income distributions and then to apply the discrepancy measure as a goodness-of-fit criterion.

### 2.1 Representation of the problem

We adopt a structure that is often applied in the income-mobility literature. Let there be an ordered set of  $n$  income classes; each class  $i$  is associated with income level  $x_i$  where  $x_i < x_{i+1}$ ,  $i = 1, 2, \dots, n - 1$ . Let  $p_i \geq 0$  be the size of class  $i$ ,  $i = 1, 2, \dots, n$  which could be an integer in the case of finite populations or a real number in the case of a continuum of persons. We will work with the associated cumulative mass  $u_i = \sum_{j=1}^i p_j$ ,  $i = 1, 2, \dots, n$ . The set of distributions is given by  $U := \{\mathbf{u} \mid \mathbf{u} \in \mathbb{R}_+^n, u_1 \leq u_2 \leq \dots \leq u_n\}$ . The aggregate discrepancy measurement problem can be characterised as the relationship between two cumulative-mass vectors  $\mathbf{u}, \mathbf{v} \in U$ ; an alternative equivalent approach is to work with  $\mathbf{z} := (z_1, z_2, \dots, z_n)$ , where each  $z_i$  is the ordered pair  $(u_i, v_i)$ ,  $i = 1, \dots, n$  and belongs to a set  $Z$ , which we will take to be a connected subset of  $\mathbb{R}_+ \times \mathbb{R}_+$ . The problem focuses on the discrepancies between the  $u$ -values and the  $v$ -values. To capture this we introduce a discrepancy function  $d : Z \rightarrow \mathbb{R}$  such that  $d(z_i)$  is strictly increasing in  $|u_i - v_i|$ . Write the vector of discrepancies as

$$\mathbf{d}(\mathbf{z}) := (d(z_1), \dots, d(z_n)).$$

The problem can then be approached in two steps.

1. We represent the problem as one of characterising a weak ordering<sup>2</sup>  $\succeq$  on

$$Z^n := \underbrace{Z \times Z \times \dots \times Z}_n.$$

where, for any  $\mathbf{z}, \mathbf{z}' \in Z^n$  the statement “ $\mathbf{z} \succeq \mathbf{z}'$ ” should be read as “the pairs in  $\mathbf{z}$  constitute at least as good a fit according to  $\succeq$  as the pairs in

---

<sup>2</sup> This implies that it has the minimal properties of completeness, reflexivity and transitivity.

$\mathbf{z}'$ .” From  $\succeq$  we may derive the antisymmetric part  $\succ$  and symmetric part  $\sim$  of the ordering.<sup>3</sup>

2. We use the function representing  $\succeq$  to generate an aggregate discrepancy index.

In the first stage of step 1 we introduce some properties for  $\succeq$ , many of which correspond to those used in choice theory and in welfare economics.

## 2.2 Basic structure

**Axiom 1 (Continuity)**  $\succeq$  is continuous on  $Z^n$ .

**Axiom 2 (Monotonicity)** If  $\mathbf{z}, \mathbf{z}' \in Z^n$  differ only in their  $i$ th component then  $d(u_i, v_i) < d(u'_i, v'_i) \iff \mathbf{z} \succ \mathbf{z}'$ .

For any  $\mathbf{z} \in Z^n$  denote by  $\mathbf{z}(\zeta, i)$  the member of  $Z^n$  formed by replacing the  $i$ th component of  $\mathbf{z}$  by  $\zeta \in Z$ .

**Axiom 3 (Independence)** For  $\mathbf{z}, \mathbf{z}' \in Z^n$  such that:  $\mathbf{z} \sim \mathbf{z}'$  and  $z_i = z'_i$  for some  $i$  then  $\mathbf{z}(\zeta, i) \sim \mathbf{z}'(\zeta, i)$  for all  $\zeta \in [z_{i-1}, z_{i+1}] \cap [z'_{i-1}, z'_{i+1}]$ .

If  $\mathbf{z}$  and  $\mathbf{z}'$  are equivalent in terms of overall discrepancy and the fit in class  $i$  is the same in the two cases then a local variation in component  $i$  simultaneously in  $\mathbf{z}$  and  $\mathbf{z}'$  has no overall effect.

**Axiom 4 (Perfect local fit)** Let  $\mathbf{z}, \mathbf{z}' \in Z^n$  be such that, for some  $i$  and  $j$ ,  $u_i = v_i$ ,  $u_j = v_j$ ,  $u'_i = u_i + \delta$ ,  $v'_i = v_i + \delta$ ,  $u'_j = u_j - \delta$ ,  $v'_j = v_j - \delta$  and, for all  $k \neq i, j$ ,  $u'_k = u_k$ ,  $v'_k = v_k$ . Then  $\mathbf{z} \sim \mathbf{z}'$ .

The principle states that if there is a perfect fit in two classes then moving  $u$ -mass and  $v$ -mass simultaneously from one class to the other has no effect on the overall discrepancy.

**Theorem 1** Given Axioms 1 to 4 (a)  $\succeq$  is representable by the continuous function given by

$$\sum_{i=1}^n \phi_i(z_i), \forall \mathbf{z} \in Z^n \quad (1)$$

where, for each  $i$ ,  $\phi_i : Z \rightarrow \mathbb{R}$  is a continuous function that is strictly decreasing in  $|u_i - v_i|$  and (b)

$$\phi_i(u, u) = a_i + b_i u. \quad (2)$$

---

<sup>3</sup> For any  $\mathbf{z}, \mathbf{z}' \in Z^n$  “ $\mathbf{z} \succ \mathbf{z}'$ ” means “[ $\mathbf{z} \succeq \mathbf{z}'$ ] & [ $\mathbf{z}' \not\succeq \mathbf{z}$ ]”; “ $\mathbf{z} \sim \mathbf{z}'$ ” means “[ $\mathbf{z} \succeq \mathbf{z}'$ ] & [ $\mathbf{z}' \succeq \mathbf{z}$ ]”.

**Proof.** Axioms 1 to 4 imply that  $\succeq$  can be represented by a continuous function  $\Phi : Z^n \rightarrow \mathbb{R}$  that is increasing in  $|u_i - v_i|$ ,  $i = 1, \dots, n$ . Using Axiom 3 part (a) of the result follows from Theorem 5.3 of Fishburn (1970). Now take  $\mathbf{z}'$  and  $\mathbf{z}$  in as specified in Axiom 4. Using (1) it is clear that  $\mathbf{z} \sim \mathbf{z}'$  if and only if

$$\phi_i(u_i + \delta, u_i + \delta) - \phi_i(u_i, u_i) - \phi_j(u_j + \delta, u_j + \delta) + \phi_j(u_j, u_j) = 0$$

which can only be true if

$$\phi_i(u_i + \delta, u_i + \delta) - \phi_i(u_i, u_i) = f(\delta)$$

for arbitrary  $u_i$  and  $\delta$ . This is a standard Pexider equation and its solution implies (2). ■

**Corollary 1** *Since  $\succeq$  is an ordering it is also representable by*

$$\phi \left( \sum_{i=1}^n \phi_i(z_i) \right) \quad (3)$$

where  $\phi_i$  is defined as in (1), (2) and  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ , continuous and strictly monotonic increasing.

This additive structure means that we can proceed to evaluate the aggregate discrepancy problem one income class at a time. The following axiom imposes a very weak structural requirement, namely that the ordering remains unchanged by some uniform scale change to both  $u$ -values and  $v$ -values simultaneously. As Theorem 2 shows it is enough to induce a rather specific structure on the function representing  $\succeq$ .

**Axiom 5 (Population scale irrelevance)** *For any  $\mathbf{z}, \mathbf{z}' \in Z^n$  such that  $\mathbf{z} \sim \mathbf{z}'$ ,  $t\mathbf{z} \sim t\mathbf{z}'$  for all  $t > 0$ .*

**Theorem 2** *Given Axioms 1 to 5  $\succeq$  is representable by*

$$\phi \left( \sum_{i=1}^n u_i h_i \left( \frac{u_i}{v_i} \right) \right) \quad (4)$$

where  $h_i$  is a real-valued function.

**Proof.** Using the function  $\Phi$  introduced in the proof of Theorem 1, Axiom 5 implies

$$\begin{aligned}\Phi(\mathbf{z}) &= \Phi(\mathbf{z}') \\ \Phi(t\mathbf{z}) &= \Phi(t\mathbf{z}')\end{aligned}$$

and so, since this has to be true for arbitrary  $\mathbf{z}, \mathbf{z}'$  we have

$$\frac{\Phi(t\mathbf{z})}{\Phi(\mathbf{z})} = \frac{\Phi(t\mathbf{z}')}{\Phi(\mathbf{z}')} = \psi(t)$$

where  $\psi$  is a continuous function  $\mathbb{R} \rightarrow \mathbb{R}$ . Hence, using the  $\phi_i$  given in (1), we have for all  $\mathbf{z}$ :

$$\phi_i(tz_i) = \psi(t) \phi_i(z_i) \quad i = 1, \dots, n.$$

or, equivalently

$$\phi_i(tu_i, tv_i) = \psi(t) \phi_i(u_i, v_i), \quad i = 1, \dots, n. \quad (5)$$

So, in view of Aczél and Dhombres (1989), page 346 there must exist  $c \in \mathbb{R}$  and a function  $h_i : \mathbb{R}_+ \rightarrow \mathbb{R}$  such that

$$\phi_i(u_i, v_i) = u_i^c h_i\left(\frac{u_i}{v_i}\right). \quad (6)$$

From (2) and (6) it is clear that

$$\phi_i(u_i, u_i) = u_i^c h_i(1) = a_i + b_i u_i, \quad (7)$$

which implies  $c = 1$  for non-trivial cases. Putting (6) with  $c = 1$  into (3) gives the result. ■

The function  $h_i$  in Theorem 2 is arbitrary and it is useful to impose more structure. This is done in Section 2.3.

### 2.3 Mass discrepancy and goodness-of-fit

We now focus on the way in which one compares the  $(u, v)$  discrepancies in different parts of the distribution. The form of (4) suggests that discrepancy should be characterised terms of proportional differences:

$$d(z_i) = \max\left(\frac{u_i}{v_i}, \frac{v_i}{u_i}\right).$$

This is the form for  $d$  that we will assume from this point onwards. We also introduce:

**Axiom 6 (Discrepancy scale irrelevance)** *Suppose there are  $\mathbf{z}_0, \mathbf{z}'_0 \in Z^n$  such that  $\mathbf{z}_0 \sim \mathbf{z}'_0$ . Then for all  $t > 0$  and  $\mathbf{z}, \mathbf{z}'$  such that  $d(\mathbf{z}) = td(\mathbf{z}_0)$  and  $d(\mathbf{z}') = td(\mathbf{z}'_0)$ :  $\mathbf{z} \sim \mathbf{z}'$ .*

The principle states this. Suppose we have two distributional fits  $\mathbf{z}_0$  and  $\mathbf{z}'_0$  that are regarded as equivalent under  $\succeq$ . Then scale up (or down) all the mass discrepancies in  $\mathbf{z}_0$  and  $\mathbf{z}'_0$  by the same factor  $t$ . The resulting pair of distributional fits  $\mathbf{z}$  and  $\mathbf{z}'$  will also be equivalent.<sup>4</sup>

**Theorem 3** *Given Axioms 1 to 6  $\succeq$  is representable by*

$$\Phi(\mathbf{z}) = \phi \left( \sum_{i=1}^n u_i^\alpha v_i^{1-\alpha} \right) \quad (8)$$

where  $\alpha \neq 1$  is a constant.<sup>5</sup>

**Proof.** Take the special case where, in distribution  $\mathbf{z}'_0$  the discrepancy takes the same value  $r$  in all  $n$  classes. If  $(u_i, v_i)$  represents a typical component in  $\mathbf{z}_0$  then  $\mathbf{z}_0 \sim \mathbf{z}'_0$  implies

$$r = \psi \left( \sum_{i=1}^n u_i h_i \left( \frac{u_i}{v_i} \right) \right) \quad (9)$$

where  $\psi$  is the solution in  $r$  to

$$\sum_{i=1}^n u_i h_i \left( \frac{u_i}{v_i} \right) = \sum_{i=1}^n u_i h_i(r). \quad (10)$$

In (10) we can take the  $u_i$  as fixed weights. Using Axiom 6 in (9) requires

$$tr = \psi \left( \sum_{i=1}^n u_i h_i \left( t \frac{u_i}{v_i} \right) \right), \text{ for all } t > 0. \quad (11)$$

Using (10) we have

$$\sum_{i=1}^n u_i h_i \left( t \psi \left( \sum_{i=1}^n u_i h_i \left( \frac{u_i}{v_i} \right) \right) \right) = \sum_{i=1}^n u_i h_i \left( t \frac{u_i}{v_i} \right) \quad (12)$$

---

<sup>4</sup> Also note that Axiom 6 can be stated equivalently by requiring that, for a given  $\mathbf{z}_0, \mathbf{z}'_0 \in Z^n$  such that  $\mathbf{z}_0 \sim \mathbf{z}'_0$ , either (a) any  $\mathbf{z}$  and  $\mathbf{z}'$  found by rescaling the  $u$ -components will be equivalent or (b) any  $\mathbf{z}$  and  $\mathbf{z}'$  found by rescaling the  $v$ -components will be equivalent.

<sup>5</sup> The following proof draws on Ebert (1988).

Define variables  $q_i$  and functions  $\psi_i, i = 1, \dots, n$  such that

$$q_i := u_i h_i \left( \frac{u_i}{v_i} \right), \quad (13)$$

$$\frac{u_i}{v_i} = \psi_i(q_i). \quad (14)$$

From (12) - (14) we obtain

$$\sum_{i=1}^n u_i h_i \left( t \psi \left( \sum_{i=1}^n q_i \right) \right) = \sum_{i=1}^n u_i h_i (t \psi_i(q_i)). \quad (15)$$

Also define  $n + 1$  functions  $\theta_i$ :

$$\theta_0(q, t) := \sum_{i=1}^n u_i h_i(t \psi(q)), \quad (16)$$

$$\theta_i(q, t) := u_i h_i(t \psi_i(q)), i = 1, \dots, n. \quad (17)$$

From (15) - (17) we obtain

$$\theta_0 \left( \sum_{i=1}^n q_i, t \right) = \sum_{i=1}^n \theta_i(q_i, t),$$

which implies (Aczél 1966, p. 142)

$$\theta_i(q, t) = b_i(t) + B(t)q, i = 0, 1, \dots, n$$

where

$$b_0(t) = \sum_{i=1}^n b_i(t).$$

Therefore:

$$h_i \left( t \frac{u_i}{v_i} \right) = \frac{b_i(t)}{u_i} + B(t) h_i \left( \frac{u_i}{v_i} \right), i = 1, \dots, n, \quad (18)$$

the solution to which is

$$h_i(v) = \begin{cases} \beta_i v^{\alpha-1} + \gamma_i, & \alpha \neq 1, \\ \beta_i \log v + \gamma_i & \alpha = 1, \end{cases} \quad (19)$$

where  $\beta_i > 0$  is an arbitrary positive number – see Eichhorn (1978), Theorem 2.7.3. Substituting for  $h_i(\cdot)$  from (19) into (2) for the case where  $\beta_i$  is the same for all  $i$  gives the result. ■

## 2.4 Aggregate discrepancy index

Theorem 3 provides some of the essential structure of an aggregate discrepancy index. To make further progress consider the behaviour of the index over the subset of  $Z$  consisting of all distributions with given total mass:

$$Z(\bar{u}, \bar{v}) := \left\{ \mathbf{z} \in Z \mid \sum_{i=1}^n z_i = (\bar{u}, \bar{v}) \right\}.$$

Clearly, for all  $\mathbf{z} \in Z(\bar{u}, \bar{v})$  the discrepancy index must take the form

$$\Phi(\mathbf{z}) = \bar{\phi} \left( \sum_{i=1}^n u_i^\alpha v_i^{1-\alpha}; \bar{u}, \bar{v} \right), \quad (20)$$

where  $\bar{u}, \bar{v}$  are parameters of the function  $\bar{\phi}$  that is the counterpart of  $\phi$  in (8). It is reasonable to require that  $\Phi(\mathbf{z})$  should take the value zero when  $\mathbf{z}$  represents a “perfect fit,” so that there is no discrepancy between the  $u$ -distribution and the  $v$ -distribution; but there is a slight ambiguity about what the meaning of this requirement is.

A narrow interpretation of zero discrepancy is that  $v_i = u_i, i = 1, \dots, n$ , in which case the form (20) requires that

$$\bar{\phi} \left( \sum_{i=1}^n u_i; \bar{u}, \bar{u} \right) = 0, \quad (21)$$

in other words we have the restriction  $\bar{\phi}(\bar{u}; \bar{u}, \bar{u}) = 0$ . However, this restriction does not actually impose much additional structure. By contrast, suppose we take a broader interpretation of zero discrepancy, namely that if the  $v$ -distribution is obtained rescaling each component in the  $u$ -distribution by a factor  $k > 0$  then there is no discrepancy; in other words suppose we say that the total mass is not relevant in the evaluation of discrepancy only the relative frequencies in each class. This interpretation requires that, if  $v_i = k u_i, i = 1, \dots, n$  (where  $k = \bar{v}/\bar{u}$ ) then, from (20), we have

$$\bar{\phi} \left( k^{1-\alpha} \sum_{i=1}^n u_i; \bar{u}, \bar{v} \right) = 0 \quad (22)$$

which implies

$$\bar{\phi}(\bar{u}^\alpha \bar{v}^{1-\alpha}; \bar{u}, \bar{v}) = 0. \quad (23)$$

This can only be true for all  $\alpha$  if  $\phi$  in (8) and  $\bar{\phi}$  in (20) can be written in the form

$$\psi \left( \sum_{i=1}^n \left[ \frac{u_i}{\mu_u} \right]^\alpha \left[ \frac{v_i}{\mu_v} \right]^{1-\alpha} \right), \quad (24)$$

where  $\mu_u := n^{-1} \sum_{i=1}^n u_i$  and  $\mu_v := n^{-1} \sum_{i=1}^n v_i$ .

A suitable cardinalisation of (24) gives the aggregate discrepancy measure

$$G_\alpha := \frac{1}{\alpha(\alpha-1)} \sum_{i=1}^n \left[ \left[ \frac{u_i}{\mu_u} \right]^\alpha \left[ \frac{v_i}{\mu_v} \right]^{1-\alpha} - 1 \right], \alpha \in \mathbb{R}, \alpha \neq 0, 1 \quad (25)$$

where we have the following limiting forms for the cases  $\alpha = 0$  and  $\alpha = 1$ , respectively

$$G_0 = - \sum_{i=1}^n \frac{v_i}{\mu_v} \log \left( \frac{u_i}{\mu_u} / \frac{v_i}{\mu_v} \right), \quad (26)$$

$$G_1 = \sum_{i=1}^n \frac{u_i}{\mu_u} \log \left( \frac{u_i}{\mu_u} / \frac{v_i}{\mu_v} \right). \quad (27)$$

Expressions (25)-(27) constitute a *family* of aggregate discrepancy measures where an individual family member is characterised by choice of  $\alpha$ : a high positive  $\alpha$  produces an index that is particularly sensitive to discrepancies where  $u$  exceeds  $v$  and a negative  $\alpha$  yields an index that is sensitive discrepancies where  $v$  exceeds  $u$ .<sup>6</sup>

## 2.5 Goodness of fit

Our approach to the goodness-of-fit problem is to use the index constructed in section 2.4 to quantify the aggregate discrepancy between an empirical

---

<sup>6</sup> There is a natural extension to the case where one is dealing with a continuous distribution on support  $Y \subseteq \mathbb{R}$ . Expressions (25) - (27) become, respectively:

$$\begin{aligned} & \frac{1}{\alpha(\alpha-1)} \left[ \int_Y \left[ \frac{F_u(y)}{\mu_u} \right]^\alpha \left[ \frac{F_v(y)}{\mu_v} \right]^{1-\alpha} dy - 1 \right], \\ & - \int_Y \frac{F_v(y)}{\mu_v} \log \left[ \frac{F_u(y)}{\mu_u} / \frac{F_v(y)}{\mu_v} \right] dy, \text{ and} \\ & \int_Y \frac{F_u(y)}{\mu_u} \log \left[ \frac{F_u(y)}{\mu_u} / \frac{F_v(y)}{\mu_v} \right] dy. \end{aligned}$$

Clearly there is a family resemblance to the Kullback and Leibler (1951) measure of relative entropy or divergence measure of  $f_2$  from  $f_1$

$$\int_Y f_1 \log \left( \frac{f_2}{f_1} \right) dy$$

but with densities  $f$  replaced by cumulative distributions  $F$ .

distribution and a model. Given a set of observations  $\{x_1, x_2, \dots, x_n\}$  the empirical distribution function (EDF) is

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \iota(x_{(i)} \leq x),$$

where  $x_{(i)}$  denotes the  $i$ th smallest observation and  $\iota$  is an indicator function such that  $\iota(S) = 1$  if statement  $S$  is true and  $\iota(S) = 0$  otherwise. Denote the proposed model distribution by  $F(\cdot; \theta)$ , where  $\theta$  is a set of parameters and let

$$\begin{aligned} u_i &= F(x_{(i)}; \theta), i = 1, \dots, n \\ v_i &= \hat{F}(x_{(i)}) = \frac{i}{n}, i = 1, \dots, n; \end{aligned}$$

then  $u_i$  is a set of non-decreasing population proportions generated by the model from the  $n$  ordered observations. As before write  $\mu_u$  for the mean value of the  $u$ ; observe that

$$\mu_v = \frac{1}{n} \sum_{i=1}^n v_i = \sum_{i=1}^n \frac{i}{n^2} = \frac{n+1}{2n}.$$

Using (25)-(27) we then find that we have a family of goodness-of-fit statistics

$$G_\alpha(F, \hat{F}) = \frac{1}{\alpha(\alpha-1)} \sum_{i=1}^n \left[ \left[ \frac{u_i}{\mu_u} \right]^\alpha \left[ \frac{2i}{n+1} \right]^{1-\alpha} - 1 \right], \quad (28)$$

where  $\alpha \in \mathbb{R} \setminus \{0, 1\}$  is a parameter.<sup>7</sup>

### 3 Inference

If the parametric family  $F(\cdot, \theta)$  is replaced by a single distribution  $F$ , then the  $u_i$  become just  $F(x_{(i)})$ , and therefore have the same distribution as the order statistics of a sample of size  $n$  drawn from the uniform  $U(0,1)$  distribution. The statistic  $G_\alpha(F, \hat{F})$  in (28) is random only through the  $u_i$ , and so, for given  $\alpha$  and  $n$ , it has a fixed distribution, independent of  $F$ . Further, as  $n \rightarrow \infty$ , the distribution converges to a limiting distribution that does not depend on  $\alpha$ .

---

<sup>7</sup> In the cases  $\alpha = 0$  and  $\alpha = 1$  we have, respectively, that  $G_0(F, \hat{F}) = -\sum_{i=1}^n \frac{2i}{n+1} \log\left(\frac{[n+1]u_i}{2i\mu_u}\right)$  and  $G_1(F, \hat{F}) = \sum_{i=1}^n \frac{u_i}{\mu_u} \log\left(\frac{[n+1]u_i}{2i\mu_u}\right)$ .

To see this, we make use of a result concerning the empirical quantile process; see van der Vaart and Wellner (1996), example 3.9.24. Let  $F$  be a distribution function with continuous positive derivative  $f$  defined on a compact support. Let  $\hat{F}_n$  be the empirical distribution function of an IID sample drawn from  $F$ , and let  $Q(p) = F^{-1}(p)$  and  $\hat{Q}_n(p) = \hat{F}_n^{-1}(p)$ ,  $p \in [0, 1]$ , be the corresponding quantile functions. Since  $\hat{F}$  is a discrete distribution,  $\hat{Q}_n(p)$  is just the order statistic indexed by  $\lceil np \rceil$  of the sample. Here  $\lceil x \rceil$  denotes the smallest integer not less than  $x$ . Then

$$\sqrt{n}(\hat{Q}_n(p) - Q(p)) \rightsquigarrow -\frac{B \circ F(Q(p))}{f(Q(p))}, \quad (29)$$

where the notation  $\rightsquigarrow$  means that the left-hand side, considered as a stochastic process defined on  $[0, 1]$ , converges weakly to the distribution of the right-hand side, where  $f$  is the density of distribution  $F$ , and where  $B(p)$  is a standard Brownian bridge, that is, a Gaussian process with covariance function

$$\text{cov}(B(t), B(s)) = \min(s, t) - st.$$

The  $U(0,1)$  distribution certainly has compact support  $[0, 1]$ , and its density is constant and equal to 1 on that interval. The result (29) in this case reduces to

$$\sqrt{n}(u_{\lceil np \rceil} - p) \rightsquigarrow B(p). \quad (30)$$

We will be chiefly interested in the arguments  $t_i$  defined as  $i/(n+1)$ ,  $i = 1, \dots, n$ . Then we see that

$$\sqrt{n}(u_i - t_i) \rightsquigarrow B(t_i). \quad (31)$$

This result expresses the asymptotic joint distribution of the uniform order statistics. Note that  $E(u_i) = t_i$ .

Write  $u_i = t_i + z_i$ , where  $E(z_i) = 0$ . From (30), we see that the variance of  $n^{1/2}z_i$  is  $t_i(1-t_i)$  plus a term that vanishes as  $n \rightarrow \infty$ <sup>8</sup>. Thus  $z_i = O_p(n^{-1/2})$  as  $n \rightarrow \infty$ . We express the statistic  $G_\alpha(F, \hat{F})$ , under the null hypothesis that the  $u_i$  do indeed have the joint distribution of the uniform order statistics, replacing  $u_i$  by  $t_i + z_i$  and discarding terms that tend to 0 as  $n \rightarrow \infty$ . We see that

$$G_\alpha(F, \hat{F}) = \frac{1}{\alpha(\alpha-1)} \frac{1}{\mu_u^\alpha(1/2)^{1-\alpha}} \sum_{i=1}^n \left[ t_i \left( 1 + \frac{z_i}{t_i} \right)^\alpha - \mu_u^\alpha(1/2)^{1-\alpha} \right]. \quad (32)$$

---

<sup>8</sup> In fact, the true variance of  $z_i$  is  $t_i(1-t_i)/(n+2)$ .

Now, by Taylor's theorem,

$$t_i \left(1 + \frac{z_i}{t_i}\right)^\alpha = t_i + \alpha z_i + \frac{\alpha(\alpha-1)}{2} \frac{z_i^2}{t_i} + \frac{\alpha(\alpha-1)(\alpha-2)}{6} \frac{(\theta_i z_i)^3}{t_i^2}, \quad (33)$$

where  $0 \leq \theta_i \leq 1$ ,  $i = 1, \dots, n$ , and so

$$\sum_{i=1}^n t_i \left(1 + \frac{z_i}{t_i}\right)^\alpha = \frac{n}{2} + n\alpha\bar{z} + \frac{\alpha(\alpha-1)}{2} \sum_{i=1}^n \frac{z_i^2}{t_i} + o_p(1), \quad (34)$$

where  $\bar{z}$  is the mean of the  $z_i$ , since it can be shown that the sum over  $i$  of the last term on the right-hand side of (33) is  $o_p(1)$ . Here, we have made use of the fact that  $\sum_{i=1}^n t_i = (n+1)^{-1} \sum_{i=1}^n i = n/2$ . By definition,

$$\mu_u = n^{-1} \sum_{i=1}^n u_i = \frac{1}{2} + n^{-1} \sum_{i=1}^n z_i = \frac{1}{2} + \bar{z}.$$

It follows that

$$\mu_u^\alpha (1/2)^{1-\alpha} = \frac{1}{2} (1 + 2\bar{z})^\alpha.$$

Using Taylor's theorem once more, we see that

$$\mu_u^\alpha (1/2)^{1-\alpha} = \frac{1}{2} \left(1 + 2\alpha\bar{z} + 2\alpha(\alpha-1)\bar{z}^2 + \frac{4\alpha(\alpha-1)(\alpha-2)}{3} (\theta_\mu \bar{z})^3\right), \quad (35)$$

with  $0 \leq \theta_\mu \leq 1$ . Now  $\bar{z}$  is the estimation error made by estimating  $1/2$  by  $\mu_u$ , and so it is  $O_p(n^{-1/2})$ . The last term above is thus of order  $n^{-3/2}$  in probability. Putting together equations (34) and (35) gives

$$\sum_{i=1}^n \left[ t_i \left(1 + \frac{z_i}{t_i}\right)^\alpha - \mu_u^\alpha \left(\frac{1}{2}\right)^{-\alpha} \right] = \frac{\alpha(\alpha-1)}{2} \left[ \sum_{i=1}^n \frac{z_i^2}{t_i} - 2n\bar{z}^2 \right] + o_p(1),$$

and so from (32) we arrive at the result

$$G_\alpha(F, \hat{F}) = \sum_{i=1}^n \frac{z_i^2}{t_i} - 2n\bar{z}^2 + o_p(1). \quad (36)$$

It is striking that the leading-order term in (36) does not depend on  $\alpha$ . For finite  $n$ ,  $G_\alpha$  does of course depend on  $\alpha$ . Simulation shows that, for  $n$  even as small as 10, the distributions of  $G_\alpha$  and of the leading term in (36) are very close indeed for  $\alpha = 2$ , but that, for  $n$  even as large as 10,000, the distributions are noticeably different for values of  $\alpha$  far enough removed

from 2. The reason for this phenomenon is of course the factor of  $\alpha - 2$  in the remainder terms in (33) and (35).

If the limiting asymptotic distribution of  $G_\alpha$  exists, it is the same as that of the approximation in (36), and, if the latter exists, it is the distribution of the limiting random variable obtained by replacing  $z_i$  by  $n^{-1/2}B(t_i)$  (see (31)) and then letting  $n$  tend to infinity. For  $\bar{z}$  first, we have

$$n^{1/2}\bar{z} = n^{-1/2} \sum_{i=1}^n z_i =_d n^{-1} \sum_{i=1}^n B(t_i) \rightsquigarrow \int_0^1 B(t) dt. \quad (37)$$

Above, the symbol  $=_d$  signifies equality in distribution, and the last step follows on noting that the second last expression is a Riemann sum that approximates the integral.

Similarly, we see that

$$\sum_{i=1}^n z_i^2/t_i =_d n^{-1} \sum_{i=1}^n \frac{B^2(t_i)}{t_i} \rightsquigarrow \int_0^1 \frac{B^2(t)}{t} dt. \quad (38)$$

From (37) and (38), we see that the limiting distribution of  $G_\alpha$  is that of

$$\int_0^1 \frac{B^2(t)}{t} dt - 2 \left( \int_0^1 B(t) dt \right)^2. \quad (39)$$

The denominator of  $t$  in the first integral may lead one to suppose that the integral may diverge with positive probability. However, notice that the expectation of the integral is

$$\int_0^1 \frac{1}{t} \mathbb{E} B^2(t) dt = \int_0^1 (1-t) dt = \frac{1}{2}.$$

A longer calculation shows that the second moment of the integral is also finite, so that the integral is finite in mean square, and so also in probability. We conclude that the limiting distribution of  $G_\alpha$  exists, is independent of  $\alpha$ , and is equal to the distribution of (39).

We now turn to the more interesting case in which  $F$  does depend on a vector  $\theta$  of parameters. The quantities  $u_i$  are now given by  $u_i = F(x_{(i)}, \hat{\theta})$ , where  $\hat{\theta}$  is assumed to be a root- $n$  consistent estimator of  $\theta$ . If  $\theta$  is the true parameter vector, then we can write  $x_{(i)} = Q(v_i, \theta)$ , where  $Q(\cdot, \theta)$  is the quantile function inverse to the cumulative distribution function (CDF)  $F(\cdot, \theta)$ , and the  $v_i$  have the distribution of the uniform order statistics. Then we have  $u_i = F(Q(v_i, \theta), \hat{\theta})$ , and

$$\mu_u = n^{-1} \sum_{i=1}^n F(Q(v_i, \theta), \hat{\theta}).$$

The statistic (28) becomes

$$G_\alpha = \frac{1}{\alpha(\alpha-1)} \frac{1}{\mu_u^\alpha (1/2)^{1-\alpha}} \sum_{i=1}^n \left[ F(Q(v_i, \theta), \hat{\theta})^\alpha t_i^{1-\alpha} - \mu_u^\alpha (1/2)^{1-\alpha} \right].$$

Let  $p(x, \theta)$  be the gradient of  $F$  with respect to  $\theta$ , and define  $g(v, \theta)$  to be  $p(Q(v, \theta), \theta)$ . As before, we let  $z_i = v_i - t_i$ . Then a short Taylor expansion gives the approximation

$$F(Q(v_i, \theta), \hat{\theta}) = t_i + z_i + g^\top(t_i, \theta)s(\theta) + O_p(n^{-1}),$$

where  $s(\theta) = \hat{\theta} - \theta$  is the estimation error, and is of order  $n^{-1/2}$ . To leading order asymptotically, a calculation exactly like that leading to (36) gives

$$G_\alpha = \sum_{i=1}^n \frac{(z_i + g^\top(t_i, \theta)s(\theta))^2}{t_i} - 2 \left( n^{-1/2} \sum_{i=1}^n (z_i + g^\top(t_i, \theta)s(\theta)) \right)^2 + o_p(1). \quad (40)$$

This asymptotic expression depends explicitly on  $\theta$ , and also on the estimator  $\hat{\theta}$  that is used. In order to show that there does exist a limiting distribution for (40), we suppose that there exists a (vector) function  $h(x, \theta)$  such that

$$n^{1/2}(\hat{\theta} - \theta) = n^{1/2}s(\theta) = n^{-1/2} \sum_{i=1}^n h(x_i, \theta) + o_p(1), \quad (41)$$

where  $E_\theta h(x, \theta) = 0$ . The function  $h$  exists straightforwardly for most commonly used estimators, including maximum likelihood and least squares. Our sample is supposed to be IID, and so in (41) we can sum over the order statistics  $x_{(i)}$ . Then a short Taylor expansion gives

$$\begin{aligned} n^{1/2}s(\theta) &= n^{-1/2} \sum_{i=1}^n h(Q(v_i, \theta), \theta) + o_p(1) \\ &= n^{-1/2} \sum_{i=1}^n h(Q(t_i + z_i, \theta), \theta) + o_p(1) \\ &= n^{-1/2} \sum_{i=1}^n \left[ h(Q(t_i, \theta), \theta) + \frac{h'(Q(t_i, \theta), \theta)}{f(Q(t_i, \theta), \theta)} z_i \right] + o_p(1), \end{aligned} \quad (42)$$

where  $f(x, \theta)$  is the density that corresponds to  $F(x, \theta)$  and  $h'$  is the derivative of  $h$  with respect to its first argument. The integration over the whole real line means in fact integration over the support of the distribution  $F$ .

Now, again by use of an argument based on a Riemann sum, we see that

$$\begin{aligned} n^{-1} \sum_{i=1}^n h(Q(t_i, \theta), \theta) &= \int_0^1 h(Q(t, \theta), \theta) dt + O(n^{-1}) \\ &= \int_{-\infty}^{\infty} h(x, \theta) dF(x, \theta) + O(n^{-1}) = O(n^{-1}), \end{aligned}$$

because  $E_{\theta} h(x, \theta) = 0$ . Thus the first term in the sum in (42) is  $O(n^{-1/2})$  and can be ignored for the asymptotic distribution. For the second term, we replace  $z_i$  as before by  $n^{-1/2} B(t_i)$  to get

$$n^{1/2} s(\theta) \rightsquigarrow \int_0^1 \frac{h'(Q(t, \theta), \theta)}{f(Q(t, \theta), \theta)} B(t) dt = \int_{-\infty}^{\infty} h'(x, \theta) B(F(x, \theta)) dx, \quad (43)$$

where for the last step we make the change of variables  $x = Q(t, \theta)$ , and note that  $dF(x, \theta) = f(x, \theta) dx$ .

Next consider the sum

$$n^{-1/2} \sum_{i=1}^n (z_i + g^{\top}(t_i, \theta) s(\theta))$$

that appears in (40). By the definition of  $g$ ,  $g(t_i, \theta) = p(Q(t_i, \theta), \theta)$ . Hence, with error of order  $n^{-1}$ , we have

$$\begin{aligned} n^{-1} \sum_{i=1}^n g(t_i, \theta) &= n^{-1} \sum_{i=1}^n p(Q(t_i, \theta), \theta) \\ &= \int_0^1 p(Q(t, \theta), \theta) dt = \int_{-\infty}^{\infty} p(x, \theta) dF(x, \theta) \equiv P(\theta). \end{aligned}$$

Using (43), we have

$$n^{-1/2} \sum_{i=1}^n g^{\top}(t_i, \theta) s(\theta) \rightsquigarrow P^{\top}(\theta) \int_{-\infty}^{\infty} h'(x, \theta) B(F(x, \theta)) dx,$$

and so

$$n^{-1/2} \sum_{i=1}^n (z_i + g^{\top}(t_i, \theta) s(\theta)) \rightsquigarrow \int_0^1 B(t) dt + P^{\top}(\theta) \int_{-\infty}^{\infty} h'(x, \theta) B(F(x, \theta)) dx. \quad (44)$$

Finally, we consider the first sum in (40). By arguments similar to those used above, we see that

$$\sum_{i=1}^n \frac{(z_i + g^\top(t_i, \theta)s(\theta))^2}{t_i} \rightsquigarrow \int_0^1 \frac{1}{t} \left[ B(t) + p^\top(Q(t, \theta), \theta) \times \int_{-\infty}^{\infty} h'(x, \theta) B(F(x, \theta)) dx \right]^2 dt. \quad (45)$$

So as to be sure that the integral converges with probability 1, we have to show that the non-random integrals

$$\int_0^1 \frac{p(Q(t, \theta), \theta)}{t} dt \quad \text{and} \quad \int_0^1 \frac{p^2(Q(t, \theta), \theta)}{t} dt$$

are finite. Observe that

$$\int_0^1 \frac{p(Q(t, \theta), \theta)}{t} dt = \int_{-\infty}^{\infty} \frac{p(x, \theta)}{F(x, \theta)} dF(x, \theta) = \int_{-\infty}^{\infty} D_\theta \log F(x, \theta) dF(x, \theta),$$

where  $D_\theta$  is the operator that takes the gradient of its operand with respect to  $\theta$ . Similarly,

$$\int_0^1 \frac{p^2(Q(t, \theta), \theta)}{t} dt = \int_{-\infty}^{\infty} (D_\theta \log F(x, \theta))^2 F(x, \theta) dF(x, \theta),$$

Clearly, it is enough to require that  $D_\theta \log(F(x, \theta))$  should be bounded for all  $x$  in the support of  $F(\cdot, \theta)$ . It is worthy of note that this condition is not satisfied if varying  $\theta$  causes the support of the distribution to change.

Under the condition just stated, the results (45) and (44) establish the existence of the limiting distribution of  $G_\alpha$ . In general, this distribution depends on the parameter vector  $\theta$ , and so, in general,  $G_\alpha$  is not asymptotically pivotal with respect to the parametric family represented by the distributions  $F(\cdot, \theta)$ . However, if the family can be interpreted as a location-scale family, then it is not difficult to check that, if  $\hat{\theta}$  is the maximum-likelihood estimator, the random quantity  $g^\top(t_i, \theta)s(\theta)$  does not in fact depend on  $\theta$ . In addition, it turns out that the lognormal family also has this property. It would be interesting to see how common the property is, since, when it holds, the bootstrap benefits from an asymptotic refinement. But, even when it does not, the existence of the asymptotic distribution provides an asymptotic justification for the bootstrap.

It may be useful to give the details here of the bootstrap procedure used in the following section in order to perform goodness-of-fit tests, in the context

both of simulations and of an application with real data. It is a parametric bootstrap procedure; see for instance Horowitz (1997) or Davidson and MacKinnon (2006). Estimates  $\theta$  of the parameters of the family  $F(\cdot, \theta)$  are first obtained, after which the statistic of interest  $\hat{\tau}$  is computed, whether it is (28) for a chosen value of  $\alpha$  or one of the other statistics studied in the next section. Bootstrap samples of the same size as the original data sample are drawn from the estimated distribution  $F(\cdot, \hat{\theta})$ . Note that this is *not* a resampling procedure. For each of a suitable number  $B$  of bootstrap samples, parameter estimates  $\theta_j^*$ ,  $j = 1, \dots, B$ , are obtained using the same estimation procedure as with the original data, and the bootstrap statistic  $\tau_j^*$  computed, also exactly as with the original data, but with  $F(\cdot, \theta_j^*)$  as the target distribution. Then a bootstrap  $P$  value is obtained as the proportion of the  $\tau_j^*$  that are more extreme than  $\hat{\tau}$ , that is, greater than  $\hat{\tau}$  for statistics like (28) which reject for large values. For well-known reasons – see Davison and Hinkley (1997) or Davidson and MacKinnon (2000) – the number  $B$  should be chosen so that  $(B + 1)/100$  is an integer. In the sequel, we set  $B = 999$  unless otherwise stated. This computation of the  $P$  value can be used to test the fit of any parametric family of distributions.

## 4 Simulations and Application

We now turn to the way the new class of goodness-of-fit statistics performs in practice. In this section, we first study the finite sample properties of our  $G_\alpha$  test statistic and those of several standard measures: in particular we examine the comparative performance of the Anderson and Darling (1952) statistic (AD), the Cramér-von-Mises statistic given by

$$\text{CVM} = \int_{-\infty}^{\infty} [\hat{F}(x) - F(x, \hat{\theta})]^2 dF(x, \hat{\theta}),$$

the Kolmogorov-Smirnov statistic

$$\text{KS} = \sup_x |\hat{F}(x) - F(x, \hat{\theta})|,$$

and the Pearson chi-square (P) goodness-of-fit statistic

$$\text{P} = \sum_{i=1}^m (O_i - E_i)^2 / E_i,$$

where  $O_i$  is the observed percentage in the  $i^{\text{th}}$  histogram interval,  $E_i$  is the expected percentage in the  $i^{\text{th}}$  histogram interval and  $m$  is the number of

histogram intervals.<sup>9</sup> Then, we provide an application using a UK data set on income distribution.

## 4.1 Tests for Normality

Consider the application of the  $G_\alpha$  statistic to the problem of providing a test for normality. It is clear from expression (28) that different members of the  $G_\alpha$  family will be sensitive to different types of divergence of the EDF of the sample data from the model  $F$ . We take as an example two cases in which the data come from a Beta distribution, and we attempt to test the hypothesis that the data are normally distributed.

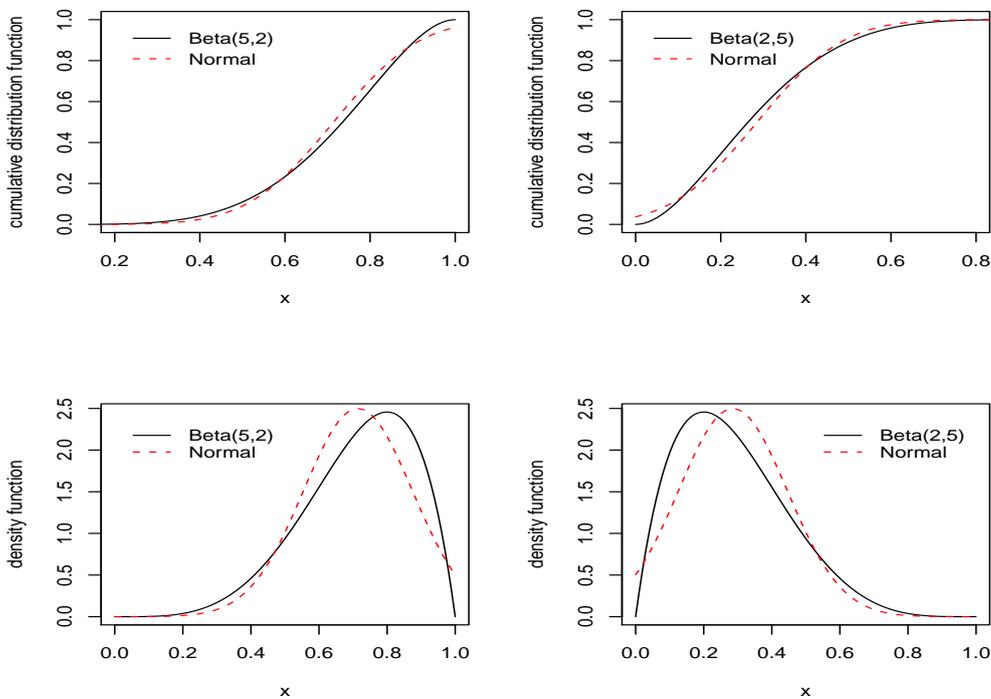


Figure 1: Different types of divergence of the data distribution from the model

Figure 1 represents the cumulative distribution functions and the density functions of two Beta distributions with their corresponding normal distributions (with equal mean and standard deviation). The parameters of the Beta distributions have been chosen to display divergence from the normal distribution in opposite directions. It is clear from Figure 1 that the Beta(5,2)

<sup>9</sup> We use the standard tests as implemented with R, the number of intervals  $m$  is due to Moore (1986). Note that G, AD, CVM and KS statistics are based on the empirical distribution function (EDF) and the P statistic is based on the density function.

distribution is skewed to the left and Beta(2,5) is skewed to the right, while the normal distribution is of course unskewed. As can be deduced from (28), in the first case the  $G_\alpha$  statistic decreases as  $\alpha$  increases, whereas in the second case it increases with  $\alpha$ .

$\alpha$	-2	-1	0	0.5	1	2	5	10
B(5,2)	2.29	2.03	1.85	1.79	1.73	1.64	1.47	1.35
B(2,5)	3.70	4.02	4.6	5.15	6.01	11.09	1.37e4	3.34e11
$t(4)$	61.35	6.83	4.17	3.99	3.94	4.02	4.74	7.30

Table 1: Normality tests with  $G_\alpha$  based on 1000 observations drawn from Beta and  $t$  distributions

These observations are confirmed by the results of Table 1, which shows normality tests with  $G_\alpha$  based on single samples of 1000 observations each drawn from the Beta(5,2) and from the Beta(2,5) distributions. Additional results are provided in the table with data generated by Student's  $t$  distribution with four degrees of freedom, denoted  $t(4)$ . The  $t$  distribution is symmetric, and differs from the normal on account of kurtosis rather than skewness. The results in Table 1 for  $t(4)$  show that  $G_\alpha$  does not increase or decrease globally with  $\alpha$ . However, as this example shows, the sensitivity to  $\alpha$  provides information on the sort of divergence of the data distribution from normality. It is thus important to compare the finite-sample performance of  $G_\alpha$  with that of other standard goodness-of-fit tests.

Table 2 presents simulation results on the size and power of normality tests using Student's  $t$  and Gamma ( $\Gamma$ ) distributions with several degrees of freedom,  $df = 2, 4, 6, \dots, 20$ . The  $t$  and  $\Gamma$  distributions provide two realistic examples that exhibit different types of departure from normality but tend to be closer to the normal as  $df$  increases. The values given in Table 2 are the percentages of rejections of the null  $H_0 : x \sim \text{Normal}$  at 5% nominal level when the true distribution of  $x$  is  $F_0$ , based on samples of 100 observations. Rejections are based on *bootstrap*  $P$  values for all tests, not just those that use  $G_\alpha$ . When  $F_0$  is the standard normal distribution (first line), the results measure the Type I error of the tests, by giving the percentage of rejections of  $H_0$  when it is true. For nominal level of 5%, we see that the Type I error is small. When  $F_0$  is not the normal distribution (other lines of the Table), the results show the power of the tests. The higher a value in the table, the better is the test at detecting departures from normality. As expected, results show that the power of all statistics considered increases as  $df$  decreases and the distribution is further from the normal distribution.

$F_0$	Standard tests				$G_\alpha$ test with $\alpha =$						
	AD	CVM	KS	P	-2	-1	0	0.5	1	2	5
N(0,1)	5.3	5.2	5.4	5.4	4.6	4.6	4.7	5.0	5.1	5.2	5.4
$t(20)$	<i>7.7</i>	7.3	6.6	5.8	<b>11.7</b>	10.4	7.3	6.6	6.7	6.5	6.2
$t(18)$	<i>8.9</i>	8.3	6.6	5.5	<b>12.4</b>	11.5	8.0	7.4	7.4	7.5	6.9
$t(16)$	<i>9.9</i>	8.9	7.1	6.3	<b>13.5</b>	12.9	9.4	8.6	8.6	8.7	8.0
$t(14)$	<i>9.8</i>	8.8	7.5	6.0	<b>15.0</b>	13.8	9.4	8.7	8.5	9.0	8.2
$t(12)$	<i>13.5</i>	12.0	8.9	6.5	<b>17.8</b>	17.8	12.7	11.8	11.7	11.9	11.0
$t(10)$	<i>15.2</i>	12.8	10.3	6.7	<b>21.8</b>	21.3	15.2	13.5	13.4	13.6	12.4
$t(8)$	<i>22.3</i>	19.0	13.4	8.2	<b>26.5</b>	26.5	20.7	19.1	19.0	19.4	17.7
$t(6)$	<i>37.5</i>	33.0	24.1	13.6	34.4	<b>37.3</b>	33.4	32.2	31.9	32.7	29.8
$t(4)$	<i>64.3</i>	59.9	48.5	28.6	49.6	<b>59.9</b>	59.4	58.5	58.7	59.5	56.6
$t(2)$	<i>98.0</i>	97.6	95.2	87.6	87.3	96.4	97.0	97.1	97.2	<b>97.3</b>	96.9
$\Gamma(20)$	<i>25.2</i>	21.9	17.8	10.2	0.1	4.5	13.8	16.1	18.4	23.2	<b>36.3</b>
$\Gamma(18)$	<i>28.3</i>	25.1	20.9	10.7	0.1	5.8	16.4	19.3	22.0	27.2	<b>40.0</b>
$\Gamma(16)$	<i>30.9</i>	27.2	21.9	12.0	0.1	7.1	18.9	22.0	24.5	29.5	<b>42.6</b>
$\Gamma(14)$	<i>34.5</i>	30.3	24.4	11.8	0.1	8.7	21.2	25.1	28.1	34.5	<b>49.3</b>
$\Gamma(12)$	<i>41.3</i>	36.6	28.5	14.5	0.1	10.7	26.4	30.3	34.0	40.6	<b>56.2</b>
$\Gamma(10)$	<i>48.9</i>	42.4	34.0	17.1	0.1	14.2	32.3	36.5	41.1	48.5	<b>64.4</b>
$\Gamma(8)$	<i>58.1</i>	51.7	41.6	22.0	0.1	19.9	41.7	47.1	51.6	59.7	<b>74.8</b>
$\Gamma(6)$	<i>72.7</i>	65.4	52.3	31.0	0.5	31.4	57.5	63.0	67.7	75.5	<b>87.8</b>
$\Gamma(4)$	<i>88.5</i>	82.1	68.8	49.7	2.0	55.7	79.6	84.0	87.0	92.1	<b>97.5</b>
$\Gamma(2)$	<i>99.8</i>	99.3	95.4	95.3	22.5	96.5	99.4	99.7	99.8	99.9	<b>100</b>

Table 2: Normality tests: percentage of rejections of  $H_0 : x \sim \text{Normal}$ , when the true distribution of  $x$  is  $F_0$ . Sample size = 100, 5000 replications, 999 bootstraps.

Among the standard goodness-of-fit tests, Table 2 shows that the AD statistic is better at detecting most departures from the normal distribution (*italic values*). The CVM statistic is close, but KS and P have poorer power. Similar results are found in Stephens (1986). Indeed, the Pearson chi-square test is usually not recommended as a goodness-of-fit test, on account of its inferior power properties.

Among the  $G_\alpha$  goodness-of-fit tests, Table 2 shows that the detection of greatest departure from the normal distribution is sensitive to the choice of  $\alpha$ . We can see that, in most cases, the most powerful  $G_\alpha$  test performs better than the most powerful standard test (*bold vs. italic values*). In addition, it is clear that  $G_\alpha$  increases with  $\alpha$  when the data are generated from the Gamma distribution. This is due to the fact that the Gamma distribution is skewed to the right.

## 4.2 Tests for other distributions

Table 3 presents simulation results on the power of tests for the lognormal distribution.<sup>10</sup> The values given in the table are the percentages of rejections of the null  $H_0 : x \sim \text{lognormal}$  at level 5% when the true distribution of  $x$  is the Singh-Maddala distribution – see Singh and Maddala (1976) – of which the CDF is

$$F_{\text{SM}}(x) = 1 - \frac{1}{(1 + ax^b)^c}$$

with parameters  $a = 100$ ,  $b = 2.8$ , and  $c = 1.7$ . We can see that the most powerful  $G_\alpha$  test ( $\alpha = 1$ ) performs better than the most powerful standard test (*bold vs. italic values*). The least powerful  $G_\alpha$  test ( $\alpha = 5$ ) performs similarly to the KS test.

nobs	Standard tests				$G_\alpha$ test with $\alpha =$						
	AD	CVM	KS	P	-2	-1	0	0.5	1	2	5
50	<i>20.4</i>	18.2	14.5	9.4	32.2	<b>33.7</b>	25.7	21.3	19.3	17.4	12.4
100	<i>33.7</i>	30.2	23.1	11.4	46.0	<b>49.0</b>	37.8	33.3	31.0	28.2	18.1
200	<i>56.2</i>	51.5	40.6	17.4	65.7	<b>70.3</b>	59.3	55.5	53.1	50.1	36.1
300	<i>73.9</i>	69.4	56.9	24.6	81.0	<b>84.3</b>	76.4	73.0	71.0	68.1	55.4
400	<i>84.3</i>	80.2	68.5	31.8	89.0	<b>91.5</b>	85.7	83.5	82.2	79.9	69.2
500	<i>90.6</i>	87.7	77.7	38.7	93.8	<b>95.0</b>	91.5	90.0	89.1	87.5	79.5

Table 3: Lognormality tests: percentage of rejections of  $H_0 : x \sim \text{lognormal}$ , when the true distribution of  $x$  is Singh-Maddala(100,2.8,1.7). 5000 replications, 499 bootstraps.

Table 4 presents simulation results on the power of tests for the Singh-Maddala distribution. The values given in the table are the percentage of rejections of the null  $H_0 : x \sim \text{SM}$  at 5% when the true distribution of  $x$  is lognormal. We can see that the most powerful  $G_\alpha$  test ( $\alpha = 5$ ) performs better than the most powerful standard test (*bold vs. italic values*).

Note that the two experiments concern the divergence between Singh-Maddala and lognormal distributions, but in opposite directions. For this reason the  $G_\alpha$  tests are sensitive to  $\alpha$  in opposite directions.

<sup>10</sup> Results under the null are close to the nominal level of 5%. For  $n = 50$ , we obtain rejection rates, for AD, CVM, KS, Pearson and G with  $\alpha = -2, -1, 0, 0.5, 1, 2, 5$  respectively, of 5.02, 4.78, 4.76, 4.86, 5.3, 5.06, 4.88, 4.6, 4.72, 5.18.

nobs	Standard tests				$G_\alpha$ test with $\alpha =$						
	AD	CVM	KS	P	-2	-1	0	0.5	1	2	5
500	<i>53.6</i>	43.3	32.3	16.7	11.3	37.3	47.7	50.2	53.0	57.4	<b>73.5</b>
600	<i>65.8</i>	52.6	37.4	20.1	18.6	51.3	60.1	62.4	64.5	68.4	<b>83.3</b>
700	<i>75.7</i>	61.8	43.7	22.8	24.9	61.4	71.5	73.3	74.4	77.9	<b>87.4</b>
800	<i>82.3</i>	69.3	53.1	27.6	37.9	72.5	79.3	80.6	82.6	85.8	<b>93.6</b>
900	<i>87.7</i>	75.9	54.8	30.6	45.8	77.5	82.9	83.9	85.6	88.5	<b>93.7</b>
1000	<i>91.2</i>	80.9	62.8	34.2	55.7	82.6	86.9	88.1	89.4	92.4	<b>96.4</b>

Table 4: Singh-Maddala tests: percentage of rejections of  $H_0 : x \sim \text{SM}$ , when the true distribution of  $x$  is lognormal(0,1). 1000 replications, 199 bootstraps.

### 4.3 Application

Finally, as a practical example, we take the problem of modelling income distribution using the UK Households Below Average Incomes 2004-5 (2006). The application uses the “before housing costs” income concept, deflated and equivalised using the OECD equivalence scale, for the cohort of ages 21-45, couples with and without children, excluding households with self-employed individuals. The variable used in the dataset is `oe.bhc`. We exclude households with self-employed individuals as reported incomes are known to be misrepresented. The empirical distribution  $\hat{F}$  consists of 3858 observations and has mean and standard deviation (398.28, 253.75).

We test the goodness-of-fit of a Singh-Maddala distribution, with 999 bootstrap samples used to compute the bootstrap  $P$  values. Table 5 presents the results with standard goodness-of-fit tests, Table 6 presents the results with  $G_\alpha$  tests. If we use standard goodness-of-fit statistics, we would not reject the Singh-Maddala distribution in most cases. Conversely, if we use  $G_\alpha$  goodness-of-fit statistics, we would reject the Singh-Maddala distribution in most cases. Our previous simulation study shows  $G_\alpha$  and AD have better finite sample properties. This leads us to conclude that the Singh-Maddala distribution is not a good fit, contrary to the conclusion from standard goodness-of-fit tests only.

test	AD	CVM	KS	P
statistic	0.569	0.047	0.009	54.4
$p$ -value	0.033	0.308	0.323	0.344

Table 5: Standard goodness-of-fit tests: bootstrap  $P$  values,  $H_0 : x \sim \text{SM}$ .

$\alpha$	-2	-1	0	0.5	1	2	5
statistic	164	1.36	0.441	0.404	0.390	0.382	0.398
$p$ -value	0.003	0.002	0.003	0.006	0.007	0.010	0.013

Table 6:  $G_\alpha$  goodness-of-fit tests: bootstrap  $P$  values,  $H_0 : x \sim \text{SM}$ .

## 5 Concluding Remarks

The family of goodness-of-fit tests presented in this paper has been seen to have excellent size and power properties as compared with other, commonly used, goodness-of-fit tests. It has the further advantage that the profile of the  $G_\alpha$  statistic as a function of  $\alpha$  can provide valuable information about the nature of the departure from the target family of distributions, when that family is wrongly specified.

We have advocated the use of the parametric bootstrap for tests based on  $G_\alpha$ . The distributions of the limiting random variables (39) and (34) exist, as shown, but cannot be conveniently used without a simulation experiment that is at least as complicated as that involved in a bootstrapping procedure. In addition, there is no reason to suppose that the asymptotic distributions are as good an approximation to the finite-sample distribution under the null as the bootstrap distribution. We rely on the mere existence of the limiting distribution in order to justify use of the bootstrap. The same reasoning applies, of course, to the conventional goodness-of-fit tests studied in Section 4. They too give more reliable inference in conjunction with the parametric bootstrap.

Of course, the  $G_\alpha$  statistics for different values of  $\alpha$  are correlated, and so it is not immediately obvious how to conduct a simple, powerful, test that works in all cases. It is clearly interesting to compute  $G_\alpha$  for various values of  $\alpha$ , and so a solution to the problem would be to use as test statistic the maximum value of  $G_\alpha$  over some appropriate range of  $\alpha$ . The simulation results in the previous section indicate that a range of  $\alpha$  from -2 to 5 should be enough to provide ample power. It would probably be inadvisable to consider values of  $\alpha$  outside this range, given that it is for  $\alpha = 2$  that the finite-sample distribution is best approximated by the limiting asymptotic distribution. That said, the maximum of  $G_\alpha$  has the same limiting distribution as  $G_\alpha$  for any fixed  $\alpha$ , because there is only one limit distribution for all  $\alpha$ . Thus bootstrapping the maximum over the chosen range is asymptotically justified, and seems likely to give a powerful test, to the extent that our simulation results can be generalised.

## References

- Aczél, J. (1966). *Lectures on Functional Equations and their Applications*. Number 9 in Mathematics in Science and Engineering. New York: Academic Press.
- Aczél, J. and J. G. Dhombres (1989). *Functional Equations in Several Variables*. Cambridge: Cambridge University Press.
- Anderson, T. W. and D. A. Darling (1952). “Asymptotic Theory of Certain ‘Goodness-of-Fit’ Criteria based on Stochastic Processes”, *Annals of Mathematical Statistics*, 23, 193–212.
- Davidson, R. and J. G. MacKinnon (2000). “Bootstrap Tests: How Many Bootstraps” *Econometric Reviews*, 19, 55–68.
- Davidson, R. and J. G. MacKinnon (2006). “Bootstrap Methods in Econometrics”, Chapter 23 of *Palgrave Handbook of Econometrics*, Volume 1, *Econometric Theory*, eds T. C. Mills and K. Patterson, Palgrave-Macmillan, London.
- Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and their Application*. Cambridge University Press.
- Ebert, U. (1988). “Measurement of inequality: an attempt at unification and generalization”, *Social Choice and Welfare* 5, 147–169.
- Eichhorn, W. (1978). *Functional Equations in Economics*. Reading Massachusetts: Addison Wesley.
- Fishburn, P. C. (1970). *Utility Theory for Decision Making*. New York: John Wiley.
- Horowitz, J. L. (1997). “Bootstrap Methods in Econometrics: Theory and Numerical Performance”, in David M. Kreps and Kenneth F. Wallis, eds, *Advances in Economics and Econometrics: Theory and Applications*, Volume 3, 188–222. Cambridge, Cambridge University Press.
- Households Below Average Income (HBAI) 1994/95–2004/05. See <http://statistics.dwp.gov.uk/asd/hbai/hbai2005/contents.asp>
- Kullback, S. and R. A. Leibler (1951). “On Information and Sufficiency”, *Annals of Mathematical Statistics*, 22(1), 79–86.
- Moore, D. S. (1986). “Tests of the chi-squared type”, in *Goodness-of-fit techniques*, eds R. B. D’Agostino and M. A. Stephens, M.A., Marcel Dekker, New York.
- Sen, A. K. (1976a). “Real national income”, *Review of Economic Studies*, 43, 19–39.

- Sen, A. K. (1976b). "Poverty: An ordinal approach to measurement", *Econometrica*, 44, 219–231.
- Singh, S. K. and G. S. Maddala (1976). "A Function for Size Distribution of Incomes", *Econometrica* 44, 963–970.
- Stephens, M. A. (1986). "Tests based on EDF statistics", in *Goodness-of-fit techniques*, eds R. B. D'Agostino and M. A. Stephens, M.A., 97-193, Marcel Dekker, New York.
- van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.