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

Russell Davidson

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

email: russell.davidson@mcgill.ca

and

Emmanuel Flachaire

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

email: emmanuel.flachaire@univ-amu.fr

February 2014
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:** D31, 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, 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.

Many tests of goodness of fit exist already, of course. Test statistics that are based on the empirical distribution function (EDF) of the sample include the Anderson-Darling statistic (see Anderson and Darling (1952)), the Cramér-von Mises statistic, and the Kolmogorov-Smirnov statistic. See Stephens (1986) for much more information on these and other statistics. The Pearson chi-square goodness-of-fit statistic, on the other hand, is based on a histogram approximation to the density; a reference more recent than Pearson’s original paper is Plackett (1983).

Here our aim is not just to add to the collection of existing goodness-of-fit statistics. Our approach 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. As examples of the axiomatic method, see Sen (1976a) on national income, Sen (1976b) on poverty, and Ebert (1988) on inequality. 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 the context of income distribution. We emphasise, however, that the approach is quite general, although in the sequel we use income distributions as our principal example.

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 of the family, 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.

A set of simulation experiments was designed to uncover the size and power
properties of bootstrap tests based on our proposed family of statistics, and to compare these properties with those of four other commonly used goodness-of-fit tests. We find that our tests have superior performance. In addition, 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 the other 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 class. Section 4 examines the performance of the goodness-of-fit criteria in practice, and uses them to analyse a UK income dataset. Section 5 concludes. All proofs are found in the Appendix.

2 Axiomatic foundation

The axiomatic approach developed in this section is in part motivated by its potential application to the analysis of income distributions.
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 \( x_i < x_{i+1} \), \( i = 1, 2, ..., n - 1 \). Let \( p_i \geq 0 \) be the size of class \( i \), \( i = 1, 2, ..., n \) 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 \( u_i = \sum_{j=1}^{i} p_j \), \( i = 1, 2, ..., n \). The set of distributions is given by \( U := \{ u \mid u \in \mathbb{R}^n_+, u_1 \leq u_2 \leq ... \leq u_n \} \). The aggregate discrepancy measurement problem can be characterised as the relationship between two cumulative-mass vectors \( u, v \in U \). An alternative equivalent approach is to work with \( z : = (z_1, z_2, ..., z_n) \), where each \( z_i \) is the ordered pair \((u_i, v_i)\), \( i = 1, \ldots, n \) 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 \( d : Z \to \mathbb{R} \) such that \( d(z_i) \) is strictly increasing in \(|u_i - v_i|\). Write the vector of discrepancies as

\[
\mathbf{d}(z) := (d(z_1), ..., d(z_n)).
\]

The problem can then be approached in two steps.
1. We represent the problem as one of characterising a weak ordering \(^1 \succeq \) on
\[
\mathbb{Z}^n := \mathbb{Z} \times \mathbb{Z} \times \ldots \times \mathbb{Z}
\]
where, for any \(z, z' \in \mathbb{Z}^n\) the statement “\(z \succeq z'\)” should be read as “the pairs in \(z'\) constitute at least as good a fit according to \(\succeq\) as the pairs in \(z\).” From \(\succeq\) we may derive the antisymmetric part \(\succ\) and symmetric part \(\sim\) of the ordering.\(^2\)

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 \(\mathbb{Z}^n\).

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

\(^1\) This implies that it has the minimal properties of completeness, reflexivity and transitivity.

\(^2\) For any \(z, z' \in \mathbb{Z}^n\) “\(z \succ z'\)” means “[\(z \succeq z'\] & \(z' \not\succeq z\)]”; and “\(z \sim z'\)” means “[\(z \succeq z'\] & \(z' \succeq z\)].”
For any $z \in Z^n$ denote by $z(\zeta, i)$ the member of $Z^n$ formed by replacing the $i$th component of $z$ by $\zeta \in Z$.

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

If $z$ and $z'$ 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 $z$ and $z'$ has no overall effect.

**Axiom 4 (Perfect local fit)** Let $z, z' \in Z^n$ be such that, for some $i$ and $j$, and for some $\delta > 0$, $u_i = v_i$, $u_j = v_j$, $u'_i = u_i + \delta$, $v'_i = v_i + \delta$, $u'_j = u_j - \delta$, $v'_j = v_j - \delta$ and, for all $k \neq i, j$, $u'_k = u_k$, $v'_k = v_k$. Then $z \sim z'$.

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

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

(1)
where, for each $i = 1, \ldots, n$, $\phi_i : \mathbb{Z} \to \mathbb{R}$ is a continuous function that is strictly increasing in $|u_i - v_i|$, with $\phi(0,0) = 0$; and

(b)

$$
\phi_i (u, u) = b_i u.
$$

(2)

Proof. In the Appendix. □

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

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

(3)

where $\phi_i$ is defined as in (1), (2) and $\phi : \mathbb{R} \to \mathbb{R}$ is continuous and strictly 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 $z, z' \in \mathbb{Z}^n$ such that $z \sim z'$, $t z \sim t z'$ for all $t > 0$. 

8
Theorem 2  Given Axioms 1 to 5 is representable by

\[ \phi \left( \sum_{i=1}^{n} u_i c_i h_i \left( \frac{v_i}{u_i} + b_i v_i \right) \right) \]  \hspace{1cm} (4)

where, for all \(i = 1, \ldots, n\), \(h_i\) is a real-valued function with \(h_i(1) = 0\), and \(b_i = 0\) unless \(c = 1\).

Proof. In the Appendix

The functions \(h_i\) in Theorem 2 are 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 in 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 \(z_0, z'_0 \in Z^n\)
such that \( z_0 \sim z_0' \). Then for all \( t > 0 \) and \( z, z' \) such that \( d(z) = td(z_0) \) and \( d(z') = td(z_0') \): \( z \sim z' \).

The principle states this. Suppose we have two distributional fits \( z_0 \) and \( z_0' \) that are regarded as equivalent under \( \succeq \). Then scale up (or down) all the mass discrepancies in \( z_0 \) and \( z_0' \) by the same factor \( t \). The resulting pair of distributional fits \( z \) and \( z' \) will also be equivalent.\(^3\)

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

\[
\Phi(z) = \phi \left( \sum_{i=1}^{n} (\delta_i u_i + c_i u_i^{1-\alpha} v_i^\alpha) \right)
\]

where \( \alpha \), the \( \delta_i \) and the \( c_i \) are constants, with \( c_i > 0 \), and \( \delta_i + c_i \) is equal to the \( b_i \) of (2) and (4).

**Proof.** In the Appendix. ■

### 2.4 Aggregate discrepancy index

Theorem 3 provides some of the essential structure of an aggregate discrepancy index. We can impose further structure by requiring that the index should be invariant to the scale of the \( u \)-distribution and to that of the \( v \)-distribution

\(^3\) Also note that Axiom 6 can be stated equivalently by requiring that, for a given \( z_0, z_0' \in Z^n \) such that \( z_0 \sim z_0' \), either (a) any \( z \) and \( z' \) found by rescaling the \( u \)-components will be equivalent or (b) any \( z \) and \( z' \) found by rescaling the \( v \)-components will be equivalent.
separately. In other words, we may say that the total mass in the \( u \)- and \( v \)-distributions is not relevant in the evaluation of discrepancy, but only the relative frequencies in each class. This implies that the discrepancy measure \( \Phi(z) \) must be homogeneous of degree zero in the \( u_i \) and in the \( v_i \) separately. But it also means that the requirement that \( \phi_i \) is increasing in \(|u_i - v_i|\) holds only once the two scales have been fixed.

**Theorem 4** If in addition to Axioms 1-6 we require that the ordering \( \succeq \) should be invariant to the scales of the masses \( u_i \) and of the \( v_i \) separately, the ordering can be represented by

\[
\Phi(z) = \phi\left(\sum_{i=1}^{n} \left[ \frac{u_i}{\mu_u} \right]^{1-\alpha} \left[ \frac{v_i}{\mu_v} \right]^{\alpha} \right),
\]

(6)

where \( \mu_u = n^{-1} \sum_{i=1}^{n} u_i \), \( \mu_v = n^{-1} \sum_{i=1}^{n} v_i \), and \( \phi(n) = 0 \).

**Proof.** In the Appendix. 

A suitable cardinalisation of (6) gives the aggregate discrepancy measure

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

(7)

The denominator of \( \alpha(\alpha - 1) \) is introduced so that the index, which otherwise would be zero for \( \alpha = 0 \) or \( \alpha = 1 \), takes on limiting forms, as follows for \( \alpha = 0 \)
and $\alpha = 1$ respectively:

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

(8)

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

(9)

Expressions (7)-(9) 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 $v$ exceeds $u$ and a negative $\alpha$ yields an index that is sensitive to discrepancies where $u$ exceeds $v$. There is a natural extension to the case in which one is dealing with a continuous distribution on support $Y \subseteq \mathbb{R}$. Expressions (7) - (9) become, respectively:

$$\frac{1}{\alpha(\alpha - 1)} \left[ \int_{Y} \left[ \frac{F_v(y)}{\mu_v} \right]^\alpha \left[ \frac{F_u(y)}{\mu_u} \right]^{1-\alpha} \, dy - 1 \right],$$

$$- \int_{Y} \frac{F_u(y)}{\mu_u} \log \left( \frac{F_v(y)}{\mu_v} \right) \left( \frac{F_u(y)}{\mu_u} \right) \, dy,$$

and

$$\int_{Y} \frac{F_v(y)}{\mu_v} \log \left( \frac{F_v(y)}{\mu_v} \right) \left( \frac{F_v(y)}{\mu_u} \right) \, dy.$$
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 \( F \).

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

\[
\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_{(i)} \leq x),
\]

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

\[ v_i = F(x_{(i)}; \theta), \quad i = 1, \ldots, n \]

\[ u_i = \hat{F}_n(x_{(i)}) = \frac{i}{n}, \quad i = 1, \ldots, n. \]

Then \( v_i \) is a set of non-decreasing population proportions generated by the model from the \( n \) ordered observations. As before write \( \mu_v \) for the mean value of the \( v_i \); observe that

\[ \mu_u = \frac{1}{n} \sum_{i=1}^{n} u_i = \sum_{i=1}^{n} \frac{i}{n^2} = \frac{n + 1}{2n}. \]

Using (7)-(9) we then find that we have a family of goodness-of-fit statistics

\[ G_\alpha(F, \hat{F}_n) = \frac{1}{\alpha(\alpha - 1)} \sum_{i=1}^{n} \left[ \left( \frac{v_i}{\mu_v} \right)^\alpha \left( \frac{2i}{n + 1} \right)^{1 - \alpha} - 1 \right], \quad (10) \]

where \( \alpha \in \mathbb{R} \setminus \{0, 1\} \) is a parameter. In the cases \( \alpha = 0 \) and \( \alpha = 1 \) we have, respectively, that

\[ G_0(F, \hat{F}_n) = -\sum_{i=1}^{n} \frac{2i}{n + 1} \log \left( \frac{[n + 1]v_i}{2i\mu_v} \right) \text{ and} \]

\[ G_1(F, \hat{F}_n) = \sum_{i=1}^{n} \frac{v_i}{\mu_v} \log \left( \frac{[n + 1]v_i}{2i\mu_v} \right). \]
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 $G_\alpha(F, \hat{F}_n)$ in (10) is random only through the $u_i$, and so, for given $\alpha$ and $n$, it has a fixed distribution, independent of $F$. Further, as $n \to \infty$, the distribution converges to a limiting distribution that does not depend on $\alpha$.

**Theorem 5** Let $F$ be a distribution function with continuous positive derivative defined on a compact support, and let $\hat{F}_n$ be the empirical distribution of an IID sample of size $n$ drawn from $F$. The statistic $G_\alpha(F, \hat{F}_n)$ in (10) tends in distribution as $n \to \infty$ to the distribution of the random variable

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

where $B(t)$ is a standard Brownian bridge, that is, a Gaussian stochastic process defined on the interval $[0,1]$ with covariance function

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

**Proof.** See the Appendix. □
The denominator of $t$ in the first integral in (11) may lead one to suppose that the integral may diverge with positive probability. However, notice that the expectation of the integral is

$$\int_0^1 \frac{1}{t}EB^2(t) \, dt = \int_0^1 (1-t) \, dt = \frac{1}{2}.$$ 

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 (11).

**Remark:** As one might expect from the presence of a Brownian bridge in the asymptotic distribution of $G_\alpha(F, \hat{F}_n)$, the proof of the theorem makes use of standard results from empirical process theory; see van der Vaart and Wellner (1996).

We now turn to the more interesting case in which $F$ does depend on a vector $\theta$ of parameters. The quantities $v_i$ are now given by $v_i = F(x_i, \hat{\theta})$, 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 $x_i = Q(u_i, \theta)$, where $Q(\cdot, \theta)$ is the quantile function inverse to the distribution function $F(\cdot, \theta)$, and the $u_i$ have the distribution of the uniform order statistics. Then we have $v_i = F(Q(u_i, \theta), \hat{\theta})$. 

16
and

\[ \mu_v = n^{-1} \sum_{i=1}^{n} F(Q(u_i, \hat{\theta}), \hat{\theta}). \]

The statistic (10) becomes

\[ G_\alpha(F, \hat{F}_n) = \frac{1}{\alpha(\alpha - 1)} \frac{1}{\mu_v^\alpha(1/2)^{1-\alpha}} \sum_{i=1}^{n} \left[ F(Q(u_i, \theta), \hat{\theta})^\alpha t_i^{1-\alpha} - \mu_v^\alpha(1/2)^{1-\alpha} \right], \]

where \( t_i = i/(n + 1) \). Let \( p(x, \theta) \) be the gradient of \( F \) with respect to \( \theta \), and make the definition

\[ P(\theta) = \int_{-\infty}^{\infty} p(x, \theta) \, dF(x, \theta). \]

Then we have

**Theorem 6** Consider a family of distribution functions \( F(\cdot, \theta) \), indexed by a parameter vector \( \theta \) contained in a finite-dimensional parameter space \( \Theta \). For each \( \theta \in \Theta \), suppose that \( F(\cdot, \theta) \) has a continuous positive derivative defined on a compact support, and that it is continuously differentiable with respect to the vector \( \theta \). Let \( \hat{F}_n \) be the EDF of an IID sample \( \{x_1, \ldots, x_n\} \) of size \( n \) drawn from the distribution \( F(\cdot, \theta) \) for some given fixed \( \theta \). Suppose that \( \hat{\theta} \) is a root-\( n \) consistent estimator of \( \theta \) such that, as \( n \to \infty \),

\[ n^{1/2}(\hat{\theta} - \theta) = n^{-1/2} \sum_{i=1}^{n} h(x_i, \theta) + o_p(1) \]

(13)
for some vector function $h$, differentiable with respect to its first argument, and where $h(x, \theta)$ has expectation zero when $x$ has the distribution $F(x, \theta)$. The statistic $G_\alpha(F, \hat{F}_n)$ given by (12) has a finite limiting asymptotic distribution as $n \to \infty$, expressible as the distribution of the random variable

$$
\int_0^1 \frac{1}{t} \left[ B(t) + p^\top (Q(t, \theta), \theta) \int_{-\infty}^\infty h'(x, \theta) B(F(x, \theta)) \, dx \right]^2 \, dt \\
-2 \left[ \int_0^1 B(t) \, dt + P^\top (\theta) \int_{-\infty}^\infty h'(x, \theta) B(F(x, \theta)) \, dx \right]^2. \tag{14}
$$

Here $B(t)$ is a standard Brownian bridge, as in Theorem 5.

**Proof.** See the Appendix.

**Remarks:** The limiting distribution is once again independent of $\alpha$.

The function $h$ exists straightforwardly for most commonly used estimators, including maximum likelihood and least squares.

So as to be sure that the integral in the first line of (14) 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

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

18
where $D_\theta$ is the operator that takes the gradient of its operand with respect to $\theta$. Similarly,

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

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.

In general, the limiting distribution given by (14) 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, then even in finite samples, the statistic $G_\alpha$ 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, preferably by maximum likelihood, after which the statistic of interest, which we denote by $\hat{\tau}$, is computed, whether it is (10) 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 $\theta_j^*$, $j = 1, \ldots, B$, 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 (10) 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 $B = 999$. 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),

$$AD = n \int_{-\infty}^{\infty} \left[ \frac{(\hat{F}(x) - F(x, \hat{\theta}))^2}{F(x, \hat{\theta})(1 - F(x, \hat{\theta}))} \right] dF(x, \hat{\theta}),$$

the Cramér-von-Mises statistic given by

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

the Kolmogorov-Smirnov statistic

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

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

$$P = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i},$$
where $O_i$ is the observed number of observations in the $i^{th}$ histogram interval, $E_i$ is the expected number in the $i^{th}$ histogram interval and $m$ is the number of histogram intervals.\(^4\) Then we provide an application using a UK data set on income distribution.

\section*{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 (10) 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 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) distribution is skewed to the left and Beta(2,5) is skewed to the right, while

\(^4\) 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.
the normal distribution is of course unskewed. As can be deduced from (10), in the first case the $G_\alpha$ statistic decreases as $\alpha$ increases, whereas in the second case it increases with $\alpha$.

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

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, $df = 2, 4, 6, \ldots, 20$. 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
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.

<table>
<thead>
<tr>
<th>$F_0$</th>
<th>Standard tests</th>
<th>$G_\alpha$ test with $\alpha =$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AD</td>
<td>CVM</td>
</tr>
<tr>
<td>N(0,1)</td>
<td>5.3</td>
<td>5.2</td>
</tr>
<tr>
<td>t(20)</td>
<td>7.7</td>
<td>7.3</td>
</tr>
<tr>
<td>t(18)</td>
<td>8.9</td>
<td>8.3</td>
</tr>
<tr>
<td>t(16)</td>
<td>9.9</td>
<td>8.9</td>
</tr>
<tr>
<td>t(14)</td>
<td>9.8</td>
<td>8.8</td>
</tr>
<tr>
<td>t(12)</td>
<td>13.5</td>
<td>12.0</td>
</tr>
<tr>
<td>t(10)</td>
<td>15.2</td>
<td>12.8</td>
</tr>
<tr>
<td>t(8)</td>
<td>22.3</td>
<td>19.0</td>
</tr>
<tr>
<td>t(6)</td>
<td>37.5</td>
<td>33.0</td>
</tr>
<tr>
<td>t(4)</td>
<td>64.3</td>
<td>59.9</td>
</tr>
<tr>
<td>t(2)</td>
<td>98.0</td>
<td>97.6</td>
</tr>
<tr>
<td>$\Gamma(20)$</td>
<td>25.2</td>
<td>21.9</td>
</tr>
<tr>
<td>$\Gamma(18)$</td>
<td>28.3</td>
<td>25.1</td>
</tr>
<tr>
<td>$\Gamma(16)$</td>
<td>30.9</td>
<td>27.2</td>
</tr>
<tr>
<td>$\Gamma(14)$</td>
<td>34.5</td>
<td>30.3</td>
</tr>
<tr>
<td>$\Gamma(12)$</td>
<td>41.3</td>
<td>36.6</td>
</tr>
<tr>
<td>$\Gamma(10)$</td>
<td>48.9</td>
<td>42.4</td>
</tr>
<tr>
<td>$\Gamma(8)$</td>
<td>58.1</td>
<td>51.7</td>
</tr>
<tr>
<td>$\Gamma(6)$</td>
<td>72.7</td>
<td>65.4</td>
</tr>
<tr>
<td>$\Gamma(4)$</td>
<td>88.5</td>
<td>82.1</td>
</tr>
<tr>
<td>$\Gamma(2)$</td>
<td>99.8</td>
<td>99.3</td>
</tr>
</tbody>
</table>

Among the standard goodness-of-fit tests, Table 2 shows that the AD statis-
tic 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.\(^5\) 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

---

\(^5\) 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.
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>
<thead>
<tr>
<th>nosb</th>
<th>AD</th>
<th>CVM</th>
<th>KS</th>
<th>P</th>
<th>$G_\alpha$ test with $\alpha =$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>20.4</td>
<td>18.2</td>
<td>14.5</td>
<td>9.4</td>
<td>32.2 33.7 25.7 21.3 19.3 17.4 12.4</td>
</tr>
<tr>
<td>100</td>
<td>33.7</td>
<td>30.2</td>
<td>23.1</td>
<td>11.4</td>
<td>46.0 49.0 37.8 33.3 31.0 28.2 18.1</td>
</tr>
<tr>
<td>200</td>
<td>56.2</td>
<td>51.5</td>
<td>40.6</td>
<td>17.4</td>
<td>65.7 70.3 59.3 55.5 53.1 50.1 36.1</td>
</tr>
<tr>
<td>300</td>
<td>73.9</td>
<td>69.4</td>
<td>56.9</td>
<td>24.6</td>
<td>81.0 84.3 76.4 73.0 71.0 68.1 55.4</td>
</tr>
<tr>
<td>400</td>
<td>84.3</td>
<td>80.2</td>
<td>68.5</td>
<td>31.8</td>
<td>89.0 91.5 85.7 83.5 82.2 79.9 69.2</td>
</tr>
<tr>
<td>500</td>
<td>90.6</td>
<td>87.7</td>
<td>77.3</td>
<td>38.7</td>
<td>93.8 95.0 91.5 90.0 89.1 87.5 79.5</td>
</tr>
</tbody>
</table>

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

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

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

the distribution function is

$$F_{SM}(x) = 1 - \left(1 + \frac{x}{b}\right)^p$$

with parameters $b = 100$, $a = 2.8$, and $p = 1.7$. 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.
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.

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 dataset. 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. Despite the name of the dataset, it covers the entire income distribution. 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 $(398.28, 253.75)$. Figure 2 shows a kernel-density estimate
Figure 2: Density of the empirical distribution of incomes

of the empirical distribution, from which it can be seen that there is a very
long right-hand tail, as usual with income distributions.

We test the goodness-of-fit of a number of distributions often used as para-
metric models of income distributions. We can immediately dismiss the Pareto
distribution, the density of which is a strictly decreasing function for arguments
greater than the lower bound of its support. First out of more serious possibil-
ities, we consider the lognormal distribution. In Table 5, we give the statistics
and bootstrap $P$ values, with 999 bootstrap samples used to compute them,
for the standard goodness-of-fit tests, and then, in Table 6, the $P$ values for
the $G_\alpha$ tests.
Table 5: Standard goodness-of-fit tests: bootstrap $P$ values, $H_0 : x \sim \text{lognormal}$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>statistic</td>
<td>1.16e21</td>
<td>9.48e8</td>
<td>7.246</td>
<td>7.090</td>
<td>7.172</td>
<td>7.453</td>
<td>8.732</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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

Every test rejects the null hypothesis that the true distribution is lognormal at any reasonable significance level.

Next, we tried the Singh-Maddala distribution, which has been shown to mimic observed income distributions in various countries, as shown by Brachman et al (1996). Table 7 presents the results for the standard goodness-of-fit tests; Table 8 results for the $G_\alpha$ tests. If we use standard goodness-of-fit statistics, we would not reject the Singh-Maddala distribution in most cases, except for the Anderson-Darling statistic at the 5% level.

Conversely, if we use $G_\alpha$ goodness-of-fit statistics, we would reject the Singh-Maddala distribution in all cases at the 5% level. Our previous simulation study shows $G_\alpha$ and AD have better finite sample properties. This leads
<table>
<thead>
<tr>
<th>test</th>
<th>AD</th>
<th>CVM</th>
<th>KS</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>statistic</td>
<td>0.644</td>
<td>0.050</td>
<td>0.010</td>
<td>13.37</td>
</tr>
<tr>
<td>p-value</td>
<td>0.028</td>
<td>0.274</td>
<td>0.305</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Table 7: Standard goodness-of-fit tests: bootstrap $P$ values, $H_0 : x \sim SM$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>statistic</td>
<td>164.3</td>
<td>1.362</td>
<td>0.441</td>
<td>0.404</td>
<td>0.390</td>
<td>0.382</td>
<td>0.398</td>
</tr>
<tr>
<td>p-value</td>
<td>0.002</td>
<td>0</td>
<td>0.006</td>
<td>0.011</td>
<td>0.013</td>
<td>0.014</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Table 8: $G_\alpha$ goodness-of-fit tests: bootstrap $P$ values, $H_0 : x \sim SM$.

us to conclude that the Singh-Maddala distribution is not a good fit, contrary to the conclusion from standard goodness-of-fit tests only.

Finally, we tested goodness of fit for the Dagum distribution, for which the distribution function is

$$F_D(x) = \left[1 + \left(\frac{b}{x}\right)^a\right]^{-p};$$

see Dagum (1977) and Dagum (1980). Both this distribution and the Singh-Maddala are special cases of the generalised beta distribution of the second kind, introduced by McDonald (1984). For further discussion, see Kleiber (1996), where it is remarked that the Dagum distribution usually fits real income distributions better than the Singh-Maddala. The results, in Tables 9

31
and 10, indicate clearly that, at the 5% level of significance, we can reject the null hypothesis that the data were drawn from a Dagum distribution on the basis of the Anderson-Darling test, the Pearson chi-square, and, still more conclusively, for all of the $G_\alpha$ tests. For this dataset, therefore, although we can reject both the Singh-Maddala and the Dagum distributions, the latter fits less well than the former.

For all three of the lognormal, Singh-Maddala, and Dagum distributions, the $G_\alpha$ statistics decrease with $\alpha$ except for the higher values of $\alpha$. This suggests that the empirical distribution is more skewed to the left than any of the distributions fitted to one of the families. Figure 3 shows kernel density estimates of the empirical distribution and the best fits from the lognormal, Singh-Maddala, and Dagum families. The range of income is smaller than
that in Figure 2, so as to make the differences clearer. The poorer fit of the lognormal is clear, but the other two families provide fits that seem reasonable to the eye. It can just be seen that, in the extreme left-hand tail, the empirical distribution has more mass than the fitted distributions.

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 (11) and (45) 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. However, simulations, not reported here, show that, even in conjunction with an appropriate bootstrap procedure, use of the maximum value leads to greater size distortion than for $G_\alpha$ for any single value of $\alpha$. 
Appendix of Proofs

Proof of Theorem 1. Axioms 1 to 4 imply that $\succeq$ can be represented by a continuous function $\Phi : \mathbb{Z}^n \rightarrow \mathbb{R}$ that is increasing in $|u_i - v_i|$, $i = 1, \ldots, n$. By Axiom 3, part (a) of the result follows from Theorem 5.3 of Fishburn (1970). This theorem says further that the functions $\phi_i$ are unique up to similar positive linear transformations; that is, the representation of the weak ordering is preserved if $\phi_i(z)$ is replaced by $a_i + b\phi_i(z)$ for constants $a_i$, $i = 1, \ldots, n$ and a constant $b > 0$. We may therefore choose to define the $\phi_i$ such that $\phi_i(0, 0) = 0$ for all $i = 1, \ldots, n$.

Now take $z'$ and $z$ in as specified in Axiom 4. From (1), it is clear that $z \sim z'$ 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 be true only 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 an instance of the first fundamental Pexider functional equation. Its solution implies that $\phi_i(u, u) = a_i + b_i u$. But above we chose to set $\phi_i(0, 0) = 0$, which implies that $a_i = 0$, and that $\phi_i(u, u) = b_i u$. 36
This is equation (2). □

**Proof of Theorem 2.** The function $\Phi$ introduced in the proof of Theorem 1 can, by virtue of (1), be chosen as

$$\Phi(z) = \sum_{i=1}^{n} \phi_i(z_i). \quad (15)$$

Then the relation $z \sim z'$ implies that $\Phi(z) = \Phi(z')$. By Axiom 5, it follows that, if $\Phi(z) = \Phi(z')$, then $\Phi(tz) = \Phi(tz')$, which means that $\Phi$ is a homothetic function. Consequently, there exists a function $\theta : \mathbb{R} \to \mathbb{R}$ that is increasing in its second argument, such that

$$\sum_{i=1}^{n} \phi_i(tz_i) = \theta(t, \sum_{i=1}^{n} \phi_i(z_i)). \quad (16)$$

The additive structure of $\Phi$ implies further that there exists a function $\psi : \mathbb{R} \to \mathbb{R}$ such that, for each $i = 1, \ldots, n$,

$$\phi_i(tz_i) = \psi(t)\phi_i(z_i). \quad (17)$$

To see this, choose arbitrary distinct values $j$ and $k$ and set $u_i = v_i = 0$ for all
\( i \neq j, k \). Then, since \( \phi_i(0,0) = 0 \), (16) becomes

\[
\phi_j(tu_j,tv_j) + \phi_k(tu_k,tv_k) = \theta(t,\phi_j(u_j,v_j) + \phi_k(u_k,v_k))
\]  

(18)

for all \( t > 0 \), and for all \((u_j,v_j),(u_k,v_k) \in Z\). Let us fix values for \( t, v_j, \) and \( v_k \), and consider (18) as a functional equation in \( u_j \) and \( u_k \). As such, it can be converted to a Pexider equation, as follows. First, let \( f_i(u) = \phi_i(tu,tv_i) \), \( g_i(u) = \phi_i(u,v_i) \) for \( i = j, k \), and \( h(x) = \theta(t,x) \). With these definitions, equation (18) becomes

\[
f_j(u_j) + f_k(u_k) = h(g_j(u_j) + g_k(u_k)).
\]

(19)

Next, let \( x_i = g_i(u_i) \) and \( \gamma_i(x) = f_i(g_i^{-1}(x)) \), for \( i = j, k \). This transforms (19) into

\[
\gamma_j(x_j) + \gamma_k(x_k) = h(x_j + x_k),
\]

which is an instance of the first fundamental Pexider equation, with solution

\[
\gamma_i(x_i) = a_0x_i + a_i, \quad i = j, k, \quad h(x) = a_0x + a_j + a_k,
\]

(20)

where the constants \( a_0, a_j, \) and \( a_k \) may depend on \( t, v_j, \) and \( v_k \). In terms of the functions \( f_i \) and \( g_i \), (20) implies that \( f_i(u_i) = a_0g_i(u_i) + a_i \), or, with all
possible functional dependencies made explicit,

\[
\phi_j(tu_j, tv_j) = a_0(t, v_j, v_k)\phi_j(u_j, v_j) + a_j(t, v_j, v_k) \quad \text{and} \quad (21)
\]

\[
\phi_k(tu_k, tv_k) = a_0(t, v_j, v_k)\phi_k(u_k, v_k) + a_k(t, v_j, v_k). \quad (22)
\]

If we construct an equation like (21) for \( j \) and another index \( l \neq j, k \), we get

\[
\phi_j(tu_j, tv_j) = d_0(t, v_j, v_l)\phi_j(u_j, v_j) + d_j(t, v_j, v_l) \quad (23)
\]

for functions \( d_0 \) and \( d_j \) that depend on the arguments indicated. But, since the right-hand sides of (21) and (23) are equal, that of (21) cannot depend on \( v_k \), since that of (23) does not. Thus \( a_j \) can depend at most on \( t \) and \( v_j \), while \( a_0 \), which is the same for both \( j \) and \( k \), can depend only on \( t \); we write \( a_0 = \psi(t) \). Thus equations (21) and (22) both take the form

\[
\phi_i(tu_i, tv_i) = \psi(t)\phi_i(u_i, v_i) + a_i(t, v_i), \quad (24)
\]

and this must be true for any \( i = 1, \ldots, n \), since \( j \) and \( k \) were chosen arbitrarily.

Now let \( u_i = v_i \), and then, since by (2) we have \( \phi_i(v_i, v_i) = b_iv_i \) and
\[ \phi_i(tv_i, tv_i) = tb_i v_i, \text{ equation (21) gives} \]

\[ a_i(t, v_i) = (t - \psi(t)) b_i v_i, \quad i = j, k. \quad (25) \]

Define the function \( \lambda_i(u_i, v_i) = \phi_i(u_i, v_i) - b_i v_i. \) This definition along with (2) implies that \( \lambda_i(u_i, u_i) = 0. \) Equation (24) can be written, with the help of (25), as

\[ \lambda_i(tu_i, tv_i) = \psi(t) \lambda_i(u_i, v_i), \]

where the function \( a_i(v_i, t) \) no longer appears. Then, in view of Aczél and Dhombres (1989), page 346 there must exist \( c \in \mathbb{R} \) and a function \( h_i : \mathbb{R}_+ \to \mathbb{R} \) such that

\[ \lambda_i(u_i, v_i) = u_i^c h_i(v_i/u_i). \quad (26) \]

From (26) it is clear that

\[ 0 = \lambda_i(u_i, u_i) = u_i^c h_i(1), \]

so that \( h_i(1) = 0. \)

We can now see that the assumption that the function \( a_i(t, v_i) \) is not identically equal to zero leads to a contradiction. For this assumption implies that neither \( \psi(t) - t \) nor \( b_i \) can be identically zero. Then, from (26) and the
definition of $\lambda_i$, we would have

$$
\phi_i(u_i, v_i) = u_i h_i(v_i/u_i) + b_i v_i.  \tag{27}
$$

With (27), equation (16) can be satisfied only if $c = 1$, as otherwise the two terms on the right-hand side of (27) are homogeneous with different degrees. But, if $c = 1$, both $\phi(u_i, v_i)$ and $\lambda_i(u_i, v_i)$ are homogeneous of degree 1, which means that $\psi(t) = t$, in contradiction with our assumption.

It follows that $a_i(t, v_i) = 0$ identically. If $\psi(t) = t$, we have $c = 1$, and equation (27) becomes

$$
\phi_i(u_i, v_i) = u_i h_i(v_i/u_i) + b_i v_i.  \tag{28}
$$

If $\psi(t)$ is not identically equal to $t$, $b_i$ must be zero for all $i$, and (27) becomes

$$
\phi_i(u_i, v_i) = u_i^c h_i(v_i/u_i).  \tag{29}
$$

Equations (28) and (29) imply the result (4).

**Proof of Theorem 3.** With Axiom 6 we may rule out the case in which the $b_i = 0$ in (4), according to which we would have $\phi_i(u_i, v_i) = u_i^c h_i(v_i/u_i)$ with $h_i(1) = 0$ for all $i = 1, \ldots, n$. To see this, note that, since we let $\phi_i(0,0) = 0
without loss of generality, and because \( \phi_i \) is increasing in \( |u_i - v_i| \), \( \phi_i(u_i, v_i) > 0 \) unless \((u_i, v_i) = (0, 0)\). Thus \( h_i(x) \) is positive for all \( x \neq 1 \), and is decreasing for \( x < 1 \) and increasing for \( x > 1 \). Now take the special case in which, in distribution \( z_0' \), the discrepancy takes the same value \( r \) in all \( n \) classes. If \((u_i, v_i)\) represents a typical component in \( z_0 \), then \( z_0 \sim z_0' \) implies that

\[
\sum_{i=1}^{n} u_i^e h_i(r) = \sum_{i=1}^{n} u_i^e h_i(v_i/u_i). 
\]

(A30)

Axiom 6 requires that, in addition,

\[
\sum_{i=1}^{n} u_i^e h_i(tr) = \sum_{i=1}^{n} u_i^e h_i(tv_i/u_i) 
\]

(A31)

Choose \( t \) such that \( tr = 1 \). Then the left-hand side of (A31) vanishes. But, since \( h_i(x) > 0 \) for \( x \neq 1 \), the right-hand side is positive, which contradicts the assumption that the \( b_i \) are zero. Consequently, the \( \phi_i \) are given by the representation (28), where \( c = 1 \). Let \( g_i(x) = h_i(x) + b_i x \), and define \( s_i = v_i/u_i \).

Then we may write (28) as \( \phi_i(u_i, v_i) = u_i g_i(s_i) \). Note that \( g_i(1) = b_i \) since \( h_i(1) = 0 \). Axiom 6 states that

\[
\sum_{i=1}^{n} u_i g_i(s_i) = \sum_{i=1}^{n} u_i g_i(r) \quad \text{implies} \quad \sum_{i=1}^{n} u_i g_i(ts_i) = \sum_{i=1}^{n} u_i g_i(tr). 
\]

(A32)
Define the function $\chi$ as the inverse in $x$ of the function $\sum_{i=1}^{n} u_i g_i(x)$. The first equation in (32) then implies that $r = \chi(\sum_{i=1}^{n} u_i g_i(s_i))$, and the second that $tr = \chi(\sum_{i=1}^{n} u_i g_i(ts_i))$. It follows that

$$\chi(\sum_{n=1}^{n} u_i g_i(ts_i)) = t\chi(\sum_{i=1}^{n} u_i g_i(s_i)).$$

Therefore the function $\chi(\sum_{i=1}^{n} u_i g_i(s_i))$ is homogeneous of degree one in the $s_i$, whence the function $\sum_{i=1}^{n} u_i g_i(s_i)$ is homothetic in the $s_i$. We have

$$\sum_{i=1}^{n} u_i g_i(ts_i) = \theta(t, \sum_{i=1}^{n} u_i g_i(s_i))$$

where $\theta(t, x) = \chi^{-1}(t\chi(x))$.

For fixed values of the $u_i$, make the definitions $f_i(s_i) = u_i g_i(ts_i)$, $e_i(s_i) = u_i g_i(s_i)$, $h(x) = \theta(t, x)$, $\gamma_i(x) = f_i(e_i^{-1}(x))$, $x_i = e_i(s_i)$. Then by an argument exactly like that in the proof of Theorem 2, we conclude that

$$\gamma_i(x_i) = a_0(t)x_i + a_i(t, u_i), \quad \text{and} \quad h(x) = a_0(t)x + \sum_{i=1}^{n} a_i(t, u_i).$$

With our definitions, this means that

$$u_i g_i(ts_i) = a_0(t)u_i g_i(s_i) + a_i(t, u_i). \quad (33)$$
Let $s_i = 1$. Then, since $g_i(1) = b_i$, (33) gives $a_i(t, u_i) = u_i(g_i(t) - a_0(t)b_i)$, and with this (33) becomes

$$g_i(ts_i) = a_0(t)(g_i(s_i) - b_i) + g_i(t)$$

(34)

as an identity in $t$ and $s_i$. The identity looks a little simpler if we define $k_i(x) = g_i(x) - b_i$, which implies that $k_i(1) = 0$. Then (34) can be written as

$$k_i(ts_i) = a_0(t)k_i(s_i) + k_i(t).$$

(35)

The remainder of the proof relies on the following lemma.

**Lemma 1** The general solution of the functional equation $k(ts) = a(t)k(s) + k(t)$ with $t > 0$ and $s > 0$, under the condition that neither $a$ nor $k$ is identically zero, is $a(t) = t^\alpha$ and $k(t) = c(t^\alpha - 1)$, where $\alpha$ and $c$ are real constants.

**Proof.** Let $t = s = 1$. The equation is $k(1) = a(1)k(1) + k(1)$, which implies that $k(1) = 0$ unless $a(1) = 0$. But if $a(1) = 0$, then the equation gives $k(s) = k(1)$ for all $s > 0$, which in turn implies that $k(1) = k(1)(a(t) + 1)$, which implies that $a(t) = 0$ identically, or that $k(1) = k(t) = 0$. Since we exclude the trivial solutions with $a$ or $k$ identically zero, we must have $a(1) \neq 0$ and $k(1) = 0$. 

44
Since \( k(ts) = k(st) \), the functional equation implies that

\[
a(t)k(s) + k(t) = a(s)k(t) + k(s), \quad \text{or} \quad k(s)(a(t) - 1) = k(t)(a(s) - 1),
\]

or

\[
\frac{k(t)}{a(t) - 1} = \frac{k(s)}{a(s) - 1} = c,
\]

for some real constant \( c \). Thus \( k(t) = c(a(t) - 1) \), and substituting this in the original functional equation and dividing by \( c \) gives

\[
a(ts) - 1 = a(t)(a(s) - 1) + a(t) - 1 = a(t)a(s) - 1,
\]

so that \( a(ts) = a(t)a(s) \). This is the fourth fundamental Cauchy functional equation, of which the general solution is \( a(t) = t^\alpha \), for some real \( \alpha \). It follows immediately that \( k(t) = c(t^\alpha - 1) \), as we wished to show. \( \blacksquare \)

**Proof of Theorem 3 (continued)**

The lemma and equation (35) imply that \( a_0(x) = x^\alpha \) and \( k_i(x) = c_i(x^\alpha - 1) \).

Since \( g_i(x) = k_i(x) + b_i = c_i(x^\alpha - 1) + b_i \) and \( \phi_i(u_i, v_i) = u_i g(v_i/u_i) \), we see that

\[
\phi_i(u_i, v_i) = u_i [\delta_i + c_i(v_i/u_i)^\alpha] = \delta_i u_i + c_i u_i^{1-\alpha} v_i^\alpha
\]

(36)
where $\delta_i = b_i - c_i$. Note that $c_i > 0$ in order that $\phi(u_i, v_i) > 0$ for all $u_i, v_i \neq (0, 0)$, but that $\delta_i$ may take on either sign, or may be zero. Equation (36) gives the result (5) of the theorem. ■

**Proof of Theorem 4.**

Let $\bar{u} = \sum_{i=1}^{n} u_i$ and $\bar{v} = \sum_{i=1}^{n} v_i$. Given the result of Theorem 3, we may write

$$\Phi(z) = \tilde{\phi} \left( \sum_{i=1}^{n} [\delta_i u_i + c_i u_i^{1-\alpha} v_i^{\alpha}] ; \bar{u}, \bar{v} \right),$$

where $\bar{u}$ and $\bar{v}$ are parameters of the function $\tilde{\phi}$ that is the counterpart of $\phi$ in (5). It is reasonable to require that $\Phi(z)$ should be zero when $z$ represents a "perfect fit". A narrow interpretation of zero discrepancy is that $v_i = u_i$, $i = 1, \ldots, n$. In this case, we see from (37) that

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

recall that $\delta_i + c_i = b_i$. Equation (38) is an identity in the $u_i$, which means that the function $\sum_{i=1}^{n} b_i u_i$ of the $u_i$ is a function of $\bar{u}$ alone for any choice of the $u_i$. This is possible only if $b_i = b$, and so the aggregate discrepancy index must be based on individual terms that all use the same value for $b_i$.

Scale invariance implies that $\Phi(kz) = \Phi(z)$ for all $k > 0$, and from (37)
this means that, identically in the $u_i$, the $v_i$, and $k$,

$$\bar{\phi}(k[b\bar{u} + \sum_{i=1}^{n} c_i(u_i^{1-\alpha}v_i^\alpha - u_i)]; k\bar{u}, k\bar{v}) = \bar{\phi}(b\bar{u} + \sum_{i=1}^{n} c_i(u_i^{1-\alpha}v_i^\alpha - u_i); \bar{u}, \bar{v}),$$

This implies that $\bar{\phi}$ is homogeneous of degree zero in its three arguments. But the value of $\Phi(z)$ is also unchanged if we rescale only the $v_i$, multiplying them by $k$, and so the expression

$$\bar{\phi}(b\bar{u} + \sum_{i=1}^{n} c_i(k^\alpha u_i^{1-\alpha}v_i^\alpha - u_i); \bar{u}, k\bar{v})$$

is equal for all $k$ to its value for $k = 1$. If $u_i = v_i$ for all $i = 1, \ldots, n$, then we have

$$\bar{\phi}(b\bar{u} + (k^\alpha - 1) \sum_{i=1}^{n} c_i u_i; \bar{u}, k\bar{u}) = 0$$

identically in the $u_i$ and $k$, and this is possible only if $c_i = c$. Thus the discrepancy index can be written as

$$\bar{\phi}((b - c)\bar{u} + c \sum_{i=1}^{n} u_i^{1-\alpha}v_i^\alpha; \bar{u}, \bar{v}).$$
that is, a function of \( \sum_{i=1}^{n} u_i^{1-\alpha} v_i^\alpha, \bar{u}, \text{ and } \bar{v} \), which we now write as

\[
\psi_1\left( \sum_{i=1}^{n} u_i^{1-\alpha} v_i^\alpha, \bar{u}, \bar{v} \right).
\]

This new function \( \psi_1 \) is still homogeneous of degree zero in its three arguments, and so it can be expressed as a function of only two arguments, as follows:

\[
\psi_2\left( \frac{1}{\bar{u}} \sum_{i=1}^{n} u_i^{1-\alpha} v_i^\alpha, \frac{\bar{v}}{\bar{u}} \right).
\] (39)

The value of \( \psi_2 \) is unchanged if we rescale the \( v_i \) while leaving the \( u_i \) unchanged, and so we have, identically,

\[
\psi_2\left( k^{\alpha} \frac{1}{\bar{u}} \sum_{i=1}^{n} u_i^{1-\alpha} v_i^\alpha, \bar{v} \right) = \psi_2\left( \frac{1}{\bar{u}} \sum_{i=1}^{n} u_i^{1-\alpha} v_i^\alpha, \frac{\bar{v}}{\bar{u}} \right),
\]

which we can write formally as a property of \( \psi_2 \): \( \psi_2(k^{\alpha}x, ky) = \psi_2(x, y) \) identically in \( k, x, \) and \( y \). Let \( \psi_3(x, y) = \psi_2(x^\alpha, y) \) be the definition of the function \( \psi_3 \), so that \( \psi_3(kx, ky) = \psi_2(k^{\alpha}x^\alpha, ky) = \psi_2(x^\alpha, y) = \psi_3(x, y) \). Thus \( \psi_3 \) is homogeneous of degree zero in its two arguments, and so we may define \( \psi_4 \) by the relation \( \psi_3(x, y) = \psi_4(x/y) \), which is equivalent to \( \psi_2(x, y) = \psi_4(x^{1/\alpha}/y) = \psi_4(x/y^\alpha) \), where we define two functions, each of one scalar argument, \( \psi_4 \) and \( \psi \).
The discrepancy index in the form (39) is therefore given by
\[ \psi \left( \frac{1}{\bar{u}} \sum_{i=1}^{n} u_i^{1-\alpha} v_i^{\alpha} \left( \frac{1}{\bar{v}} - 1 \right) \right) = \psi \left( \sum_{i=1}^{n} \left[ \frac{u_i}{\bar{u}} \right]^{1-\alpha} \left[ \frac{v_i}{\bar{v}} \right]^{\alpha} \right). \]

The result (6) follows if the function \( \phi \) is defined so that \( \phi(x) = \psi(nx) \). In order for the discrepancy index to be zero for a perfect fit with \( u_i = v_i \), we require that \( \psi \left( \frac{1}{\bar{u}} \sum_{i=1}^{n} u_i \right) = \psi(1) = 0 \), or \( \phi(n) = 0 \). \( \blacksquare \)

Proof of Theorem 5.

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
\[ \sqrt{n} \left( \hat{Q}_n(p) - Q(p) \right) \rightsquigarrow - \frac{B \circ F(Q(p))}{f(Q(p))}, \]
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 as defined in the statement of the theorem.

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 (40) in this case reduces to

\[ \sqrt{n}(u_{\lfloor np \rfloor} - p) \rightsquigarrow B(p). \] (41)

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

\[ \sqrt{n}(u_i - t_i) \rightsquigarrow B(t_i). \] (42)

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 (41), we see that the variance of \( n^{1/2}z_i \) is \( t_i(1 - t_i) \) plus a term that vanishes as \( n \to \infty \). Thus \( z_i = O_p(n^{-1/2}) \) as \( n \to \infty \). 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 \( n \to \infty \). We see

\(^6\) In fact, the true variance of \( z_i \) is \( t_i(1 - t_i)/(n + 2) \).
that

\[ G_{\alpha}(F, \hat{F}) = \frac{1}{\alpha(\alpha - 1)} \mu_u(1/2)^{1-\alpha} \sum_{i=1}^{n} \left[ t_i \left(1 + \frac{z_i}{t_i}\right)^\alpha - \mu_u(1/2)^{1-\alpha}\right]. \quad (43) \]

Now, by Taylor’s theorem,

\[ t_i \left(1 + \frac{z_i}{t_i}\right)^\alpha = t_i + \alpha z_i + \frac{\alpha(\alpha - 1)}{2} \frac{z_i^2}{t_i} + \frac{\alpha(\alpha - 1)(\alpha - 2)}{6} \frac{(\theta_i z_i)^3}{t_i^2}, \quad (44) \]

where \(0 \leq \theta_i \leq 1, i = 1, \ldots, n\), and so

\[ \sum_{i=1}^{n} t_i \left(1 + \frac{z_i}{t_i}\right)^\alpha = \frac{n}{2} + n\alpha \bar{z} + \frac{\alpha(\alpha - 1)}{2} \sum_{i=1}^{n} \frac{z_i^2}{t_i} + o_p(1), \quad (45) \]

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 (44) 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

$$\mu_u^{(1/2)^{1-\alpha}} = \frac{1}{2} \left( 1 + 2\alpha \bar{z} + 2\alpha(\alpha - 1)\bar{z}^2 + \frac{4\alpha(\alpha - 1)(\alpha - 2)}{3}(\theta_{\mu}\bar{z})^3 \right), \quad (46)$$

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 (45) and (46) 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 (43) we arrive at the result

$$G_\alpha(F, \hat{F}) = \sum_{i=1}^n \frac{z_i^2}{t_i} - 2n\bar{z}^2 + o_p(1). \quad (47)$$

It is striking that the leading-order term in (47) does not depend on $\alpha$. For finite $n$, $G_\alpha$ does of course depend on $\alpha$. Simulation shows that, even for $n$ as small as 10, the distributions of $G_\alpha$ and of the leading term in (47) 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 (44) and (46).
If the limiting asymptotic distribution of $G_\alpha$ exists, it is the same as that of the approximation in (47), 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 (42)) and then letting $n$ tend to infinity. For $\bar{z}$ first, we have

$$n^{1/2}\bar{z} = n^{-1/2}\sum_{i=1}^{n} z_i =_d n^{-1} \sum_{i=1}^{n} B(t_i) \sim \int_0^1 B(t) \, dt.$$  \hspace{1cm} (48)

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

$$\sum_{i=1}^{n} \frac{z_i^2}{t_i} =_d n^{-1} \sum_{i=1}^{n} \frac{B^2(t_i)}{t_i} \sim \int_0^1 \frac{B^2(t)}{t} \, dt.$$  \hspace{1cm} (49)

From (48) and (49), we see that the limiting distribution of $G_\alpha$ is that of

$$\int_0^1 \frac{B^2(t)}{t} \, dt - 2\left( \int_0^1 B(t) \, dt \right)^2,$$  \hspace{1cm} (50)

in agreement with (11) in the statement of the theorem. ■

**Proof of Theorem 6.**

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 (47) gives

\[
G_\alpha = \sum_{i=1}^{n} \left(\frac{z_i + g^\top(t_i, \theta)s(\theta)}{t_i}\right)^2 - 2\left(n^{-1/2} \sum_{i=1}^{n} (z_i + g^\top(t_i, \theta)s(\theta))\right)^2 + o_p(1). \tag{51}
\]

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 (51), note that, by the definition of the function \(h\), we have

\[
n^{1/2}(\hat{\theta} - \theta) = n^{1/2}s(\theta) = n^{-1/2} \sum_{i=1}^{n} h(x_i, \theta) + o_p(1). \tag{52}
\]

Our sample is supposed to be IID, and so in (52) we can sum over the order
statistics \( x(i) \). Then a short Taylor expansion gives

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

(53)

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.

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

\[
n^{-1} \sum_{i=1}^{n} h(Q(t_i, \theta), \theta) = \int_{0}^{1} h(Q(t, \theta), \theta) \, dt + O(n^{-1}) \\
= \int_{-\infty}^{\infty} h(x, \theta) \, dF(x, \theta) + O(n^{-1}) = O(n^{-1}),
\]

because the expectation of \( h(x, \theta) \) is zero. (The integration over the whole real line means in fact integration over the support of the distribution \( F \).) Thus the first term in the sum in (53) 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

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

(54)

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

\[
n^{-1} \sum_{i=1}^n g(t_i, \theta) = n^{-1} \sum_{i=1}^n p(Q(t_i, \theta), \theta)
\]

\[
= \int_0^1 p(Q(t,\theta),\theta) \, dt = \int_{-\infty}^\infty p(x,\theta) \, dF(x,\theta) = P(\theta).
\]

Using (54), we have

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

\[ n^{-1/2} \sum_{i=1}^{n} \left( z_i + g^\top (t_i, \theta) s(\theta) \right) \sim \int_{0}^{1} B(t) \, dt + P^\top(\theta) \int_{-\infty}^{\infty} h'(x, \theta) B(F(x, \theta)) \, dx. \]

(55)

Finally, we consider the first sum in (51). By arguments similar to those used above, we see that

\[ \sum_{i=1}^{n} \frac{(z_i + g^\top (t_i, \theta) s(\theta))^2}{t_i} \sim \int_{0}^{1} \frac{1}{t} \left[ B(t) + p^\top(Q(t, \theta), \theta) \times \int_{-\infty}^{\infty} h'(x, \theta) B(F(x, \theta)) \, dx \right]^2 \, dt. \]

(56)

By combining (51), (55), and (56), we get (14). □

References


Anderson, T. W. and D. A. Darling (1952). “Asymptotic Theory of Cer-


