TESTS FOR NORMAL MIXTURES BASED ON THE EMPIRICAL CHARACTERISTIC FUNCTION

Bernhard Klar and Simos G. Meintanis [1]

Institut für Mathematische Stochastik, Universität Karlsruhe,
Englerstr. 2, 76128 Karlsruhe, Germany
and
Department of Economics, National and Kapodistrian University of Athens,
8 Pesmazoglou Street, 105 59 Athens, Greece

Abstract. A goodness–of–fit test for two–component homoscedastic and homothetic mixtures of normal distributions is proposed. The tests are based on a weighted L2–type distance between the empirical characteristic function and its population counterpart, where in the latter, parameters are replaced by consistent estimators. Consequently the resulting tests are consistent against general alternatives. When moment estimation is employed and as the decay of the weight function tends to infinity the test statistics approach limit values, which are related to the first nonvanishing moment equation. The new tests are compared via simulation to other omnibus tests for mixtures of normal distributions, and are applied to several real data sets.

Keywords. Characteristic function, Goodness-of-fit test, Mixtures of Normal Distributions

1 Introduction and Summary

Mixtures of normal distributions have a long history in Statistics, dating back to the late nineteenth century and the writings of S. Newcomb and K. Pearson. Since then, they appear as models in diverse areas of applied research. Typically leptokurtic, skewed and multimodal data are modeled by considering an appropriate normal mixture. However,

[1]Corresponding author. Tel.: ++302103689814; fax: ++302103689809.
E–Mail address: simosmei@econ.uoa.gr.

even in the simplest of cases, the two–component normal mixture, one encounters serious theoretical as well as computational difficulties when attempting to perform basic statistical analysis such as parameter estimation and goodness–of–fit. A full–book treatment on the subject of mixture distributions (with particular emphasis on normal mixtures) is provided by Titterington et al. (1985). For recent developments on mixtures (including estimation of parameters, testing for the number of components, and mixture modeling), the reader is referred to the special issue of Computational Statistics & Data Analysis: CSDA (2003), 41, No. 3–4. Goodness–of–fit against general alternatives is an outstanding problem, and in Thode (2002) some methods are reviewed, most of them heuristic. Finite–sample results on the Kolmogorov–Smirnov statistic for homoscedastic two–component normal mixtures may be found in Neus and Mendell (2001).

Apart from moment–based or likelihood inference, recent years have witnessed an increasing interest in using empirical transforms as tools for statistical inference. Inference procedures based on statistical transforms, such as the empirical characteristic function and the empirical moment generating function, were introduced by Press (1972) (parameter estimation) and Heathcote (1972) (goodness–of–fit). More recent work includes, among others, Koutrouvelis and Meintanis (2002), Gürtler and Henze (2000a), Epps (1999), Kankainen and Ushakov (1998), and Koutrouvelis and Canavos (1997). A large part of the literature on the empirical characteristic function is covered in Ushakov (1999). For normal mixtures in particular, Quandt and Ramsey (1978) and Schmidt (1982) employ the empirical moment generating function, while Bryant and Paulson (1983) and Tran (1998) employ the empirical characteristic function in parameter estimation. The present paper studies a new class of omnibus tests for two–component normal mixtures which are based on the empirical characteristic function (ecf). To be specific, let $\mathcal{NM}$ denote the collection of all two–component normal mixtures. The characteristic function $\varphi(t;\vartheta)$ of any member in $\mathcal{NM}$, is given by

$$\varphi(t;\vartheta) = (1-\lambda)\exp\left(i\mu_1 t - \frac{1}{2}\sigma_1^2 t^2\right) + \lambda \exp\left(i\mu_2 t - \frac{1}{2}\sigma_2^2 t^2\right), \tag{1.1}$$

with $\vartheta = (\lambda, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)'$, where $0 \le \lambda < 1$ denotes the mixing parameter, $\mu_k \in \mathbb{R}$ and $\sigma_k^2 > 0$, $k = 1, 2$, denote the means and variances, respectively, of the normal components. The class $\mathcal{NM}_\sigma$, denotes the subset of homoscedastic two–component normal mixtures which results from $\mathcal{NM}$ when it is assumed that $\sigma_1^2 = \sigma_2^2 = \sigma^2$. In turn, the class $\mathcal{NM}_\mu$ denotes the subset of homothetic two–component normal mixtures which results from $\mathcal{NM}$ when it is assumed that $\mu_1 = \mu_2 = \mu$.

In what follows we use (1.1) to develop goodness–of–fit tests for the null hypotheses,

$$H_{0\sigma}: \text{ The law of } X \text{ is in } \mathcal{NM}_\sigma, \text{ for some } \lambda \in [0,1), \mu_1, \mu_2 \in \mathbb{R}, \text{ and } \sigma^2 > 0,$$

and

$$H_{0\mu}: \text{ The law of } X \text{ is in } \mathcal{NM}_\mu, \text{ for some } \lambda \in [0,1), \mu \in \mathbb{R}, \text{ and } \sigma_1^2, \sigma_2^2 > 0,$$

against general alternatives, where X is a random variable with characteristic function $\phi(t) = E[\exp(itX)]$. In particular, let

$$\phi_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(itX_j) = C_n(t) + iS_n(t),$$

be the *ecf* based on independent copies $X_1, X_2, \ldots, X_n$ of X. A general class of test statistics for assessing $H_{0\sigma}$ (resp. $H_{0\mu}$) is given by

$$T_n \;=\; n\int_{-\infty}^{\infty} |\phi_n(t) - \varphi(t;\hat{\vartheta})|^2 \, w(t)\, dt, \tag{1.2}$$

where $\varphi(t;\vartheta)$ results from (1.1) for $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (resp. $\mu_1 = \mu_2 = \mu$), $\hat{\vartheta}$ denotes an estimator of ϑ, and $w(t)$ is a nonnegative integrable weight function. Many properties of the proposed goodness–of–fit tests which reject $H_{0\sigma}$ and $H_{0\mu}$ for large values of T_n, such as consistency under fixed alternatives and the existence of a limit distribution of T_n under the null hypothesis, hold under very general conditions for the weight function $w(t)$. Here however we will focus on a parametric class of weight functions for which T_n takes a simple form. Specifically, we use $w(t) = \exp(-at^2)$, where $a > 0$ is a parameter, the role of which we discuss later. Denoting by $\Sigma_{n,a}$ the test statistic corresponding to $H_{0\sigma}$ and by $M_{n,a}$ the test statistic corresponding to $H_{0\mu}$, straightforward algebra yields

$$\Sigma_{n,a} = \frac{1}{n}\sqrt{\frac{\pi}{a}} \sum_{j,k=1}^{n} e^{-\frac{(X_j - X_k)^2}{4a}}$$

$$+n\sqrt{\frac{\pi}{a+\hat{\sigma}^2}}\left[\{1 - 2\hat{\lambda}(1-\hat{\lambda})\} + 2\hat{\lambda}(1-\hat{\lambda})e^{-\frac{(\hat{\mu}_1-\hat{\mu}_2)^2}{4(a+\hat{\sigma}^2)}}\right]$$

$$-2\sqrt{\frac{2\pi}{2a+\hat{\sigma}^2}}\sum_{j=1}^{n}\left[(1-\hat{\lambda})e^{-\frac{(X_j-\hat{\mu}_1)^2}{4a+2\hat{\sigma}^2}}+\hat{\lambda}e^{-\frac{(X_j-\hat{\mu}_2)^2}{4a+2\hat{\sigma}^2}}\right],$$

and

$$M_{n,a}=\frac{1}{n}\sqrt{\frac{\pi}{a}}\sum_{j,k=1}^{n}e^{-\frac{(X_j-X_k)^2}{4a}}$$

$$+n\left[(1-\hat{\lambda})^2\sqrt{\frac{\pi}{a+\hat{\sigma}_1^2}}+\hat{\lambda}^2\sqrt{\frac{\pi}{a+\hat{\sigma}_2^2}}+2\hat{\lambda}(1-\hat{\lambda})\sqrt{\frac{2\pi}{2a+\hat{\sigma}_1^2+\hat{\sigma}_2^2}}\right]$$

$$-2\sum_{j=1}^{n}\left[(1-\hat{\lambda})\sqrt{\frac{2\pi}{2a+\hat{\sigma}_1^2}}\,e^{-\frac{(X_j-\hat{\mu})^2}{4a+2\hat{\sigma}_1^2}}+\hat{\lambda}\sqrt{\frac{2\pi}{2a+\hat{\sigma}_2^2}}\,e^{-\frac{(X_j-\hat{\mu})^2}{4a+2\hat{\sigma}_2^2}}\right],$$

where in the case of testing $H_{0\sigma}$ (resp. $H_{0\mu}$), $\hat{\vartheta}=(\hat{\lambda},\hat{\mu}_1,\hat{\mu}_2,\hat{\sigma}^2)'$ (resp. $\hat{\vartheta}=(\hat{\lambda},\hat{\mu},\hat{\sigma}_1^2,\hat{\sigma}_2^2)'$) denotes an estimator of the vector parameter $\vartheta=(\lambda,\mu_1,\mu_2,\sigma^2)'$ (resp. $\vartheta=(\lambda,\mu,\sigma_1^2,\sigma_2^2)'$). Obviously these are readily amenable for computational purposes. Moreover if the estimators of the parameters satisfy the usual properties, i.e. that location estimators ($\hat{\mu}_1,\hat{\mu}_2$ and $\hat{\mu}$) are affine equivariant, scale estimators ($\hat{\sigma}$ and $\hat{\sigma}_1,\hat{\sigma}_2$) are location invariant and scale equivariant, and that the shape estimator $\hat{\lambda}$ is affine invariant, then both $\Sigma_{n,a}$ and $M_{n,a}$ are location invariant. An affine invariant version of the statistic for testing $H_{0\sigma}$ (resp. $H_{0\mu}$) results when $\Sigma_{n,a}$ (resp. $M_{n,a}$) is replaced by $\hat{\sigma}\Sigma_{n,a\hat{\sigma}^2}$ (resp. $\hat{\sigma}_1 M_{n,a\hat{\sigma}_1^2}$).

It should be noted at this point that the tests proposed herein are not confined to the case of two–component normal mixtures. In principle they are applicable to any type of mixture with two or more components. In fact, it is straightforward to generalize (1.2) to include many–component mixtures, with normal or non–normal components (Bryant and Paulson (1983) have provided such a general formula). However, it will be seen in the sequel that efficient parameter estimation remains a serious problem, weighing heavily on the performance of the proposed tests. Also, given the estimator of the parameter, the feasibility of closed form expressions for the test statistic would have to be considered individualy for each particular mixture, depending on the distribution of components. It is the functional form of the component characteristic function which will indicate the type of weight functions, if any, that renders the resulting test statistic in closed form.

A general problem when testing the hypothesis that the underlying distribution is some mixture distribution with unspecified parameters is that the null distribution of a test statistic usually depends on the unknown parameter vector ϑ. In order to have a test that maintains a nominal level of significance irrespective of the value of ϑ, it would be possible to use special tables or formulas for critical values, obtained by extensive simulations or by a numerical approximation of the asymptotic null distribution. These formulas differ with the nominal level of significance and the statistic used, and they depend either on the sample size and the estimated value of ϑ from given data, or in case that the asymptotic null distribution is used, only on the estimated value of ϑ.

Such formulas are used in the simulation study of Neus and Mendell (2001) for a two component homoscedastic normal mixture using the Kolmogorv-Smirnov test statistic with estimated parameters.

A different, and, in our opinion, preferable method is to use a parametric bootstrap. This idea of simulating the null distribution of a test statistic is now well-established, and it does not only lead to critical values, but also to approximate p-values. Hence, all simulations and data examples in this paper are performed using a parametric bootstrap procedure.

The paper is organized as follows. In Section 2, from $\Sigma_{n,a}$ and $M_{n,a}$ we derive the limit statistics as $a \to \infty$. The consistency of the test that rejects $H_{0\sigma}$ and $H_{0\mu}$ for large values of T_n against general alternatives, is deferred to the Appendix. The results of a Monte Carlo study are presented in Section 3, where the new tests are compared to other goodness–of–fit tests for normal mixtures. The paper concludes with real data examples, given in Section 4.

2 Limit Test Statistics

From (1.2) and assuming $w(t) = w(-t)$, straightforward algebra yields the representation

$$T_n \;=\; \int_{-\infty}^{\infty} Z_n^2(t;\hat{\vartheta})\, w(t)\, dt, \tag{2.1}$$

where

$$Z_n(t;\vartheta) = \sqrt{n}\left[\frac{1}{n}\sum_{j=1}^{n}[(\cos tX_j + \sin tX_j) - cs(t;\vartheta)]\right], \quad t \in \mathbb{R},$$

with

$$cs(t;\vartheta) = e^{-\frac{1}{2}\sigma^2 t^2}\left[(1-\lambda)(\cos\mu_1 t + \sin\mu_1 t) + \lambda(\cos\mu_2 t + \sin\mu_2 t)\right],$$

in the case of testing for $H_{0\sigma}$, and

$$cs(t;\vartheta) = [(1-\lambda)e^{-\frac{1}{2}\sigma_1^2 t^2} + \lambda e^{-\frac{1}{2}\sigma_2^2 t^2}]\,(\cos\mu t + \sin\mu t),$$

in the case of testing for $H_{0\mu}$. For $w(t) = \exp(-at^2)$, the value of the parameter a controls the rate of decay of the weight function. The representation in (2.1) suggests that this value influences to a large extent the value of the test statistic. Large values of a let the weight function decay rapidly, and consequently the test statistic is dominated by the behavior of $|\phi_n(t) - \varphi(t)|$ near $t = 0$. Since the tail behavior of a distribution is reflected on the behavior of its characteristic function around zero, putting most of the weight around $t = 0$ renders the test statistics powerful against distributions with markedly different tail behaviour than those in $\mathcal{NM}$.

It should be noted at this point that for any symmetric interval around the origin, there exist distinct distributions for which the corresponding characteristic functions coincide in this interval. It is not clear whether these distributions could have widely differing distribution functions. However there exist novel results which impose an upper bound on the Kolmogorov–Smirnov distance between pairs of distributions, and this bound is related to the distance between their characteristic functions. There may exist cases of distributions where the bound is conservative, and consequently the Kolmogorov–Smirnov distance between these distributions is well below the corresponding distance between their characteristic functions. In such cases, any ecf–based test will be much less informative that standard tests utilizing the empirical distribution function. The above remarks not withstanding and however interesting these counterexamples may be, they are in our opinion exceptional. For more information on this theoretical issue the interested reader is referred to Ushakov (1999), Section 2.9 and Appendix A.

Interestingly for large values of the parameter a, the test statistics $\Sigma_{n,a}$ and $M_{n,a}$ approach limit values. In fact, from (2.1) with $w(t) = \exp(-at^2)$ we have

$$T_n = \int_0^\infty g(t)\, e^{-at}\, dt,$$

where $g(t) = Z_n^2(\sqrt{t}; \hat{\vartheta})\, t^{-1/2}$. A tedious but straightforward Taylor expansion yields,

$$\begin{aligned} Z_n(\sqrt{t};\vartheta) = \sqrt{n}[t^{1/2}(&\bar{X}_n - \{(1-\lambda)\mu_1 + \lambda\mu_2\}) \\ &-\frac{t}{2!}(\overline{X_n^2} - \{(1-\lambda)\mu_1^2 + \lambda\mu_2^2 + \sigma^2\}) \\ &-\frac{t^{3/2}}{3!}(\overline{X_n^3} - \{(1-\lambda)\mu_1^3 + \lambda\mu_2^3 + 3(1-\lambda)\mu_1\sigma^2 + 3\lambda\mu_2\sigma^2\}) \\ &+\frac{t^2}{4!}(\overline{X_n^4} - \{(1-\lambda)\mu_1^4 + \lambda\mu_2^4 + 6(1-\lambda)\mu_1^2\sigma^2 + 6\lambda\mu_2^2\sigma^2 + 3\sigma^4\}) \\ &+\frac{t^{5/2}}{5!}(\overline{X_n^5} - \{(1-\lambda)\mu_1^5 + \lambda\mu_2^5 + 10(1-\lambda)\mu_1^3\sigma^2 + 10\lambda\mu_2^3\sigma^2 \\ &\qquad +15(1-\lambda)\mu_1\sigma^4 + 15\lambda\mu_2\sigma^4\})] \;+\; o(t^{5/2}), \end{aligned}$$

in the case of testing for $H_{0\sigma}$, with $\bar{X}_n = n^{-1}\sum_{j=1}^n X_j$, and $\overline{X_n^k} = n^{-1}\sum_{j=1}^n X_j^k$, $k = 2, \ldots$. If moment estimation is employed, see for example Cohen (1967), then the first four terms in $Z_n(t;\hat{\vartheta})$ vanish, as they represent the first four moment equations from which $\hat{\vartheta}$ is obtained. Hence

$$g(t) = \frac{1}{(5!)^2} t^{9/2} (\sqrt{n}\, \tilde{\mathcal{M}}_5)^2 + o(t^{9/2}), \tag{2.2}$$

where the coefficient of $t^{5/2}/(5!)$ in the expansion of $Z_n(\sqrt{t};\cdot)$ is denoted by $\tilde{\mathcal{M}}_5$ and coincides with the fifth moment equation. Likewise, in the case of testing for $H_{0\mu}$ we have

$$\begin{aligned} Z_n(\sqrt{t};\vartheta) = \sqrt{n}[t^{1/2}(\bar{X}_n - \mu) - \frac{t}{2!}(\overline{X_n^2} - \{(1-\lambda)\sigma_1^2 + \lambda\sigma_2^2 + \mu^2\}) \\ -\frac{t^{3/2}}{3!}(\overline{X_n^3} - \{3(1-\lambda)\mu\sigma_1^2 + 3\lambda\mu\sigma_2^2 + \mu^3\})] + o(t^{3/2}). \end{aligned}$$

If moment estimation is employed, then the first two terms in $Z_n(t;\hat{\vartheta})$ vanish, as they represent the first two moment equations from which $\hat{\vartheta}$ is obtained. (Note that since homothetic normal mixtures are symmetric, all central moments of odd order equal to zero, and therefore moment estimators are produced by considering the first, second, fourth and sixth moment equation). Hence

$$g(t) = \frac{1}{(3!)^2} t^{5/2} (\sqrt{n}\, \tilde{\mathcal{M}}_3)^2 + o(t^{5/2}), \tag{2.3}$$

where the coefficient of $t^{3/2}/(3!)$ in the expansion of $Z_n(\sqrt{t};\cdot)$ is denoted by $\tilde{\mathcal{M}}_3$ and coincides with the third moment equation.

An Abelian theorem for Laplace transforms (see Proposition 1.1 in Baringhaus et al. (2000)), along with (2.2) and (2.3) yields

$$\lim_{a\to\infty} a^{11/2}\ \Sigma_{n,a} \ = \ \frac{\Gamma(11/2)}{(5!)^2} \Sigma_{n,\infty}, \quad \text{with} \quad \Sigma_{n,\infty} = n\, \tilde{\mathcal{M}}_5^2, \tag{2.4}$$

$$\lim_{a\to\infty} a^{7/2}\ M_{n,a} \ = \ \frac{\Gamma(7/2)}{(3!)^2} M_{n,\infty}, \quad \text{with} \quad M_{n,\infty} = n\, \tilde{\mathcal{M}}_3^2. \tag{2.5}$$

Thus as $a \to \infty$, each test statistic, when suitably rescaled, and apart from an irrelevant constant factor, approaches a 'limit statistic'. Notice that since $\tilde{\mathcal{M}}_5$ (resp. $\tilde{\mathcal{M}}_3$) in $\Sigma_{n,\infty}$ (resp. $M_{n,\infty}$) is equivalent to the fifth (resp. third) moment equation, it vanishes under $H_{0\sigma}$ (resp. $H_{0\mu}$), as $n \to \infty$.

3 Simulations

This section presents the results of a Monte Carlo study conducted at an α–nominal level with 1000 replications to assess the power of the new tests. In order to avoid reliance on asymptotic critical values, and since the null distribution of the test statistic depends on the (unknown) value of the parameter ϑ, we used a parametric bootstrap procedure to obtain the critical point p_n of the test as follows:

Conditionally on the observed value of $\hat{\vartheta}$, generate B bootstrap samples (in all cases $\tilde{B} = (1-\alpha)B$ was an integer) from $\mathcal{NM}$, with $\vartheta = \hat{\vartheta}$. Calculate the value of the test statistic, say T_b^*, $(b = 1, 2, ..., B)$, for each bootstrap sample. Obtain p_n as $T^*_{(\tilde{B})}$, where $T^*_{(b)}$, $b = 1, 2, ..., B$, denotes the ordered values T_b^*–values, and reject H_0 if $T_n > p_n$. We have adapted the choice in Gürtler and Henze (2000b) and used the modified critical point $\tilde{p}_n = T^*_{(\tilde{B})} + (1-\alpha)(T^*_{(\tilde{B}+1)} - T^*_{(\tilde{B})})$, which leads to a more accurate empirical level of the test.

The new tests are compared with classical procedures that utilize the empirical distribution function (*edf*). Among the *edf* tests we considered the *Kolmogorov–Smirnov (KS)* test, the *Cramér–von Mises (CM)* test, the *Anderson–Darling (AD)* test and the *Watson (WA)* test. Denote by $F(x;\vartheta)$ the cdf of a distribution in $\mathcal{NM}$, and let $\hat{F}(x) = F(x,\hat{\vartheta})$. Writing $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ for the order statistics of $X_1, X_2, ..., X_n$, the KS-statistic is

$$KS = \max\{D^+, D^-\},$$

where

$$D^+ = \max_{j=1,2,\dots,n} \left\{\frac{j}{n} - \hat{F}(X_{(j)})\right\}, \quad D^- = \max_{j=1,2,\dots,n} \left\{\hat{F}(X_{(j)}) - \frac{j-1}{n}\right\}.$$

The statistics of Cramér-von Mises and Anderson-Darling are given by

$$CM = \frac{1}{12n} + \sum_{j=1}^{n} \left(\hat{F}(X_{(j)}) - \frac{2j-1}{2n}\right)^2,$$

and

$$AD = -n - \frac{1}{n}\sum_{j=1}^{n} \left[(2j-1)\log \hat{F}(X_{(j)}) + (2(n-j)+1)\log(1-\hat{F}(X_{(j)}))\right],$$

respectively, whereas the WA-statistic takes the form

$$WA = CM - n\left[\sum_{j=1}^{n} \frac{\hat{F}(X_{(j)})}{n} - \frac{1}{2}\right]^2.$$

Critical points for the *edf* tests were obtained by the same parametric bootstrap as in the case of the *ecf* based tests. All calculations were done using double precision arithmetic in FORTRAN and routines from the IMSL library, whenever available.

The following alternative distributions are considered:

- the two–component normal mixtures $\mathcal{NM}_\sigma$, with $\mu_1 = 0$ and $\sigma^2 = 1$ denoted by $HS(\lambda, \mu_2)$, and $\mathcal{NM}_\mu$, with $\mu = 0$ and $\sigma_1^2 = 1$, denoted by $HT(\lambda, \sigma_2^2)$,
- the gamma distribution $G(\theta)$, with density $\Gamma(\theta)^{-1}x^{\theta-1}\exp(-x)$,
- the Weibull distribution with density $\theta x^{\theta-1}\exp(-x^\theta)$, denoted by $W(\theta)$,
- the lognormal law $LN(\theta)$, with density $(\theta x)^{-1}(2\pi)^{-1/2}\exp[-(\log x)^2/(2\theta^2)]$,

- the normal–lognormal mixture $NLN(\theta)$, where $X = N(0, \sigma^2)$, with $\sigma^2 \sim LN(\theta)$,
- the Tukey g–h distribution, where $X = e^{hZ^2/2}((e^{gZ} - 1)/g)$, with $Z \sim N(0, 1)$, denoted by $GH(g, h)$,
- Student's t–distribution with ν degrees of freedom, denoted by t_ν,
- the Laplace distribution with density $(1/2)e^{-|x|}$,
- The logistic distribution with density $\exp(-x)[1 + \exp(-x)]^{-2}$,
- the uniform distribution in (-1,1),
- the triangular distribution in (-1,1),

Apart from the uniform and the triangular, these distributions comprise models widely used as alternatives to normal mixtures, in order to accommodate data that are skewed and/or heavy tailed. In Tables 1–2 (level results) and in Tables 3–4 (power results), the percentage of rejection is shown (rounded to the nearest integer) for all tests considered, with sample size n. Results for $a = \infty$ correspond to the limit statistics $\Sigma_{n,\infty}$ and $M_{n,\infty}$ defined in (2.4) and (2.5), respectively. An asterisk denotes power 100%. In the case of testing for $H_{0\sigma}$, we employ sample sizes $n = 100$ and $n = 200$, and $B = 500$ bootstrap samples are drawn. In the case of testing for $H_{0\mu}$, calculating the moment estimators involves the sixth sample moment, which renders the results for all tests considered highly variable. For example, when sampling from $\mathcal{NM}_\mu$, and for many samples with small sample size n, the resulting moment estimator did not belong to the parameter space ($\hat{\sigma}_1^2$ or $\hat{\sigma}_2^2 < 0$ was calculated). Therefore we employ sample size $n = 1000$. Even with a sample size as large as this, the problem persisted in 4.76% of the samples. This was the average percentage in five out of the 15 cases of distributions reported in Table 2. Each of the samples with $\hat{\sigma}_1^2$ or $\hat{\sigma}_2^2 < 0$ was discarded and replaced by a new sample. Consequently, and in order to save CPU time, only $B = 100$ bootstrap samples were generated. The main conclusions that can be drawn from the simulation results are the following:

(λ, μ_2)	CM	AD	KS	WA	0.5	1.0	2.0	3.0	5.0	10.0	∞
(0.1,2.0)	10	10	12	10	12	12	12	12	11	11	10
(0.1,4.0)	11	11	11	11	12	11	12	11	11	10	10
(0.1,6.0)	10	10	10	11	11	11	11	11	11	10	10
(0.2,2.0)	11	10	12	11	12	11	10	10	9	9	8
(0.2,4.0)	9	9	10	9	10	10	10	10	10	10	9
(0.2,6.0)	9	9	10	9	12	11	10	10	10	10	10
(0.3,2.0)	9	8	11	9	10	10	9	8	8	8	7
(0.3,4.0)	8	8	9	9	10	9	10	10	11	10	10
(0.3,6.0)	9	8	9	9	11	11	10	10	10	10	11
(0.4,2.0)	8	9	10	9	10	10	9	8	8	8	7
(0.4,4.0)	8	8	9	8	10	10	10	10	10	10	10
(0.4,6.0)	8	9	9	8	9	9	10	9	10	9	9
(0.5,2.0)	9	10	10	10	10	10	10	9	9	9	8
(0.5,4.0)	9	9	9	9	10	9	9	9	9	9	8
(0.5,6.0)	9	9	9	9	10	10	10	10	9	9	8
(0.1,2.0)	10	10	11	11	11	11	9	9	8	8	7
(0.1,4.0)	11	10	10	11	12	12	10	10	10	10	9
(0.1,6.0)	10	10	10	10	10	11	11	11	10	10	10
(0.2,2.0)	9	9	10	9	11	10	10	9	8	8	7
(0.2,4.0)	10	11	11	10	11	11	10	10	10	9	9
(0.2,6.0)	10	11	10	10	11	11	11	11	10	11	10
(0.3,2.0)	9	9	9	9	9	10	8	8	7	7	6
(0.3,4.0)	10	10	10	10	11	10	9	9	9	9	9
(0.3,6.0)	9	10	10	10	10	9	9	9	9	9	9
(0.4,2.0)	9	9	9	9	9	9	9	9	8	8	8
(0.4,4.0)	9	9	9	9	9	10	10	9	9	9	9
(0.4,6.0)	8	8	9	8	9	9	9	9	9	9	9
(0.5,2.0)	9	9	11	9	10	10	9	9	9	8	8
(0.5,4.0)	10	11	10	10	12	11	11	11	10	9	9
(0.5,6.0)	10	10	10	9	10	10	10	10	10	10	10

Table 1: Percentage of 1000 Monte Carlo samples from homoscedastic normal mixtures declared significant with sample size $n = 100$ (upper part) and $n = 200$ (lower part) by various tests for $H_{0\sigma}$ (level of significance $\alpha = 0.10$)

(λ, σ_2)	CM	AD	KS	WA	0.5	1.0	2.0	3.0	5.0	10.0	∞
(0.1,2.0)	6	6	7	6	9	9	10	11	12	13	12
(0.1,3.0)	1	1	1	1	1	4	9	11	12	13	13
(0.1,4.0)	1	1	1	1	4	7	10	11	12	14	13
(0.2,2.0)	3	3	4	3	5	9	11	10	11	11	10
(0.2,3.0)	2	2	2	2	6	9	9	9	10	10	10
(0.2,4.0)	4	8	4	5	11	13	13	10	9	10	8
(0.3,2.0)	3	2	3	2	5	7	10	10	10	10	9
(0.3,3.0)	5	6	4	4	10	11	10	7	8	9	8
(0.3,4.0)	3	7	4	3	4	3	4	4	6	7	7
(0.4,2.0)	3	3	3	3	6	7	9	9	9	9	9
(0.4,3.0)	3	5	4	3	5	6	9	9	9	9	9
(0.4,4.0)	4	6	5	4	4	4	4	5	7	7	7
(0.5,2.0)	3	4	4	3	6	8	9	10	10	9	9
(0.5,3.0)	4	4	5	4	4	4	6	7	8	8	8
(0.5,4.0)	6	6	6	6	6	6	6	6	7	8	7

Table 2: Percentage of 1000 Monte Carlo samples from homothetic normal mixtures declared significant with sample size $n = 1000$ by various tests for $H_{0\mu}$ (level of significance $\alpha = 0.10$)

1. Testing for $H_{0\sigma}$: The bootstrap empirical level satisfactorily recovers the nominal level of significance, for all tests. In terms of power the ecf test outperforms all *edf* tests uniformly over all values of the weight parameter a, with the exception of testing against a uniform or a triangular distribution. However such short–tailed symmetric distributions are rarely considered as alternatives to homoscedastic normal mixtures.

2. Testing for $H_{0\mu}$: All *edf* tests do not accurately recover the nominal level of significance, and there does not seem to exist a good choice among them, in this respect. The same inaccuracy is observed in the case of the ecf test. However the compromise choice $a = 2.0$ or $a = 3.0$, results in a test that in most cases of homothetic mixtures holds the empirical level close to its nominal value. With regards to power, and in absolute terms, all tests seem to lack power against several alternatives, mostly symmetric (or nearly symmetric) one's. All *edf* tests exhibit a similar performance, with the CM and the KS tests having a slight edge. In the case of the ecf test, there does

alternative	CM	AD	KS	WA	0.5	1.0	2.0	3.0	5.0	10.0	∞
$HT(0.2,4)$	29	32	26	29	35	40	41	41	41	40	39
$HT(0.2,9)$	73	74	66	73	77	78	78	77	77	76	74
$\Gamma(2)$	88	94	76	88	97	98	98	97	96	95	94
$\Gamma(3)$	64	72	51	63	81	84	87	86	84	81	79
$LN(0.25)$	21	25	17	22	28	27	27	27	27	26	27
$LN(0.50)$	78	83	64	75	93	91	90	89	89	89	89
$W(0.50)$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
$W(2.0)$	36	42	30	37	47	46	44	44	44	43	43
$Uniform$	79	87	58	81	2	2	1	1	1	1	1
$Triangul.$	23	23	21	23	1	1	1	1	1	1	1
$HT(0.2,4)$	48	51	40	49	56	61	62	62	60	59	56
$HT(0.2,9)$	97	97	94	97	98	97	97	97	97	96	95
$\Gamma(2)$	$\star$	$\star$	98	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
$\Gamma(3)$	94	98	85	94	99	99	99	99	99	99	98
$LN(0.25)$	40	44	31	40	53	52	51	51	51	49	50
$LN(0.50)$	98	99	94	98	$\star$	$\star$	99	99	99	99	99
$W(0.50)$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
$W(2.0)$	69	76	53	69	81	78	77	76	75	75	75
$Uniform$	99	$\star$	92	99	14	5	4	2	2	2	2
$Triangul.$	39	40	31	40	4	4	2	2	2	2	2

Table 3: Percentage of 1000 Monte Carlo samples declared significant with sample size $n = 100$ (upper part) and $n = 200$ (lower part) by various tests for $H_{0\sigma}$ (level of significance $\alpha = 0.10$)

not seem to exist a clear–cut choice for the value of a in terms of power. For some alternatives small values are better, while for other alternatives larger values result in a more powerful test. This fact combined with the level results discussed earlier indicates that the choice $a = 2.0$, renders the resulting ecf test the most accurate among all tests considered in terms of level and, although not very powerful in absolute terms, overall competitive to the best *edf* test.

3. The limit statistic $\Sigma_{n,\infty}$ is most of the times more conservative with regards to assessing the nominal level of significance. In terms of power it retains the general characteristics of $\Sigma_{n,a}$. Consequently, and despite the fact that $\Sigma_{n,\infty}$ is somewhat less

alternative	CM	AD	KS	WA	0.5	1.0	2.0	3.0	5.0	10.0	∞
$HS(0.1, 2)$	87	92	76	74	98	99	99	99	$\star$	99	99
$HS(0.1, 4)$	99	99	96	87	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
Laplace	37	34	38	38	26	18	12	10	9	9	10
Logistic	6	6	7	6	6	6	10	13	10	10	10
t_3	17	14	18	19	17	11	7	5	6	8	10
$NLN(0.5)$	6	6	6	6	10	10	11	11	11	11	10
$NLN(1.0)$	36	33	35	38	28	19	10	8	8	8	8
$GH(0.001, 0.2)$	28	24	28	30	24	14	7	6	7	8	9
$GH(0.5, 0.1)$	92	87	92	81	92	93	95	96	98	99	94
$GH(0.5, 0.2)$	60	42	62	37	54	64	75	81	87	95	87

Table 4: Percentage of 1000 Monte Carlo samples declared significant with sample size $n = 1000$ by various tests for $H_{0\mu}$ (level of significance $\alpha = 0.10$)

powerful than $\Sigma_{n,a}$ (for most values of a), it may be employed as a moment–based test, which is easy to implement. For the null hypothesis $H_{0\mu}$, the limit statistic $M_{n,\infty}$ shares the same power problems with the other tests considered here. It is however more accurate than $M_{n,a}$ (for several values of a) in recovering the nominal level of significance.

Since the results of the test for $H_{0\mu}$ using moment estimates are not very satisfactory, we tried other estimation procedures. First, we used a simple Newton algorithm yielding a local maximum and several methods of choosing starting values. However, the tests do not maintain the level in an acceptable way: the empirical level was around 3 or 5% (nominal level 10%) for sample sizes between 20 and 200. As a further drawback, nothing is known about the theoretical properties of these estimates.

Then, we tried global maximization of the likelihood function. To this end, we used a Simulated Annealing (SA) algorithm available in the optimization routine `optim` in the R-language. Since this SA algorithm allows no constraints, we transformed the parameters to unconstrained parameter space before maximization. SA is very slow. We only used sample size $n = 50$, and $B = 50$ bootstrap samples were drawn. The average CPU time for parameter estimation in a single sample was over 1 second. Therefore despite the rather small sample size, the computing time for Table 5 was about 10 days, and a larger sample size was not feasible. The choice $B = \max(n, [1/a])$

(λ, σ_2)	CM	AD	KS	WA	0.01	0.1	0.5	1.0	5.0	10.0
(0.5,2)	9	9	9	10	8	7	7	7	6	6
(0.5,3)	9	9	9	9	9	7	6	5	6	6
(0.5,4)	9	9	10	12	10	10	6	6	6	6
(0.2,2)	10	10	8	10	9	9	7	8	8	8
(0.2,3)	7	7	8	11	9	6	5	5	4	5
(0.2,4)	7	7	8	8	8	5	4	2	3	4
$HS(0.5, 2)$	10	11	12	20	13	14	12	12	13	12
$HS(0.5, 4)$	77	79	72	96	56	62	23	20	22	23
$HS(0.5, 8)$	100	100	100	100	70	74	74	23	23	20
$HS(0.5, 16)$	100	100	100	100	71	73	74	23	21	19
$HS(0.2, 8)$	97	99	90	49	38	66	94	92	71	57
$HS(0.2, 16)$	98	99	91	60	50	74	94	92	60	51

Table 5: Percentage of 1000 Monte Carlo samples declared significant with sample size $n = 50$ by various tests for $H_{0\mu}$ using maximum likelihood estimation (level of significance $\alpha = 0.10$)

has been suggested by Baringhaus and Henze (1992).

The results given in the first part of Table 5 show that the bootstrap empirical level is sufficiently close to the nominal level for all *edf* statistics; for the *ecf* test it is still acceptable for small values of a. For larger values of a, and particularly for $\lambda = 0.2$, the test is conservative. Looking at the power results in Table 4, it has to be expected that all tests will have low power for $n = 50$ against most alternatives, which is indeed the case. Only clearly separated bimodal densities can be detected, and only these cases are shown in the second part of Table 5. For these particular alternatives the power of the *edf* tests is higher than power of the *ecf* tests in most cases. As main conclusion of these limited simulation with a global optimization procedure one can state that the reason for the problems with moment estimation in case of $H_{0\mu}$ are due to the estimation procedure; the *edf* tests as well as the *ecf* tests also work for smaller sample sizes.

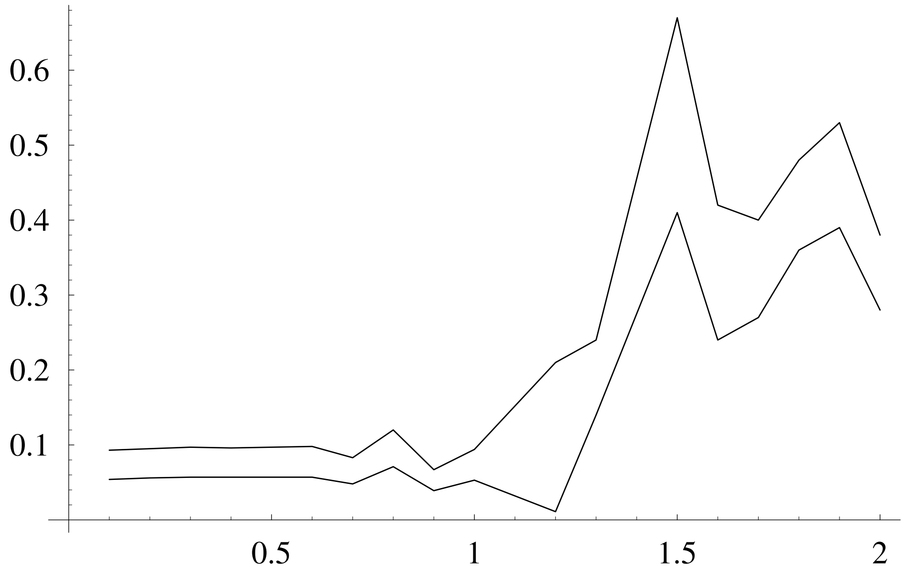

Figure 1: (Test statistic/critical point) ratio corresponding to the 5% (lower curve) and 10% (upper curve) $H_{0\sigma}$ test for the SLC data set

4 Real data examples

In order to investigate the performance of the proposed tests to practical situations, we have applied the *ecf* tests to real data sets. Our first illustration concerns the behaviour of the test based on $\Sigma_{n,a}$ when applied to the data set of Table 1 in Roeder (1994). This data set contains red blood sodium–lithium countertransport (SLC) activity for 190 individuals from six large English kindreds. The author claims that the square root of this data set supports a two–component normal mixture. In Figure 1, the values of the $\Sigma_{n,a}/\tilde{p}_n$ ratio is reported for $0 < a \leq 2.0$. Since this ratio is always less than one, we can not reject the hypothesis that the $\sqrt{\mathrm{SLC}}$ data set may well have arisen from a two–component homoscedastic normal mixture parent population.

The second example consists of data sets on logarithmic stock returns. Mixtures of normal distributions have been associated with such financial data both in the past, but also quite recently (see for example Kon (1984), Hall *et al*. (1989), Tucker (1992) and Tran (1998)). Financial researchers also use the stable (Paretian) distribution when modeling stock returns. The stable model can accommodate leptokurtic as well as skewed data, and enjoys the so–called ‘closure under summation’ property. These features are highly relevant in financial markets. However, certain drawbacks in employing the stable model, such as the lack of closed form expression for the general stable density, and the non–existence of higher moments, render the issue of the modeling potential of stable distributions controversial. For example, Tucker (1992) and McDonald (1996) point out that empirical findings on financial data are often not consistent with the stable model, and that alternative distributions should be employed. In fact it is believed that normal mixtures are most appropriate in order to accommodate, apart from the observed leptokurtosis, certain discontinuities in stock returns such as the 'weekend effect', the (turn–of–the) 'month effect' and the 'January effect'.

The first data set on returns corresponds to the company Bethlehem Steel over the period Oct. 3, 1984–Oct. 22, 1990. Logarithmic return values registered during the so–called October 'crash' period (Sept. 23–Nov. 3, 1987), correspond to a particularly extreme situation, and therefore were removed, thereby reducing the sample size to 1500 stock return values. In Figure 2, the values of the $M_{n,a}/\tilde{p}_n$ ratio is reported for $0.5 < a \leq 3.0$, and are in support of the hypothesis of a scale mixture of two normal distributions.

The second set of returns corresponds to ten companies traded in the Athens Stock Exchange over the period is Jan. 1, 1999–Dec. 31, 2002. The number of trading days in this period yields a sample size $n = 998$ observations for each company. We applied the method of moments on these returns. In seven out of the total ten data sets, the estimates lie outside the parameter space, and consequently for these seven companies the assumption of a homothetic normal mixture is rejected. The remaining data sets correspond to 'The National Bank of Greece', 'Alpha Bank' (a private bank), and to the chain of fast–food restaurants 'Goodys'. Returns for these three firms yield moment estimates within the parameter space, and the value of the resulting test statistic is in each case (and regardless of the value of the weight parameter a) highly insignificant, indicating that the hypothesis of a two–component homothetic normal mixture for these data can not be rejected.

In view of the difficulties with the moment estimates for these ten data sets, we also applied the tests using maximum likelihood estimation. This method yields reasonable estimates for each data set. The hypothesis of a homothetic normal mixture

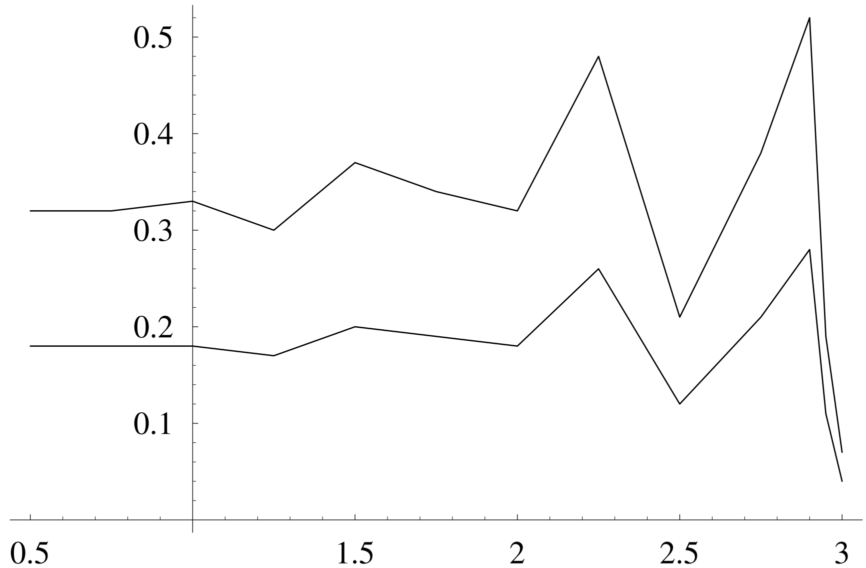

Figure 2: (Test statistic/critical point) ratio corresponding to the 5% (lower curve) and 10% (upper curve) $H_{0\mu}$ test for the Bethlehem Steel data

was clearly rejected for the seven data sets for which the method of moments failed. For the remaining three data sets, the p-values for the different *edf* and ecf tests are shown in Table 6. Hence, the hypothesis is rejected for the third data set. For the first two data sets, the p-value depends heavily on the *edf* statistic or on the value of a when applying the ecf based tests; however, one can conclude that the normal mixture model is acceptable for these two data sets.

Acknowledgements: The authors wish to thank the associate editor and a referee for their constructive critisism. Research on this topic started while the second

	CM	AD	KS	WA	0.01	0.1	0.5	1.0	5.0	10.0
Bank of Greece	0.23	0.11	0.47	0.26	0.08	0.17	0.44	0.34	0.11	0.08
Alpha Bank	0.16	0.02	0.43	0.60	0.46	0.67	0.42	0.40	0.07	0.02
Goodys	0.04	0.06	0.00	0.00	0.00	0.00	0.02	0.05	0.09	0.11

Table 6: p-values for the different *edf* and ecf tests

author was visiting the University of Karlsruhe. Simos Meintanis would like to thank the staff of the Institute of Stochastics for their hospitality and strong support. Both authors are grateful to Norbert Henze. Without his help this work would not have been completed. Some of the data sets where kindly made available with the help of K. Fragiadakis, a graduate student in the University of Athens.

5 Appendix

We consider the asymptotic behavior of T_n defined by (1.2), in a nonparametric setting. The following result implies the consistency of a goodness–of–fit test that rejects H_0 for large values of T_n against general alternatives.

Theorem 5.1 *Let X be a random variable with characteristic function $\phi(t)$ and assume that, at least for large n, the estimator $\hat{\vartheta}$ exists and converges in probability to (say) ϑ. Then*

$$n^{-1}T_n \xrightarrow{P} \Delta := \int_{-\infty}^{\infty} |\phi(t) - \varphi(t;\vartheta)|^2 \, w(t) \, dt.$$

PROOF. Recall that

$$\frac{T_n}{n} = \int_{-\infty}^{\infty} |\phi_n(t) - \varphi(t;\hat{\vartheta})|^2 \, w(t) \, dt,$$

and note that $|\phi_n(t) - \varphi(t;\hat{\vartheta})|^2 \leq 4$. Then due to the consistency of $\phi_n(t)$ and the continuity of $\varphi(t;\cdot)$, the proof of the theorem follows by an application of Lebesgue's theorem of dominated convergence. ∎

In view of (1.1), Δ is equal to zero if $\phi(t) \equiv \varphi(t;\vartheta)$, for some ϑ, but positive otherwise. Consequently, a level α–test that rejects H_0 for large values of T_n is consistent against each fixed alternative distribution which satisfies the conditions of Theorem 5.1.

Remark 5.2 There exist cases of distributions with finite moments of all orders, where even the moment equations do not admit solutions, at least asymptotically. For example, the solution of the moment equations assuming a homoscedastic two–component normal mixture, when a sample of size n from a (standard) Laplace distribution is at hand, does not exist as $n \to \infty$. The same problem appears when attempting to solve the moment equations assuming a homothetic two–component normal mixture, with a sample from the uniform or the triangular distribution in (-1,1) at hand.

References

Baringhaus, L., Gürtler, N., & Henze, N. 2000. Weighted integral test statistics and components of smooth tests of fit. *Austral. & New Zeal. J. Statist.*, *42*, 179–192.

Baringhaus, L., & Henze, N. 1992. A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statist. & Probab. Lett.*, *13*, 269–274.

Bryant, J.L., & Paulson, A.S. 1983. Estimation of mixing proportions via distance between characteristic functions. *Commun. Statist.–Theor. Meth.*, *12*, 1009–1029.

Cohen, A.C. 1967. Estimation in mixtures of two normal distributions. *Technometrics*, *9*, 15–28.

Epps, T.W. 1999. Limiting behavior of the ICF test for normality under Gram-Charlier alternatives. *Statist. & Probab. Lett.*, *42*, 175–184

Gürtler, N., & Henze, N. 2000a. Goodness–of–fit tests for the Cauchy distribution based on the empirical characteristic function. *Ann. Inst. Statist. Math.*, *52*, 267–286.

Gürtler, N., & Henze, N. 2000b. Recent and classical goodness–of–fit tests for the Poisson distribution. *J. Statist. Plann. & Infer.*, *90*, 207–225.

Hall, J.A., Brorsen, B.W. , & Irwin, S.H. 1989. The distribution of future prices: A test of the stable Paretian and the mixture of normals hypotheses. *J. Financ. Quantit. Anal.*, *24*, 105–116.

Heathcote, C.R. 1972. A test of goodness of fit for symmetric random variables. *Austral. J. Statist.*, *14*, 172-181.

Kankainen, A. , & Ushakov, N.G. 1998. A consistent modification of a test for independence based on the empirical characteristic function. *J. Math. Sciences*, *89*, 1486–1494.

Kon, S.J. 1984. Models of stock returns–A comparison. *The Journal of Finance*, *39*, 147–165.

Koutrouvelis, I. A., & Canavos, G.C. 1997. Estimation in the three–parameter gamma distribution based on the empirical moment generating function. *J. Statist. Comput. & Simulat.*, *59*, 47–62.

Koutrouvelis, I.A., & Meintanis, S.G. 2002. Estimating the parameters of Poisson–exponential models. *Austral. & New Zeal. J. Statist.*, *44*, 233–245.

McDonald, J.B. 1996. Probability distributions for financial models. In: G.S. Maddala and C.R. Rao (eds.), Statistical Methods in Finance, *Handbook of Statistics, vol. 14*, 427–461.

Neus, J.L., & Mendell, N.R. 2001. A Lilliefors test of fit to the two–component homoscedastic normal mixture. *Commun. Statist.–Simul. Comput.*, *30*, 277–289.

Press, S.J. 1972. Estimation in univariate and multivariate stable distributions. *J. Amer. Statist. Assoc.*, *67*, 842–846.

Quandt, R.E. , & Ramsey, J.B. 1978. Estimating mixtures of normals and switching regressions. *J. Amer. Statist. Assoc.*, *73*, 730–738.

Roeder, K. 1994. A graphical technique for determining the number of components in a mixture of normals. *J. Amer. Statist. Assoc.*, *89*, 487–495.

Schmidt, P. 1982. An improved version of the Quandt–Ramsey MGF estimator for mixtures of normals and switching regressions *Econometrica*, *50*, 501–516.

Thode, H.C. 2002. *Testing for Normality.* Marcel Dekker Inc., New York.

Titterington, D.M., Smith, A.F.M. , & Makov, U.E. 1985. *Statistical Analysis of Finite Mixture Distributions.* J. Wiley & Sons, New York.

Tucker, A.L. 1992. A reexamination of finite– and infinite–variance distributions as models of daily stock returns. *J. Busin. & Econom. Statist.* , *10*, 73–81.

Tran, K.C. 1998. Estimating mixtures of normal distributions via empirical characteristic function. *Economtr. Rev.* , *17*, 167–183.

Ushakov, N.G. 1999. *Selectic Topics in Characteristic Functions.* VSP, Utrecht.