Scientific
Research
Publishing

Simulated Minimum Hellinger Distance Estimation for Some Continuous Financial and Actuarial Models

Andrew Luong*, Claire Bilodeau

École d'actuariat, Université Laval, Québec, Canada
Email: *Andrew.Luong@act.ulaval.ca, Claire.Bilodeau@act.ulaval.ca

Abstract

Minimum Hellinger distance (MHD) estimation is extended to a simulated version with the model density function replaced by a density estimate based on a random sample drawn from the model distribution. The method does not require a closed-form expression for the density function and appears to be suitable for models lacking a closed-form expression for the density, models for which likelihood methods might be difficult to implement. Even though only consistency is shown in this paper and the asymptotic distribution remains an open question, our simulation study suggests that the methods have the potential to generate simulated minimum Hellinger distance (SMHD) estimators with high efficiencies. The method can be used as an alternative to methods based on moments, methods based on empirical characteristic functions, or the use of an expectation-maximization (EM) algorithm.

Keywords

Infinitely Divisible Distribution, Mixture Distribution, Hellinger Distance, Robustness

1. Introduction

In actuarial science or finance, we often encounter the problem of fitting distributions to data where the distributions have no closed-form expressions for their densities. These distributions are often infinitely divisible and they happen to be the distributions of the regularly spaced increments of Lévy processes. Beside infinitely divisible distributions, mixture distributions created using a mixing mechanism also provide examples of continuous densities without a closed-form expression. These types of distributions are often encountered in actuarial

science. A few examples will be provided as illustrations subsequently.

Likelihood methods might be difficult to implement in such cases, due to the lack of a closed-form expression for the density function. To handle such a situation, we can consider the following approaches:

1) Expectation-maximization (EM) algorithm. Only under special conditions can the EM algorithm be used as it requires some conditional distributions, and these conditional distributions might be difficult to obtain; see McNeil, Frey and Embrechts [1] (pages 81-85) or McLachlan and Krishnan [2].

2) Method of moments. Even though the model density has no closed form, if the model moments can be expressed in closed form, then the method of moments can be used. The main drawback of the method of moments is that estimators thus obtained might not be efficient nor robust for models with three or more parameters as the estimators will depend on a polynomial of degree three or higher, making the methods very sensitive to data which are contaminated; see Küchler and Tappe [3] [4] for method of moments estimation.

3) The k-L procedure. Even if the density has no closed form, if the model characteristic function has a closed-form expression, then we can select points from the real and imaginary parts of the empirical characteristic function and match them with their model counterparts at the chosen points. This is the k-L procedure as proposed by Feuerverger and McDunnough [5] (pages 22-24).

4) Indirect inference. These methods are based on simulations and they require two steps. First, we need to choose a proxy model to obtain the estimators which are biased. Second, we remove the bias using simulations. See Garcia, Renault and Veredas [6] for this method. The proxy models from which the estimators are obtained affect the efficiencies of the estimators. For some models, it is difficult to know which proxy model will generate estimators with high efficiencies.

When implementing these methods for distributions without closed-form densities, there are some drawbacks which motivate us in this paper to extend minimum Hellinger distance methods originally proposed by Beran [7] to a simulated version (version S) which consists in replacing the model density $f_{\boldsymbol{\theta}}(x)$ by a density estimate $f_{\boldsymbol{\theta}}^{S}(x)$ using a random sample drawn from $f_{\boldsymbol{\theta}}(x)$ and minimizing

$$Q_n(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \left(\left[f_n(x) \right]^{\frac{1}{2}} - \left[f_{\boldsymbol{\theta}}^{S}(x) \right]^{\frac{1}{2}} \right)^2 \mathrm{d}x \tag{1}$$

to obtain the simulated minimum Hellinger distance (SMHD) estimators, where $f_n(x)$ is an empirical density estimate based on the observed data with the property $f_n(x) \overset{p}{\longrightarrow} f_{\boldsymbol{\theta}_0}(x)$ where $\boldsymbol{\theta}_0$ is the true vector of parameters. This consistency property will imply $\int_{-\infty}^{\infty} \left(\left[f_n(x) \right]^{\frac{1}{2}} - \left[f_{\boldsymbol{\theta}_0}(x) \right]^{\frac{1}{2}} \right)^2 \mathrm{d}x \overset{p}{\longrightarrow} 0$ as $n \to \infty$; see section 3 (page 224) of Tamura and Boos [8].

Clearly, the new method proposed here will avoid the problem of arbitrariness

in the choice of points for the k-L procedure based on characteristic functions. Unlike indirect inference, the proposed method does not need a proxy model. Furthermore, the estimators obtained using the proposed method might be more robust and efficient than method of moments estimators. Besides, the proposed method does not require conditioning, which can be difficult, whereas the EM algorithm does.

It appears that the proposed method, which originally combines simulation with Hellinger distance, adds to the set of statistical techniques that can be useful for financial and actuarial data, yet many of which do not receive much attention in the actuarial literature. SMHD methods depend on being able to draw samples from the parametric family; in general, this is indeed possible. Consequently, SMHD methods also add to the existing literature on simulated inference which is relatively new; see comments by Davidson and MacKinnon [9] (page 393).

The new method is built on the classical version (version D) of Hellinger distance as proposed by Beran [7] which consists in minimizing

$$Q_n(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \left(\left[f_n(x) \right]^{\frac{1}{2}} - \left[f_{\boldsymbol{\theta}}(x) \right]^{\frac{1}{2}} \right)^2 \mathrm{d}x \qquad (2)$$

to obtain the minimum Hellinger distance (MHD) estimators. The MHD estimators have been known to have nice robustness properties with breakdown point greater than 0. Also, they are consistent with, in general, less stringent conditions for consistency than maximum likelihood (ML) estimators. However, more restrictions are placed upon the underlying parametric family for the MHD estimators to attain full efficiency, such assuming $f_{\boldsymbol{\theta}}(x)$ having a compact support for example. Despite this drawback, simulation studies often show that the methods perform well across many models. For a literature review of Hellinger distance (HD) methods, see chapters 3 and 10 of the book by Basu, Shioya and Park [10]. From the literature, it can be seen that HD methods still do not receive proper attention for their use in actuarial science and finance, especially in actuarial science.

In this paper, we introduce a simulated version of HD methods and show that the SMHD estimators are consistent. However, the question of asymptotic normality is still not resolved for the time being. Further work should generate results on asymptotic distributions for the SMHD estimators that shall then be presented in a subsequent paper. In this paper, the methods are presented with fewer technicalities and we relate them with the traditional likelihood methods. In doing so, we wish to encourage practitioners to use these methods for their applied works in their fields. In the next paragraphs, we will consider a few examples for illustrations of the types of distributions without closed-form expressions often encountered in finance and actuarial science where the new simulated method can be particularly useful.

Example 1

We present here the class of normal mean-variance mixture distributions

where the random variable X can be represented using equality in distribution as

$$X =^d \theta + \mu W + \sigma \sqrt{W} Z,\tag{3}$$

where

1) θ, μ and σ are parameters with $-\infty < \theta < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$;

2) W is a nonnegative random variable with an infinitely divisible (ID) distribution;

3) Z follows a standard normal distribution $N(0,1)$ and is independent of W.

The generalized hyperbolic, variance-gamma, and normal-inverse Gaussian distributions belong to this class; see McNeil, Frey and Embrechts [1] (pages 77-79). By conditioning on W first, the moment generating function (mgf) for X can be obtained and given by $M_X(s) = e^{\theta s} M_W\left(\mu s + \frac{1}{2}\sigma^2 s^2\right)$, where the moment generating functions of X and W are given respectively by $M_X(s)$ and $M_W(s)$. Distributions of the increments observed at regular intervals of a subordinated Brownian motion process belong to this class. It can easily be seen that the density function of X depends on the density function of W. Consequently, the density function of X might not have a closed-form expression in general. Closely related to the variance-gamma distribution is the generalized normal-Laplace (GNL) distribution which is introduced by Reed [11] and is given in the next example.

Example 2

A random variable X follows a GNL distribution if it can be represented as

$$X =^d \rho\mu + \sigma\sqrt{\rho}Z + \frac{1}{\alpha}G_1 - \frac{1}{\beta}G_2,\tag{4}$$

where

1) the parameters are μ, σ, ρ, α and β, with $-\infty < \mu < \infty$, $\sigma > 0$, $\rho > 0$, $\alpha > 0$, and $\beta > 0$;

2) the random variables G_1 and G_2 are independent and follow a common gamma distribution with density function $g(x;\rho) = \frac{1}{\Gamma(\rho)}x^{\rho-1}e^{-x}, x > 0, \rho > 0$;

3) Z follows a standard normal distribution, $N(0,1)$, with Z being independent of G_1 and G_2.

The distribution is infinitely divisible and can display asymmetry and fatter tail than the normal distribution. It will be symmetric if $\alpha = \beta$. The vector of parameters is $\boldsymbol{\theta} = (\mu, \sigma, \rho, \alpha, \beta)'$ and the mgf for X can be obtained using the representation given by Equation (4) and is given by

$$M_X(s) = e^{\rho\left(\mu s + \frac{1}{2}\sigma^2 s^2\right)}\left(\frac{\alpha}{\alpha - s}\right)^\rho \left(\frac{\beta}{\beta + s}\right)^\rho.\tag{5}$$

From the cumulant generating function, the mean and variance are given re-

spectively by

$$E(X) = \rho\left(\mu + \frac{1}{\alpha} - \frac{1}{\beta}\right) \tag{6}$$

and

$$V(X) = \rho\left(\sigma^2 + \frac{1}{\alpha^2} + \frac{1}{\beta^2}\right). \tag{7}$$

Higher cumulants are

$$\kappa_r = \rho(r-1)!\left(\frac{1}{\alpha^r} + (-1)^r\frac{1}{\beta^r}\right) \text{ for } r > 2. \tag{8}$$

Due to the lack of a closed-form expression for the density function, Reed [11] (page 477) has proposed using the method of moments and matching the empirical cumulants with the model cumulants to estimate the parameters. He applied the method to data collected on stocks. In the particular case with four parameters, where $\alpha = \beta$, moment estimators can be obtained explicitly. However, for the general case with five parameters, the moment equations must be solved numerically. The moment estimators will be discussed in more detail in section 3 and we shall compare their efficiencies with the efficiencies of the SMHD estimators based on simulated samples.

For more on Lévy processes and infinitely divisible distributions used in finance, see chapter 6 of the book by Schoutens [12] (pages 73-83). For nonnegative infinitely divisible distributions used in actuarial science, see Dufresne and Gerber [13], and Luong [14]. For mixtures of distributions without closed-form density functions, for which the proposed estimators can also be used, see Klugman, Panjer and Willmot [15] (pages 62-65). We shall consider HD estimation in all those cases.

Assume that we have a random sample of observations $X_1, \cdots, X_n$ and they are independent and identically distributed as the random variable X which is continuous with model density given by $f_{\boldsymbol{\theta}}(x)$. The vector of parameters is denoted by $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)'$. In his seminal paper, Beran [7] proposes to estimate $\boldsymbol{\theta}$ by the minimum Hellinger distance estimators denoted by $\hat{\boldsymbol{\theta}}$ which minimize, with respect to $\boldsymbol{\theta}$, the Hellinger distance between a consistent empirical density estimate f_n and the parametric family $f_{\boldsymbol{\theta}}$ with the property $f_n(x) \xrightarrow{p} f_{\boldsymbol{\theta}_0}(x)$ pointwise. It leads to minimize the objective function

$$Q_n(\boldsymbol{\theta}) = \int_{-\infty}^{\infty}\left(\left[f_n(x)\right]^{\frac{1}{2}} - \left[f_{\boldsymbol{\theta}}(x)\right]^{\frac{1}{2}}\right)^2 \mathrm{d}x. \tag{9}$$

Beran [7] also noted that, intuitively, the methods are robust as data are smoothed by a kernel density estimator f_n, and hence the effects of outliers are mitigated. It has been confirmed in various models that the asymptotic breakdown points of the estimators are around $\frac{1}{2}$ and it is well-known that the sample mean has a breakdown point of 0. See Hogg, McKean and Craig [16] (pages

594-595), and Maronna, Martin and Yohai [17] (page 58) for the notions of finite sample and asymptotic breakdown points as measures of robustness of estimators. See Lindsay [18] for the discussions on robustness and efficiencies of MHD estimators. We also note that, since

$$Q_n(\boldsymbol{\theta}) = 2 - 2\int_{-\infty}^{\infty} \left[f_n(x)\right]^{\frac{1}{2}} \left[f_{\boldsymbol{\theta}}(x)\right]^{\frac{1}{2}} dx \qquad (10)$$

and, using the Cauchy-Schwarz inequality, $\int_{-\infty}^{\infty} \left[f_n(x)\right]^{\frac{1}{2}} \left[f_{\boldsymbol{\theta}}(x)\right]^{\frac{1}{2}} dx \leq 1$, we find

$$0 \leq Q_n(\boldsymbol{\theta}) \leq 2. \qquad (11)$$

Moreover, since $\int_{-\infty}^{\infty} \left[f_n(x)\right]^{\frac{1}{2}} \left[f_{\boldsymbol{\theta}}(x)\right]^{\frac{1}{2}} dx = 1$ if and only if $f_n(x) = f_{\boldsymbol{\theta}}(x)$ almost everywhere, it implies $Q_n(\boldsymbol{\theta}) = 0$ if and only if $f_n(x) = f_{\boldsymbol{\theta}}(x)$ almost everywhere.

The objective function is stable and bounded. This might explain why, intuitively, minimizing such an objective function, we obtain estimators that are also stable and therefore robust in some sense.

Kernel density estimators are often used to define $f_n(x)$. One of the simplest kernel density estimators is the rectangular kernel density estimator which generalizes the usual histogram estimator. In general, kernel density estimators have the form

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \omega\left(\frac{x - x_i}{h_n}\right), \qquad (12)$$

where

a) h_n is the bandwidth with the property that $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$;

b) $\omega(x)$ is a density function.

The property specified by a) guarantees the consistency of $f_n(x)$; see Corollary 6.4.1 given by Lehmann [19] (pages 406-408). Subsequently, we implicitly assume that density estimates used with the SMHD method meet the requirements specified by a) and b).

For the rectangular kernel density, the following symmetric density around 0 is chosen with $\omega(x) = \frac{1}{2}$ for $-1 < x < 1$. The kernel $\omega(x)$ has a compact support. The density estimate at x is then the average of rectangles located within h_n units from x. For other kernels and their implementation using the package R, see chapter 10 of the book by Rizzo [20] (pages 281-318). For Hellinger distance estimation, it is preferable to use a symmetric kernel with a compact support and twice differentiable for meeting the regularity conditions of Theorem 4 as given by Beran [7] (pages 450-451); also see the discussions by Basu, Shioya and Park [10] (pages 78-83). In this paper, we only need univariate kernel density estimates but multivariate density estimates based on kernels can also be defined similarly; see Toma [21] and Scott [22].

If $f_{\boldsymbol{\theta}}(x)$ has no closed-form expression but random samples can be drawn from the distribution with density $f_{\boldsymbol{\theta}}(x)$, clearly we can use the same type of

kernel density estimator, used to define $f_n(x)$, to estimate $f_{\theta_0}(x)$. In other words, in order to estimate $f_\theta(x)$, we similarly define $f_\theta^S(x)$ as being the kernel density estimator based on a random sample of size $U = \tau n$. Note that $U \to \infty$ as $n \to \infty$ and τ needs to be reasonably large so that there is little loss of efficiencies due to simulations; we recommend $\tau \geq 10$.

Consequently, for the simulated version, we shall minimize the objective function given by

$$Q_n(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \left(\left[f_n(x) \right]^{\frac{1}{2}} - \left[f_\theta^S(x) \right]^{\frac{1}{2}} \right)^2 \mathrm{d}x \qquad (13)$$

to obtain the SMHD estimators.

For terminology, we shall call the classical version, which is deterministic in terms of $f_\theta(x)$, version D, and the simulated version, version S. Since $Q_n(\boldsymbol{\theta})$, as given by Equation (13), is not differentiable, a direct simplex search method which is derivative-free is recommended. The R package already has a built-in function for performing the Nelder-Mead simplex method which is a derivative-free method to minimize a function. Also, there is a built-in function to handle density estimates using various kernels. These features will facilitate the implementation of SMHD methods for applied works by practitioners. Furthermore, because the densities $f_n(x)$ and $f_\theta^S(x)$ based on a rectangular or triangular kernel are positive only in some finite interval and zero elsewhere, this makes the integration for evaluating Equation (13) easy to handle. A trapezoid quadrature method will suffice to find the SMHD estimators. Note that for the simulated version, we still have

$$0 \leq Q_n(\boldsymbol{\theta}) \leq 2. \qquad (14)$$

As data are also smoothed, intuitively, these features will again make the simulated version robust.

The paper is organized as follows. In Section 2, we will look into the asymptotic properties of MHD estimators. More precisely, we shall briefly review the asymptotic properties of the classical MHD estimators in Section 2.1 and establish the consistency of SMHD estimators in Section 2.2. Also in Section 2.2, an estimator for the Fisher information matrix is proposed with the use of SMHD estimators. In Section 3, we use a limited simulation study to compare the efficiencies of the SMHD estimators with those of method of moments estimators, using the GNL distribution. Despite being limited, the study seems to show that the SMHD estimators are more efficient than the method of moments estimators. This seems to point to the potential of SMHD methods to generate estimators with good efficiency and further justify their use in actuarial science and finance.

2. Asymptotic Properties

2.1. Asymptotic Properties of the Classical MHD Estimators

MHD estimators can be seen to be consistent in general for version D and version S. In fact, the conditions are even less restrictive than the conditions for

maximum likelihood estimators to be consistent. Since we aim for applications, we only consider asymptotic properties under the strict parametric model, *i.e.*, assuming the observations come from the parametric density family $f_{\boldsymbol{\theta}}(x)$, where $\boldsymbol{\theta} \in \Omega$, and the parameter space Ω is assumed to be compact.

Let

$$\left\| (f_1)^{\frac{1}{2}} - (f_2)^{\frac{1}{2}} \right\| = \left[\int_{-\infty}^{\infty} \left(\left[f_1(x) \right]^{\frac{1}{2}} - \left[f_2(x) \right]^{\frac{1}{2}} \right)^2 \mathrm{d}x \right]^{\frac{1}{2}}, \tag{15}$$

where $f_1(x)$ and $f_2(x)$ are density functions. Note that $\|\cdot\|$ is a norm in the density functional space and it will respect the triangular inequality.

Tamura and Boos [8] (page 224) have noted that, if $f_n(x) \overset{p}{\longrightarrow} f_{\boldsymbol{\theta}_0}(x)$, then $\left\| (f_n)^{\frac{1}{2}} - (f_{\boldsymbol{\theta}_0})^{\frac{1}{2}} \right\| \overset{p}{\longrightarrow} 0$, and if $\left\| (f_n)^{\frac{1}{2}} - (f_{\boldsymbol{\theta}})^{\frac{1}{2}} \right\| > 0$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ in probability, it is sufficient for the MHD estimators given by the vector $\hat{\boldsymbol{\theta}}$ obtained by minimizing Equation (10) to be consistent, *i.e.*, $\hat{\boldsymbol{\theta}} \overset{p}{\longrightarrow} \boldsymbol{\theta}_0$, assuming the parameter space Ω is compact. See Theorem 3.1 by Tamura and Boos [8] (page 224). Comparing with the regularity conditions for ML estimators as given by Theorem 2.5 of Newey and McFadden [23] (page 2131), the regularity conditions for MHD estimation do not require that $E\left(\sup_{\boldsymbol{\theta}} \left| \log f_{\boldsymbol{\theta}}(x) \right| \right) < \infty$ as in likelihood estimation. This makes the MHD estimators consistent in general even with fewer restrictions than ML estimators.

However, for asymptotic normality, they require more stringent conditions to be as efficient as ML estimators. They are found in Theorem 4 given by Beran [7] (pages 450-451), which is summarized in Theorem 1 below, focusing on the strict parametric model. Beran [7] (pages 450-451) allows the bandwidth of the kernel to be randomly chosen with $h_n = c_n s_n$, where c_n is a sequence of constants but s_n is a sequence of random variable with $s_n \overset{p}{\longrightarrow} s$. It also requires a compact support K for both $\dfrac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}}(x)$. Despite these restrictions, empirical studies often show that the estimators have high efficiencies in many models without the condition of compact support for the parametric family met. The regularity conditions of Beran's Theorem 4 when restricted to the strict parametric model are stated using Theorem 1 below. We also require the vector of true parameters $\boldsymbol{\theta}_0$ to be in Ω, where Ω is compact. Theorem 1 can be viewed as a corollary of Theorem 4 as given by Beran [7] and the proofs have been given there.

Theorem 1

Suppose

1) The kernel density $\omega(x)$ is symmetric about 0 and has a compact support.

2) The function $\omega(x)$ is twice differentiable and its second derivative is bounded on the compact support.

3) $\dfrac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}}(x)$ have a compact support K and $f_{\boldsymbol{\theta}}(x) > 0$ on K.

4) $f_{\boldsymbol{\theta}}(x)$ is twice absolutely continuous with its second derivative with respect to x being bounded.

5) $\lim_{n\to\infty} n^{\frac{1}{2}} c_n = \infty$, $\lim_{n\to\infty} n^{\frac{1}{2}} c_n^2 = 0$, and $\lim_{n\to\infty} c_n = 0$.

6) There exists a positive constant s which might depend on $f_{\boldsymbol{\theta}_0}(x)$ such that $\sqrt{n}(s_n - s)$ is bounded in probability.

Then $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} N(0, I(\boldsymbol{\theta}_0)^{-1})$ where $I(\boldsymbol{\theta}_0)$ is the Fisher information matrix with elements given by

$$E\left(\frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \theta_j} \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \theta_i}\right) = -E\left(\frac{\partial^2 \log f_{\boldsymbol{\theta}}(x)}{\partial \theta_j \partial \theta_i}\right), i = 1, \cdots, m, j = 1, \cdots, m \quad (16)$$

and assumed to exist.

We just give an outline establishing the results of Theorem 1 and focus only on the strict parametric model for applications with the aim that it might help practitioners in the applied fields to follow more easily the arguments needed to develop the new method subsequently.

Note that, beside the rectangular kernel, the triangular kernel with $\omega(x) = 1 - |x|$ for $-1 \le x \le 1$ and the Epanechnikov kernel with $\omega(x) = \dfrac{3}{4}(1 - x^2)$ for $-1 \le x \le 1$ meet conditions 1 and 2 as required by Theorem 1 and are available in the package R.

For establishing asymptotic normality results for the estimators as indicated by Theorem 1, we can consider a Taylor expansion of the system of equations $D(\hat{\boldsymbol{\theta}}) = \dfrac{\partial Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = 0$ around the true vector of parameters $\boldsymbol{\theta}_0$. The system of equations implies

$$D(\hat{\boldsymbol{\theta}}) = \int_{-\infty}^{\infty} \left(\left[f_n(x)\right]^{\frac{1}{2}} - \left[f_{\hat{\boldsymbol{\theta}}}(x)\right]^{\frac{1}{2}}\right) \frac{\dfrac{\partial f_{\hat{\boldsymbol{\theta}}}(x)}{\partial \boldsymbol{\theta}}}{\sqrt{f_{\hat{\boldsymbol{\theta}}}(x)}} \,\mathrm{d}x = 0 \quad (17)$$

with $\dfrac{\partial f_{\hat{\boldsymbol{\theta}}}(x)}{\partial \boldsymbol{\theta}} = \dfrac{\partial f_{\boldsymbol{\theta}}(x)}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$ and $f_{\hat{\boldsymbol{\theta}}}(x) = f_{\boldsymbol{\theta}}(x)\big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$.

We proceed to perform a Taylor expansion by noting

$$D(\boldsymbol{\theta}_0) = \int_{-\infty}^{\infty} \left(\left[f_n(x)\right]^{\frac{1}{2}} - \left[f_{\boldsymbol{\theta}_0}(x)\right]^{\frac{1}{2}}\right) \frac{\dfrac{\partial f_{\boldsymbol{\theta}_0}(x)}{\partial \boldsymbol{\theta}}}{\sqrt{f_{\boldsymbol{\theta}_0}(x)}} \,\mathrm{d}x, \quad (18)$$

$$\dot{D}(\boldsymbol{\theta}_0) = \frac{\partial D(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} = -\frac{1}{2}\int_{-\infty}^{\infty} \left(\frac{\partial \log f_{\boldsymbol{\theta}_0}(x)}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log f_{\boldsymbol{\theta}_0}(x)}{\partial \boldsymbol{\theta}}\right)' f_{\boldsymbol{\theta}_0}(x)\,\mathrm{d}x + o_p(1), (19)$$

assuming $D(\boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$ and

$$\int_{-\infty}^{\infty} \left(\left[f_n(x) \right]^{\frac{1}{2}} - \left[f_{\theta_0}(x) \right]^{\frac{1}{2}} \right) \frac{\partial s_{\theta_0}(x)}{\partial \theta} \mathrm{d}x \xrightarrow{p} 0 , \quad \text{with} \quad s_{\theta_0}(x) = \frac{\dfrac{\partial f_{\theta_0}(x)}{\partial \theta}}{\sqrt{f_{\theta_0}(x)}} , \quad \text{using}$$

the compact support assumption for $\{ f_\theta \}$. As a result, we can write that

$$\dot{D}(\theta_0) = -\frac{1}{2} I(\theta_0) + o_p(1) . \tag{20}$$

Therefore, with the regularity conditions met, we will have the representation

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left[\dot{D}(\theta_0) \right]^{-1} \sqrt{n} D(\theta_0) + o_p(1) , \tag{21}$$

where $o_p(1)$ is the remainder term which converges to 0 in probability, which can be re-expressed using the following equality which holds in law,

$$\sqrt{n}(\hat{\theta} - \theta_0) =^d 2 \left[I(\theta_0) \right]^{-1} \sqrt{n} \int_{-\infty}^{\infty} \left(\left[f_n(x) \right]^{\frac{1}{2}} - \left[f_{\theta_0}(x) \right]^{\frac{1}{2}} \right) \frac{\dfrac{\partial f_{\theta_0}(x)}{\partial \theta}}{\sqrt{f_{\theta_0}(x)}} \mathrm{d}x . \tag{22}$$

Using the argument given by Beran [7] (page 451) allows us to establish the equality in probability,

$$\begin{aligned} & \sqrt{n} \int_{-\infty}^{\infty} \left(\left[f_n(x) \right]^{\frac{1}{2}} - \left[f_{\theta_0}(x) \right]^{\frac{1}{2}} \right) \frac{\dfrac{\partial f_{\theta_0}(x)}{\partial \theta}}{\sqrt{f_{\theta_0}(x)}} \mathrm{d}x \\ & = \sqrt{n} \int_{-\infty}^{\infty} \frac{1}{2} \frac{\left(f_n(x) - f_{\theta_0}(x) \right)}{\sqrt{f_{\theta_0}(x)}} \frac{\dfrac{\partial f_{\theta_0}(x)}{\partial \theta}}{\sqrt{f_{\theta_0}(x)}} \mathrm{d}x + o_p(1) \end{aligned} \tag{23}$$

This can be viewed as a form of generalized delta method to establish equality of the left-hand side and the right-hand side of Equation (23).

Consequently, Equation (22) can be re-expressed, using the equality in distribution, as

$$\sqrt{n}(\hat{\theta} - \theta_0) =^d \left[I(\theta_0) \right]^{-1} \sqrt{n} \int_{-\infty}^{\infty} \left(f_n(x) - f_{\theta_0}(x) \right) \frac{\partial \log f_{\theta_0}(x)}{\partial \theta} \mathrm{d}x . \tag{24}$$

Note that

$$\int_{-\infty}^{\infty} \left(f_n(x) - f_{\theta_0}(x) \right) \frac{\partial \log f_{\theta_0}(x)}{\partial \theta} \mathrm{d}x = \int_{-\infty}^{\infty} f_n(x) \frac{\partial \log f_{\theta_0}(x)}{\partial \theta} \mathrm{d}x \tag{25}$$

as, in general, $\int_{-\infty}^{\infty} f_{\theta_0}(x) \frac{\partial \log f_{\theta_0}(x)}{\partial \theta} \mathrm{d}x = 0$. Furthermore,

$$\sqrt{n} \int_{-\infty}^{\infty} f_n(x) \frac{\partial \log f_{\theta_0}(x)}{\partial \theta} \mathrm{d}x = \sqrt{n} \int_{-\infty}^{\infty} \frac{\partial \log f_{\theta_0}(x)}{\partial \theta} \mathrm{d}F_n(x) + o_p(1) , \tag{26}$$

where $F_n(x)$ is the commonly used sample distribution function. This allows the following representation:

$$\sqrt{n}(\hat{\theta} - \theta_0) =^d \left[I(\theta_0) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f_{\theta_0}(x_i)}{\partial \theta} . \tag{27}$$

Therefore, $\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) \xrightarrow{L} N\left(0, I\left(\boldsymbol{\theta}_0\right)^{-1}\right)$.

For the simulated version, *i.e.*, version S, we can only obtain results for consistency and they will be given in the next section. As for asymptotic normality, we cannot conclude for the time being whether or not conditions of Theorem 7.1 given by Newey and McFadden [23] (pages 2185-2186) for the asymptotic normality of estimators obtained from a non-smooth function can be met. We hope to have more results on this issue in the future and would present them in a subsequent paper. This does not prevent SMHD estimation from being used as an alternative to methods of moments if the primary interests are in point estimation.

2.2. Asymptotic Properties of the SMHD Estimators

For version S, we minimize

$$Q_n(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \left(\left[f_n(x) \right]^{\frac{1}{2}} - \left[f_{\boldsymbol{\theta}}^S(x) \right]^{\frac{1}{2}} \right)^2 dx. \tag{28}$$

We recommend using the same seed across different values of $\boldsymbol{\theta}$ if possible and the simulated sample size $U = \tau n$ such that $U \to \infty$ at the same rate as $n \to \infty$. These recommendations conform with other simulated methods of inference such as the method of simulated moments as discussed by Davidson and McKinnon [9] (page 284) or simulated quasi-likelihood found in Smith [24] (page S68). The condition of the same seed being used is not necessary for consistency, but it allows $Q_n(\boldsymbol{\theta})$ to have the value for each $\boldsymbol{\theta}$ fixed each time we want to evaluate $Q_n(\boldsymbol{\theta})$; otherwise, the values might differ slightly due to the fact that simulations are needed to evaluate $Q_n(\boldsymbol{\theta})$. With the same seed, $Q_n(\boldsymbol{\theta})$ behaves like a non-random function with respect to $\boldsymbol{\theta}$.

Let

$$\left\| G_n(\boldsymbol{\theta}) \right\| = \left(Q_n(\boldsymbol{\theta}) \right)^{\frac{1}{2}}, \tag{29}$$

with $Q_n(\boldsymbol{\theta})$ as defined by Equation (28). The following Theorem, which is essentially Theorem (3.1) given by Pakes and Pollard [25] (page 1038) with the assumption of compactness of the parameter space added, can be used to establish the consistency of SMHD estimators. The proofs of the following Theorem have been given by Pakes and Pollard [25] (page 1038) using the Euclidean norm. Their proofs are still valid with the norm as defined by Equation (29) and discussed in Section 2.1. It is implicitly assumed that there is no identification problem for the parametric family, *i.e.*, if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $f_{\boldsymbol{\theta}_1}(x) \neq f_{\boldsymbol{\theta}_2}(x)$ except on a set of measure zero.

Theorem 2

Suppose

1) The parameter space $\boldsymbol{\Omega}$ is compact, and $\boldsymbol{\theta}_0 \in \boldsymbol{\Omega}$.

2) $\hat{\boldsymbol{\theta}}^S$ minimizes $\left\| G_n(\boldsymbol{\theta}) \right\|$ or equivalently $Q_n(\boldsymbol{\theta})$.

3) $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta} \left\| G_n(\boldsymbol{\theta}) \right\|^{-1}$ for each $\delta > 0$, is bounded in probability, where $\|\cdot\|$

denotes the norm being used.

Then $\hat{\boldsymbol{\theta}}^S \xrightarrow{\ p\ } \boldsymbol{\theta}_0$.

Clearly, we have consistency for $\hat{\boldsymbol{\theta}}^S$ as $0 \le Q_n(\boldsymbol{\theta}) \le 2$ and $Q_n(\boldsymbol{\theta}) \xrightarrow{\ p\ } 0$ only at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

For the time being, we cannot assert that $\hat{\boldsymbol{\theta}}^S$ follows a multivariate normal distribution asymptotically as we cannot verify the regularity conditions of Theorem 7.1 given by Newey and McFadden [23] (pages 2185-2186) for estimators obtained from a non-smooth objective function. For the simulated unweighted minimum chi-square, Pakes and Pollard [25] (page 1049) find the asymptotic covariance to be $\left(1 + \dfrac{1}{\tau}\right)V$, with V being the asymptotic covariance matrix of the estimators without using simulations. Conforming with other simulated methods which typically give the same type of asymptotic covariance formula, we recommend choosing $\tau \ge 10$ to minimize the loss of efficiency due to simulations. The matrix $\left(1 + \dfrac{1}{\tau}\right)V$, where $V = I(\boldsymbol{\theta}_0)^{-1}$, can be viewed as a form of benchmark for the approximate asymptotic covariance matrix for $\hat{\boldsymbol{\theta}}^S$ if indeed asymptotic normality can be shown. In the absence of a rigorous proof, we have to rely on simulations to evaluate the efficiency of $\hat{\boldsymbol{\theta}}^S$, just as for version D when the support of the distribution is not compact. Further asymptotic results to be obtained in the future will be presented in a subsequent paper.

Since we have estimates for densities, it is natural that we can estimate the Fisher information matrix. Clearly, if the model density has a closed-form expression, then the following matrix

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\dfrac{\partial f_{\hat{\boldsymbol{\theta}}^S}(x_i)}{\partial \boldsymbol{\theta}}}{f_{\hat{\boldsymbol{\theta}}^S}(x_i)}\right)\left(\frac{\dfrac{\partial f_{\hat{\boldsymbol{\theta}}^S}(x_i)}{\partial \boldsymbol{\theta}}}{f_{\hat{\boldsymbol{\theta}}^S}(x_i)}\right)' \tag{30}$$

can be used to estimate $I(\boldsymbol{\theta}_0)$. Instead of $f_{\hat{\boldsymbol{\theta}}^S}(x_i)$, if it is not available, we can use the kernel density estimate of $f_{\hat{\boldsymbol{\theta}}^S}(x_i)$, and, following a method given by Pakes and Pollard [25] (page 1043), we can use

$$\frac{\Delta f_{\hat{\boldsymbol{\theta}}^S}(x_i)}{\Delta \theta_j} = \frac{f_{\hat{\boldsymbol{\theta}}^S + \epsilon_n e_j}(x_i) - f_{\hat{\boldsymbol{\theta}}^S}(x_i)}{\epsilon_n}, \tag{31}$$

with $\epsilon_n \to 0$ at the rate $\epsilon_n = o(n^{-\delta})$, where $\delta \le \dfrac{1}{2}$, to estimate $\dfrac{\partial f_{\hat{\boldsymbol{\theta}}^S}(x_i)}{\partial \theta_j}, j = 1, \cdots, m$, assuming $f_{\hat{\boldsymbol{\theta}}^S}(x_i) > 0, i = 1, \cdots, n$. The vector e_j is a unit vector with 1 in its j-th place and 0 elsewhere. Replacing $f_{\hat{\boldsymbol{\theta}}^S}(x_i)$ and $\dfrac{\partial f_{\hat{\boldsymbol{\theta}}^S}(x_i)}{\partial \theta_j}$ by these estimates will give an estimator for the information matrix. An estimate of the information matrix is useful as the information matrix is related to the Cramer-Rao lower bound.

3. Limited Simulation Study

In this study, we shall compare the efficiencies of the moment estimators for the case with $\alpha = \beta$, *i.e.*, the GNL distribution with only four parameters. Reed [11] (page 477) has given the expressions for the moment estimators using the first six empirical cumulants $k_r, r = 2, \cdots, 6$, with the sample mean $\overline{X} = k_1$. They can be obtained using central empirical moments $m_r = \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2, r = 2, \cdots, 6$ as they follow the same type of relationships which exist between model cumulants κ_r and model central moments. Let $\mu_r = E\left(X - \mu\right)^2, r > 2$, and $\mu = E\left(X\right)$. The following relationships can be found in Stuart and Ord [26] (pages 90-91) and they are given by

$$\kappa_1 = \mu,$$
$$\kappa_2 = \mu_2,$$
$$\kappa_3 = \mu_3,$$
$$\kappa_4 = \mu_4 - 3\mu_2^2,$$
$$\kappa_5 = \mu_5 - 10\mu_3\mu_2,$$
$$\kappa_6 = \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3. \tag{32}$$

Explicitly, the moments estimators are

$$\tilde{\alpha} = \tilde{\beta} = \left(20\frac{k_4}{k_6}\right)^{\frac{1}{2}}, \quad \tilde{\rho} = \frac{100}{3}\frac{k_4^3}{k_6^2}, \quad \tilde{\mu} = \frac{k_1}{\tilde{\rho}} \quad \text{and} \quad \tilde{\sigma}^2 = \frac{k_2}{\tilde{\rho}} - \frac{2}{\tilde{\alpha}^2}. \tag{33}$$

Reed [11] (page 477) also notes that method of moments estimators (MM estimators) can take on negative values for positive parameters, and it is not easy to include constraints in method of moments estimation. Also, the use of EM algorithm does not appear to be straightforward for the GNL distribution. SMHD estimation can handle constraints by minimizing the objective function, which is given by Equation (28), with constraints.

A limited simulation study using parameters for the symmetric GNL distribution with four parameters, focusing on parameters in the ranges $\mu = 0$, $\sigma = 0.008$, $0.1 \le \rho \le 5.0$, $30 \le \alpha \le 40$, has been carried out and the relevant results are summarized in Table 1. The ranges of parameters as indicated are chosen accordingly and conform with the empirical study conducted by Reed [11] (page 481) using stock data. The simulated sample size for data is $n = 1000$ and the simulated sample size drawn from the model for SMHD estimation is $U = 10000$, hence with $\tau = 10$. It takes about twenty minutes on a laptop computer to obtain the estimators for $M = 50$ samples and, due to the limited computer capacity, we fix $M = 50$ samples for each combination of parameters for our study. As we only have access to laptop computers, the scale of the study is limited.

We noticed that the method of moments estimator for σ^2 is often negative and we set it equal to zero whenever this is the case, and the comparisons of effi-

ciencies use this version of the method of moments estimator. The density estimate is based on the built-in function of the package R with a rectangular kernel and default bandwidth based on the normal distribution. The overall asymptotic relative efficiency (*ARE*) used for comparisons is

$$ARE = \frac{MSE\left(\hat{\mu}^{S}\right)+MSE\left(\hat{\sigma}^{S}\right)+MSE\left(\hat{\alpha}^{S}\right)+MSE\left(\hat{\rho}^{S}\right)}{MSE\left(\tilde{\mu}\right)+MSE\left(\tilde{\sigma}\right)+MSE\left(\tilde{\alpha}\right)+MSE\left(\tilde{\rho}\right)}, \quad (34)$$

with $MSE\left(\hat{\theta}\right)$ being the commonly used mean square error of the estimator $\hat{\theta}$ and it is estimated using $M = 50$ samples for estimating the expression for *ARE* and the values of the estimated *ARE*'s using different sets of parameters are displayed in Table 1.

Despite the scope of the study being limited, it suggests that SMHD estimators perform much better than method of moments estimators overall for the ranges of parameters used in finance. The method of moments estimator for θ tends to perform better for small values of ρ and deteriorates rapidly as ρ grows larger with $ARE \to 0$ even for various parameter values that we tested which lie outside the ranges indicated above and not shown in Table 1. Table 1 is used for illustration and provides a summary of the key findings of the study. Also, in the ranges considered, the method of moments estimator for μ tends to perform better than its SMHD counterpart, but the overall efficiency of MM estimators still falls behind the overall efficiency of SMHD estimators in general as shown in Table 1. Clearly, more work needs to be done numerically and theoretically, but it shows the potential efficiencies of SMHD methods.

Table 1. Asymptotic relative efficiencies to compare SMHD estimators with MM estimators with $\sigma = 0.008$.

$\alpha\because\rho$	0.1	0.2	0.3	0.4	1.0	5.0
30	0.591517	0.230986	0.108643	0.091801	0.000000	0.002654
32	0.424564	0.086852	0.334888	0.034317	0.000003	0.000022
34	0.786910	0.382450	0.321474	0.059785	0.000011	0.014531
36	0.618782	0.991033	0.200291	0.001051	0.000020	0.000062
38	0.449053	0.317762	0.121347	0.069595	0.000001	0.005445
40	0.434306	0.472453	0.144194	0.008689	0.000000	0.000002

Note: Tabulated values are estimates of the asymptotic relative efficiencies of the SMHD estimators versus the MM estimators.

Individual ratios of mean square errors for some sets of parameters

$$\boldsymbol{\theta} = \left(\mu = 0, \sigma = 0.008, \alpha = 30, \rho = 0.1\right)'$$

$$\frac{MSE\left(\hat{\mu}^{S}\right)}{MSE\left(\tilde{\mu}\right)} = 22.9437, \quad \frac{MSE\left(\hat{\sigma}^{S}\right)}{MSE\left(\tilde{\sigma}\right)} = 0.8584, \quad \frac{MSE\left(\hat{\alpha}^{S}\right)}{MSE\left(\tilde{\alpha}\right)} = 0.5918,$$

$$\frac{MSE\left(\hat{\rho}^{S}\right)}{MSE\left(\tilde{\rho}\right)} = 0.0222, \quad ARE = 0.5915$$

$$\boldsymbol{\theta} = \left(\mu = 0, \sigma = 0.008, \alpha = 34, \rho = 0.3 \right)'$$

$$\frac{MSE\left(\hat{\mu}^S\right)}{MSE\left(\tilde{\mu}\right)} = 925.3334 \,, \quad \frac{MSE\left(\hat{\sigma}^S\right)}{MSE\left(\tilde{\sigma}\right)} = 0.6064 \,, \quad \frac{MSE\left(\hat{\alpha}^S\right)}{MSE\left(\tilde{\alpha}\right)} = 0.3240 \,,$$

$$\frac{MSE\left(\hat{\rho}^S\right)}{MSE\left(\tilde{\rho}\right)} = 0.0151 \,, \quad ARE = 0.3215$$

$$\boldsymbol{\theta} = \left(\mu = 0, \sigma = 0.008, \alpha = 40, \rho = 1 \right)'$$

$$\frac{MSE\left(\hat{\mu}^S\right)}{MSE\left(\tilde{\mu}\right)} = 1.3739 \,, \quad \frac{MSE\left(\hat{\sigma}^S\right)}{MSE\left(\tilde{\sigma}\right)} = 0.0503 \,, \quad \frac{MSE\left(\hat{\alpha}^S\right)}{MSE\left(\tilde{\alpha}\right)} = 0.0004 \,,$$

$$\frac{MSE\left(\hat{\rho}^S\right)}{MSE\left(\tilde{\rho}\right)} = 0.0000 \,, \quad ARE = 0.0000$$

4. Conclusion

As SMHD estimators remain consistent with minimum regularity conditions and despite the lack of results on asymptotic normality, the proposed method appears to be useful for fitting actuarial and financial models using continuous infinitely divisible distributions which arise from Lévy processes or continuous mixture distributions constructed using mixing operations, whenever it is not difficult to simulate from these distributions but the density functions of these distributions have no closed-form expressions. In many models, the proposed method appears to be more efficient than traditional methods such as the method of moments. The proposed method is not difficult to implement but methods based on simulations do not seem to receive much attention in finance and actuarial science. They might be considered as additional robust statistical techniques for analyzing empirical data, especially if point estimation is the main interest.

Acknowledgements

References

[1] McNeil, A.J., Frey, R. and Embrechts, P. (2005) Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, Princeton.

[2] McLachlan, G.J. and Krishnan, T. (2008) The EM Algorithm and Extensions. 2nd Edition, Wiley, Hoboken. https://doi.org/10.1002/9780470191613

[3] Küchler, U. and Tappe, S. (2008) Bilateral Gamma Distributions and Processes in Financial Mathematics. *Stochastic Processes and their Applications*, **118**, 261-283. https://doi.org/10.1016/j.spa.2007.04.006

[4] Küchler, U. and Tappe, S. (2013) Tempered Stable Distributions and Processes. *Stochastic Processes and Their Applications*, **123**, 4256-4293.

https://doi.org/10.1016/j.spa.2013.06.012

[5] Feuerverger, A. and McDunnough, P. (1981) On the Efficiency of Empirical Characteristic Function Procedures. *Journal of the Royal Statistical Society, Series B*, **43**, 20-27.

[6] Garcia, R., Renault, E. and Veredas, D. (2011) Estimation of Stable Distributions by Indirect Inference. *Journal of Econometrics*, **161**, 325-337.
https://doi.org/10.1016/j.jeconom.2010.12.007

[7] Beran, R. (1977) Minimum Hellinger Distance Estimates for Parametric Models. *The Annals of Statistics*, **5**, 445-463. https://doi.org/10.1214/aos/1176343842

[8] Tamura, R.N. and Boos, D.D. (1986) Minimum Hellinger Distance Estimation for Multivariate Location and Covariance. *Journal of the American Statistical Association*, **81**, 223-229. https://doi.org/10.1080/01621459.1986.10478264

[9] Davidson, R. and MacKinnon, J.G. (2004) Econometric Theory and Methods. Oxford University Press, New York.

[10] Basu, A., Shioya, H. and Park, C. (2011) Statistical Inference: The Minimum Distance Approach. Chapman and Hall, Boca Raton.

[11] Reed, W.J. (2007) Brownian-Laplace Motion and Its Use in Financial Modelling. *Communications in Statistics—Theory and Methods*, **36**, 473-484.
https://doi.org/10.1080/03610920601001766

[12] Schoutens, W. (2003) Lévy Processes in Finance: Pricing Financial Derivatives. Wiley, New York. https://doi.org/10.1002/0470870230

[13] Dufresne, F. and Gerber, H.U. (1993) The Probability of Ruin for the Inverse Gaussian and Related Processes. *Insurance: Mathematics and Economics*, **12**, 9-22.
https://doi.org/10.1016/0167-6687(93)90995-2

[14] Luong, A. (2016) Cramér-Von Mises Distance Estimation for Some Positive Infinitely Divisible Parametric Families with Actuarial Applications. *Scandinavian Actuarial Journal*, **2016**, 530-549. https://doi.org/10.1080/03461238.2014.977817

[15] Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2012) Loss Models: From Data to Decisions. 4th Edition, Wiley, Hoboken.

[16] Hogg, R.V., McKean, J.W. and Craig, A.T. (2013) Introduction to Mathematical Statistics. 7th Edition, Pearson, Boston.

[17] Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006) Robust Statistics: Theory and Methods. Wiley, Chichester. https://doi.org/10.1002/0470010940

[18] Lindsay, B.G. (1994) Efficiency versus Robustness: The Case for Minimum Hellinger Distance and Related Methods. *The Annals of Statistics*, **22**, 1081-1114.
https://doi.org/10.1214/aos/1176325512

[19] Lehmann, E.L. (1999) Elements of Large Sample Theory. Springer, New York.
https://doi.org/10.1007/b98855

[20] Rizzo, M.L. (2008) Statistical Computing with R. Chapman and Hall, Boca Raton.

[21] Toma, A. (2008) Minimum Hellinger Distance Estimators from the Johnson System. *Journal of Statistical Planning and Inference*, **138**, 803-816.
https://doi.org/10.1016/j.jspi.2007.05.033

[22] Scott, D.W. (2014) Multivariate Density Estimation: Theory, Practice and Visualization. 2nd Edition, Wiley, Hoboken.

[23] Newey, W.K. and McFadden, D. (1994) Large Sample Estimation and Hypothesis Testing. In: Engle, R.F. and McFadden, D.L., Eds., *Handbook of Econometrics*, Volume 4, North Holland, Amsterdam, 2111-2245.

[24] Smith Jr, A.A. (1993) Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions. *Journal of Applied Econometrics*, **8**, S63-S84. https://doi.org/10.1002/jae.3950080506

[25] Pakes, A. and Pollard, D. (1989) Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, **57**, 1027-1057. https://doi.org/10.2307/1913622

[26] Stuart, A. and Ord, K. (1994) Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory. 6th Edition, Edward Arnold, London.