Scientific Research Publishing

Asymptotic Normality Distribution of Simulated Minimum Hellinger Distance Estimators for Continuous Models

Andrew Luong, Claire Bilodeau

École d'actuariat, Université Laval, Québec, Canada

Email: Andrew.Luong@act.ulaval.ca, Claire.Bilodeau@act.ulaval.ca

Abstract

Certain distributions do not have a closed-form density, but it is simple to draw samples from them. For such distributions, simulated minimum Hellinger distance (SMHD) estimation appears to be useful. Since the method is distance-based, it happens to be naturally robust. This paper is a follow-up to a previous paper where the SMHD estimators were only shown to be consistent; this paper establishes their asymptotic normality. For any parametric family of distributions for which all positive integer moments exist, asymptotic properties for the SMHD method indicate that the variance of the SMHD estimators attains the lower bound for simulation-based estimators, which is based on the inverse of the Fisher information matrix, adjusted by a constant that reflects the loss of efficiency due to simulations. All these features suggest that the SMHD method is applicable in many fields such as finance or actuarial science where we often encounter distributions without closed-form density.

Keywords

Continuous Distribution, Kernel Density Estimate, Continuity in Probability, Differentiability in Probability, Hellinger Distance

1. Introduction

In actuarial science and finance, we often have to fit data with a distribution that is continuous. In several instances, though the distribution does not have a closed-form density, it is not complicated to simulate from it. Such distribution can be infinitely divisible. Also, new distributions can be created by means of a mixing mechanism.

For those distributions, Luong and Bilodeau [1] have introduced simulated minimum Hellinger distance (SMHD) estimation. They have shown that the SMHD estimators are consistent in general with less regularity conditions needed than the maximum likelihood estimators and that they have the potential to be robust and have high efficiency.

It is conjectured that, asymptotically, the SMHD estimators could attain the lower bound given by the Fisher information matrix adjusted by a factor which is a constant reflecting the loss of efficiency due to simulations from the parametric models. This constant can be expressed as $\left(1+\dfrac{1}{\tau}\right)$, with $\tau=\dfrac{U}{n}$ assumed to remain constant, where n is the original sample size of the data and U is the simulated sample size used to estimate the model density function or distribution. This factor also appears in various methods of estimation based on simulations and reflects the loss of efficiency due to the model density or distribution having to be estimated using a simulated sample drawn from the model distribution. Section 2 of the paper further discusses this factor which appears in simulated unweighted minimum chi-square method and simulated quasi-likelihood method.

In this paper, which can be viewed as a follow-up to the previous paper, we shall show that, indeed, under some regularity conditions, the SMHD estimators will follow an asymptotic normal distribution, and the asymptotic covariance matrix is given by the inverse of the Fisher information matrix adjusted by the constant $\left(1+\dfrac{1}{\tau}\right)$, as conjectured in Luong and Bilodeau [1]. Consequently, the SMHD estimators are fully efficient among the class of simulated estimators, just as the maximum likelihood (ML) estimators are in the classical set-up.

We shall closely follow the work of Tamura and Boos [2] and restrict ourselves to the case of the univariate parametric family to establish asymptotic normality of the SMHD estimators. Under those restrictions, Tamura and Boos worked with kernel density estimates with nonrandom bandwidths and the assumption that the parametric family has a closed-form density. Under those assumptions, they can relax the requirement that the parametric family needs a compact support, as given by Beran [3] in his seminal paper. Rather, they have obtained the result that, in general, if parametric families have positive integer moments of all orders, then the minimum Hellinger distance (MHD) estimators in the univariate case will have the same efficiency as the ML estimators and the asymptotic covariance matrix can be based on the Fisher information matrix just as for the ML estimators.

We shall call version D the version based on a parametric family having a closed-form density. We extend the results to a simulated version, version S, where the parametric family requires a density estimate using a random sample drawn from the parametric family as introduced in Luong and Bilodeau [1]. The results we obtain in this paper can be summarized as follows: under the same conditions required by Tamura and Boos [2], the SMHD estimators are fully ef-

ficient among the class of simulated estimators just as the MHD estimators are fully efficient when they are based on a parametric family with a closed-form density. It only requires an adjustment by a constant which depends on τ, the ratio between the sample size U drawn from the parametric family and the original sample size n of the data. Since the constant τ can be controlled, the efficiency of SMHD estimators will be close to that of the MHD estimators of the classical version D.

Furthermore, since minimum distance estimators are in general robust, it makes SMHD estimators applicable whenever there is a need for robustness and evidence that data are contaminated.

In actuarial science and finance, there are many useful densities without closed forms with semi-heavy tails which satisfy the requirements needed for asymptotic efficiency of the SMHD estimators. For examples in actuarial science, see Klugman, Panjer and Wilmot [4], or Luong [5]. For examples in finance, see Schoutens [6], or Grigoletto and Provasi [7].

To establish asymptotic normality, we shall also make use of Theorem 7.1, given by Newey and McFadden [8], that provides conditions and results for estimators obtained by minimizing or maximizing a nonsmooth objective function. In addition, we need the concepts of continuity in probability and differentiability in probability, as they will be used in this paper to justify the asymptotic distribution of SMHD estimators.

For count data, Luong, Bilodeau and Blier-Wong [9] used a similar approach to establish asymptotic normality for the SMHD estimators in the discrete case.

The paper is organized as follows. The classical version, version D, is re-examined in section 2. We also extract the relevant results given by Tamura and Boos [2] which are subsequently needed for developing the simulated version, *i.e.*, version S. Version S is studied in section 3. In section 3.1, we define the notions of continuity in probability and differentiability in probability. Those two notions are needed to apply the Theorem 7.1 given by Newey and McFadden [8]. Section 3.2 shows that the Theorem 7.1 is applicable to SMHD estimation. Hence, though the objective function to be minimized is nonsmooth, we can establish asymptotic normality for the SMHD estimators and show that the SMHD estimators attain the lower bound within the class of simulated estimators just as, for the parametric model, the MHD or ML estimators attain the lower bound based on the Fisher information matrix when simulations are not needed because the parametric model density has a closed-form expression.

2. Hellinger Distance Estimation: Classical Results

In this section, we shall review some of the results already established by Tamura and Boos [2] but will focus on the univariate set-up for Hellinger distance estimation. Subsequently, building on their work, we shall establish the asymptotic normality of Hellinger distance estimators for the simulated version, *i.e.*, asymptotic normality of the SMHD estimators, in section 3. SMHD estimators

have been introduced and consistency for version S has been established in our previous paper; see Luong and Bilodeau [1]. Hence, section 3 will complete the results already obtained.

We shall define some notation before restating Theorem 4.1 given by Tamura and Boos [2] as Theorem 1 below. Their theorem is for the multivariate case but, when restricted to the univariate case, some simplifications can be made as the bias term in their theorem will converge to zero in probability, *i.e.*, $\sqrt{n}B_n \xrightarrow{p} 0$, and thus can be ignored. Also, we want to use the notation that is directly related to the notions of Fisher information matrix and ML estimation.

We assume we have independent and identically distributed observations $X_1, \cdots, X_n$ from a parametric family $\{f_{\boldsymbol{\theta}}\}$, with $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)'$, and the true vector of parameters is denoted by $\boldsymbol{\theta}_0$. For version D as considered by Tamura and Boos [2], it leads to minimizing the following objective function:

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

where

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \omega \left(\frac{x - x_i}{h_n} \right) \tag{2}$$

is a kernel density estimate based on the sample with nonrandom bandwidth h_n, and ω is the kernel density used to obtain MHD estimators. This is version D and we shall state the relevant results in this section.

For version S, which we will consider in section 3, the model density $f_{\boldsymbol{\theta}}(x)$ is replaced by a density estimate $f_{\boldsymbol{\theta}}^S(x)$ that is constructed similarly to $f_n(x)$ but using a random sample of size $U = \tau n$ drawn from $f_{\boldsymbol{\theta}}(x)$ instead of the original sample given by the data. SMHD estimators are obtained by minimizing

$$Q_n^S(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \left\{ \left[f_n(x) \right]^{1/2} - \left[f_{\boldsymbol{\theta}}^S(x) \right]^{1/2} \right\}^2 \mathrm{d}x \tag{3}$$

and will be discussed in section 3.

Let $s_{\boldsymbol{\theta}} = (f_{\boldsymbol{\theta}})^{1/2}$ and denote the vector of its first partial derivatives by $\dot{s}_{\boldsymbol{\theta}}$ and the matrix of its second partial derivatives by $\ddot{s}_{\boldsymbol{\theta}}$. All the partial derivatives are assumed to be continuous with respect to $\boldsymbol{\theta}$.

It is easy to see that, if we can interchange the order of integration and differentiation, $\int_{-\infty}^{\infty} \ddot{f}_{\boldsymbol{\theta}}(x) \mathrm{d}x$ is an $m \times m$ matrix with entries 0, where $\ddot{f}_{\boldsymbol{\theta}}$ is the matrix of second partial derivatives of $f_{\boldsymbol{\theta}}$. It is assumed implicitly that this requirement is met subsequently.

From the definitions of $\dot{s}_{\boldsymbol{\theta}}$ and $\ddot{s}_{\boldsymbol{\theta}}$, and with the previous assumption, we have

$$\frac{\dot{s}_{\boldsymbol{\theta}}}{2(f_{\boldsymbol{\theta}})^{1/2}} = \frac{1}{4} \frac{\partial \log f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \tag{4}$$

and

$$-\int_{-\infty}^{\infty} \ddot{s}_{\boldsymbol{\theta}} \left[f_{\boldsymbol{\theta}}(x) \right]^{1/2} \mathrm{d}x = \int_{-\infty}^{\infty} (\dot{s}_{\boldsymbol{\theta}})(\dot{s}_{\boldsymbol{\theta}})' \mathrm{d}x = \frac{I(\boldsymbol{\theta})}{4} \,, \tag{5}$$

where $I(\boldsymbol{\theta})$ is the commonly used Fisher information matrix for ML estimation.

The following equalities might be used to derive Equations (4) and (5). Using differentiability rules, we have

$$\dot{s}_{\boldsymbol{\theta}} = \frac{1}{2(f_{\boldsymbol{\theta}})^{1/2}} \frac{\partial f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}$$

and

$$\ddot{s}_{\boldsymbol{\theta}} = -\frac{1}{4(f_{\boldsymbol{\theta}})^{3/2}} \frac{\partial f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \frac{\partial f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}'} + \frac{1}{2(f_{\boldsymbol{\theta}})^{1/2}} \frac{\partial^2 f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$$

if

$$E_{\boldsymbol{\theta}} \left[\frac{\partial^2 f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] = 0$$

and

$$E_{\boldsymbol{\theta}} \left[\frac{\partial \log f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \frac{\partial \log f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}'} \right] = I(\boldsymbol{\theta}),$$

by assuming that we can interchange the order of integration and differentiation, and $E_{\boldsymbol{\theta}}[\cdot]$ is the expectation of the expression inside the brackets under $f_{\boldsymbol{\theta}}$.

Now, we consider the following expression,

$$\boldsymbol{\psi}_{\boldsymbol{\theta}} = -\left\{ \int_{-\infty}^{\infty} \ddot{s}_{\boldsymbol{\theta}} \left[f_{\boldsymbol{\theta}}(x) \right]^{1/2} \mathrm{d}x \right\}^{-1} \frac{\dot{s}_{\boldsymbol{\theta}}}{2(f_{\boldsymbol{\theta}})^{1/2}}, \tag{6}$$

as given by Tamura and Boos [2] (page 225), which can be re-expressed as

$$\boldsymbol{\psi}_{\boldsymbol{\theta}} = \left[I(\boldsymbol{\theta}) \right]^{-1} \frac{\partial \log f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}, \tag{7}$$

where $I(\boldsymbol{\theta})$ is the Fisher information matrix and $\dfrac{\partial \log f_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}$ is the vector of the score functions.

Tamura and Boos [2] also define the expression $\rho_{\boldsymbol{\theta}} = \boldsymbol{\psi}_{\boldsymbol{\theta}} 2(f_{\boldsymbol{\theta}})^{1/2}$, which can be re-expressed as

$$\rho_{\boldsymbol{\theta}} = 4 \left[I(\boldsymbol{\theta}) \right]^{-1} \dot{s}_{\boldsymbol{\theta}}. \tag{8}$$

With these equalities and simplifications, we shall restate Theorem 4.1 given by Tamura and Boos [2] (pages 225-226) as Theorem 1, which is version D as given below. We shall also highlight some of the results in their proof to be used for version S in the next section together with Theorem 7.1 given by Newey and McFadden [8]. The proof of Theorem 1 can be found in the proof of Theorem 4.1 by Tamura and Boos [2].

Theorem 1

If we can find a sequence of positive numbers $\{\alpha_n\}$ with $\alpha_n \to \infty$ as $n \to \infty$, then, provided the following conditions 1 - 8 are met, the MHD estimators $\hat{\boldsymbol{\theta}}$ obtained by minimizing Equation (1) have an asymptotic normal dis-

tribution and attain the Cramer-Rao lower bound based on the Fisher information matrix, *i.e.*,

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

and $\hat{\boldsymbol{\theta}}$ is first-order as efficient as $\hat{\boldsymbol{\theta}}_{ML}$, where $\hat{\boldsymbol{\theta}}_{ML}$ is the vector of classical ML estimators.

Here are the eight conditions to meet:

1) The kernel density ω used to construct the density estimate has a compact support W and the bandwidth h_n used satisfies the property
$h_n + (nh_n)^{-1} \to 0$ as $n \to \infty$.

2) The parameter space Θ is compact and $\boldsymbol{\theta}_0$ is an interior point.

3) The parameterization of the model has no problem of identification, *i.e.*, if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $f_{\boldsymbol{\theta}_1} \neq f_{\boldsymbol{\theta}_2}$.

4) $n \sup_{t \in W} \Pr\left[\left|X - h_n t\right| > \alpha_n\right] \to 0$ as $n \to \infty$.

5) The ratio $\dfrac{\int_{-\alpha_n}^{\alpha_n} \left|\dfrac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \boldsymbol{\theta}}\right| dx}{\sqrt{nh_n}} \to 0$ as $n \to \infty$.

6) The sequence M_n with $M_n = \sup_{|x| \leq \alpha_n} \sup_{t \in W} \dfrac{f_{\boldsymbol{\theta}}(x + h_n t)}{f_{\boldsymbol{\theta}}(x)}$ is bounded as $n \to \infty$.

7) The Fisher information matrix $I(\boldsymbol{\theta})$ exists and we can interchange the order of differentiation and integration so that Equation (5) holds.

8) The function $s_{\boldsymbol{\theta}} = (f_{\boldsymbol{\theta}})^{1/2}$ has first partial derivatives vector $\dot{s}_{\boldsymbol{\theta}}$ and second partial derivatives matrix $\ddot{s}_{\boldsymbol{\theta}}$, and all the partial derivatives are continuous with respect to $\boldsymbol{\theta}$.

Conditions 1, 2 and 3 are standard and easily satisfied. Regarding condition 4, Tamura and Boos [2] commented that it is almost equivalent to $E\left[\left|X\right|^{1/l}\right] < \infty$, $l > 0$. It is not difficult to see this equivalence as we often restrict our attention to $\alpha_n = n^l$ and the use of Markov's type of inequality allows us to obtain this equivalence. Furthermore, Tamura and Boos [2] also commented that conditions 4 and 5 will be satisfied if X has all positive integer order moments. Despite this restriction, many useful distributions in finance fall into this category, even when the distribution has semi-heavy tails as mentioned earlier. Condition 6, imposed directly on the parametric family, is often met for parametric families encountered in practice. As for the bandwidth h_n, we can choose h_n such that $h_n \to 0$ and $\sqrt{n}h_n \to \infty$ as $n \to \infty$, thus meeting the requirements set in condition 1 for the bandwidth.

We shall follow these recommendations and will show, in section 3.2, that, for version S, the SHMD estimators given by the vector $\hat{\boldsymbol{\theta}}^S$ which minimizes the objective function as given by Equation (3) will have the following asymptotic normality distribution:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^S - \boldsymbol{\theta}_0\right) \xrightarrow{L} N\left(0, \left(1 + \frac{1}{\tau}\right)\left[I(\boldsymbol{\theta})\right]^{-1}\right).$$

The factor $\left(1+\dfrac{1}{\tau}\right)$ also appears in other simulated methods of inference. It is used to discount the efficiency of the minimum unweighted chi-square method to obtain the efficiency of the simulated version; see Pakes and Pollard [10] (page 1069). Similarly, it is used as a discount factor for obtaining the efficiency of the simulated quasi-likelihood method by discounting the efficiency of the related quasi-likelihood method; see Smith [11] (page S69). This factor can be interpreted as a universal adjusting factor when the true distribution is replaced by an estimate using simulated samples. With this interpretation, the SMHD estimators can be viewed as estimators which attain the lower bounds among the class of estimators based on simulated techniques.

Before we proceed, we would like to extract a few results given by Tamura and Boos [2] in their proof of Theorem 4.1. These results will be needed to prove asymptotic normality for version S. For version D, from the proof of their Theorem 4.1 and by taking into account the bias $\sqrt{n}B_n \overset{p}{\longrightarrow} 0$ for the univariate case, we have the following results using equality in distribution:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) =^d \sqrt{n}\int_{-\infty}^{\infty} \boldsymbol{\rho_\theta} \left\{\left[f_n(x)\right]^{1/2} - \left[f_{\boldsymbol{\theta}}(x)\right]^{1/2}\right\}\mathrm{d}x \tag{9}$$

$$=^d \sqrt{n}\int_{-\infty}^{\infty} \frac{\boldsymbol{\rho_\theta}}{2\left[f_{\boldsymbol{\theta}}(x)\right]^{1/2}}\left[f_n(x) - f_{\boldsymbol{\theta}}(x)\right]\mathrm{d}x \tag{10}$$

$$=^d \sqrt{n}\int_{-\infty}^{\infty} \boldsymbol{\psi_\theta}\left[f_n(x) - f_{\boldsymbol{\theta}}(x)\right]\mathrm{d}x \tag{11}$$

$$=^d \sqrt{n}\int_{-\infty}^{\infty} \boldsymbol{\psi_\theta}\,\mathrm{d}\left(F_n - F_{\boldsymbol{\theta}}\right), \tag{12}$$

where F_n is the commonly used sample distribution function and $F_{\boldsymbol{\theta}}$ is the model distribution function.

Now, under the commonly used assumption $E_{\boldsymbol{\theta}}\left[\dfrac{\partial \log f_{\boldsymbol{\theta}}(X)}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}$, justified if interchanging the order of integration and differentiation is permissible, the last equality can be re-expressed as

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) =^d \frac{1}{\sqrt{n}}\left[I(\boldsymbol{\theta})\right]^{-1}\sum_{i=1}^{n}\frac{\partial \log f_{\boldsymbol{\theta}}(X_i)}{\partial \boldsymbol{\theta}}, \tag{13}$$

from which we can see easily that $\hat{\boldsymbol{\theta}}$ is as efficient as $\hat{\boldsymbol{\theta}}_{ML}$. Besides, $\hat{\boldsymbol{\theta}}$ is robust, whereas that may not be the case for $\hat{\boldsymbol{\theta}}_{ML}$. We can also see that, using Equations (9)-(13), we have the following equalities:

$$\sqrt{n}\,2\int_{-\infty}^{\infty} \dot{\boldsymbol{s}}_{\boldsymbol{\theta}}\left\{\left[f_n(x)\right]^{1/2} - \left[f_{\boldsymbol{\theta}}(x)\right]^{1/2}\right\}\mathrm{d}x$$
$$=^d \sqrt{n}\int_{-\infty}^{\infty} \frac{\dot{\boldsymbol{s}}_{\boldsymbol{\theta}}}{\left[f_{\boldsymbol{\theta}}(x)\right]^{1/2}}\left[f_n(x) - f_{\boldsymbol{\theta}}(x)\right]\mathrm{d}x \tag{14}$$

$$=^d \sqrt{n}\int_{-\infty}^{\infty} \frac{1}{2}\frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \boldsymbol{\theta}}\left[f_n(x) - f_{\boldsymbol{\theta}}(x)\right]\mathrm{d}x \tag{15}$$

$$=^d \frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f_{\boldsymbol{\theta}}(X_i)}{\partial \boldsymbol{\theta}} \tag{16}$$

Beran [3] had obtained these results earlier but using a compact support assumption for the parametric family and random bandwidths; see Equations (3.7) and (3.12) given by Beran [3] (pages 451-452).

3. Asymptotic Normality Distribution for SMHD Estimators

The equalities given by Equations (9)-(16) will be used for establishing asymptotic normality for SMHD estimators in section 3.2. A few notions, namely the notions of continuity in probability and differentiability in probability, which extend the related notions in classical real analysis for nonrandom functions, are needed and they will be presented in section 3.1.

3.1. Some Preliminary Notions

These notions have been introduced and discussed for SMHD estimation for count data, see Luong, Bilodeau and Blier-Wong [9] (pages 201-203). They are reproduced below to make it easier to follow the results of this paper and make the paper more self-contained.

Definition 1 (Continuity in probability)

A sequence of random functions $\{g_n(\boldsymbol{\theta})\}$ is continuous in probability at $\boldsymbol{\theta}^*$ if $g_n(\boldsymbol{\theta}) \xrightarrow{p} g_n(\boldsymbol{\theta}^*)$ whenever $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$. Equivalently, for any $\varepsilon > 0$ and $\delta_1 > 0$, there exist $\delta \geq 0$ and n_0 such that

$$\Pr\left[\left|g_n(\boldsymbol{\theta}) - g_n(\boldsymbol{\theta}^*)\right| \leq \varepsilon\right] \geq 1 - \delta_1, \text{ for } n \geq n_0,$$

whenever $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \leq \delta$. This can be viewed as a stochastic version, or an extension, of the classical definition of continuity in real analysis.

It is well known that the supremum of a continuous function on a compact domain is attained at a point in the compact domain; see Davidson and Donsig [12] (page 81) or Rudin [13] (page 89) for this classical result. The equivalent property for a random function which is only continuous in probability is that the supremum of the random function is attained at a point in the compact domain in probability. This property will be given as property 1 given below.

In order to use Theorem 7.1 of Newey and McFadden [8], we need to consider the compact domain of the form

$$S(\boldsymbol{\theta}_0, \delta_n) = \left\{\boldsymbol{\theta} \middle| \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta_n\right\},$$

and we note that, as $n \to \infty$, $\delta_n \to 0$, and $S(\boldsymbol{\theta}_0, \delta_n) \to \boldsymbol{\theta}_0$.

Property 1

The random function $g_n(\boldsymbol{\theta})$, which is continuous in probability and bounded in probability on a compact set Θ, will attain its supremum on a point of Θ in probability.

The justification of this property is similar to the deterministic case, which is a classical result in real analysis. For the random case, again, it suffices to pick a sequence $\{\boldsymbol{\theta}_j\}$ in Θ with the property that $g_n(\boldsymbol{\theta}_j) \xrightarrow{p} \sup_{\boldsymbol{\theta} \in \Theta} g_n(\boldsymbol{\theta})$. Since Θ is compact, we can extract a subsequence $\{\boldsymbol{\theta}_{j_k}\}$ from $\{\boldsymbol{\theta}_j\}$ having the property $\boldsymbol{\theta}_{j_k} \to \boldsymbol{\theta}^{**}$ which belongs to Θ. Then $g_n(\boldsymbol{\theta}_{j_k}) \xrightarrow{p} g_n(\boldsymbol{\theta}^{**})$ and

$\sup_{\boldsymbol{\theta} \in \Theta} g_n(\boldsymbol{\theta}) =^p g_n(\boldsymbol{\theta}^{**})$.

Beside the concept of continuity in probability, we also need the concept of differentiability in probability which is given below.

Definition 2 (Differentiability in probability)

A sequence of random functions $\{g_n(\boldsymbol{\theta})\}$ is differentiable with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$ in probability if $\lim_{\varepsilon \to 0} \dfrac{g_n(\boldsymbol{\theta} + \varepsilon e_j) - g_n(\boldsymbol{\theta})}{\varepsilon} =^p v_n^{(j)}(\boldsymbol{\theta})$, for $j = 1, 2, \cdots, m$, where $e_j = (0, 0, \cdots, 0, 1, 0, \cdots, 0)'$ with 1 appearing only in the jth entry. We also require that $v_n^{(j)}(\boldsymbol{\theta})$ be continuous in probability for $j = 1, 2, \cdots, m$.

We can let the derivatives vector be denoted as $v_n(\boldsymbol{\theta}) = \left(v_n^{(1)}(\boldsymbol{\theta}), \cdots, v_n^{(m)}(\boldsymbol{\theta}) \right)'$.

From definition 2, we can see that differentiability in probability is a notion which parallels the classical notion of differentiability, where each partial derivative of the nonrandom function is required to be continuous.

A similar notion of differentiability in probability has been used in the stochastic processes literature; see Gusak *et al.* [14]. A more stringent differentiability notion, namely differentiability in quadratic means, has been introduced to study the local asymptotic normality property for a parametric family; see Keener [15] (page 326). Also, see Pollard [16] for the notion of stochastic differentiability, which is also more stringent than differentiability in probability.

Below are the assumptions we need to make to establish asymptotic normality for SMHD estimators in section 3.2, and they appear to be satisfied in general.

For the simulated version, we implicitly assume that the sample size U used to draw samples from the parametric family $\{f_{\boldsymbol{\theta}}\}$ is proportional to the sample size n, *i.e.*, $U = \tau n$. Moreover, the same seed is used across different values of $\boldsymbol{\theta}$ to draw the simulated samples. Under those assumptions, $\{f_{\boldsymbol{\theta}}^S\}$ can be viewed as a proxy model for $\{f_{\boldsymbol{\theta}}\}$. However, unlike other methods of simulated inference that require that we look for another parametric model that is different from the model $\{f_{\boldsymbol{\theta}}\}$, the proxy model here is directly based on the parametric model.

Assumption 1

The density of the parametric model has the continuity property with $(f_{\boldsymbol{\theta}})^{1/2} \to (f_{\boldsymbol{\theta}^*})^{1/2}$ whenever $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$.

Assumption 2

The simulated counterpart has the continuity in probability property with $(f_{\boldsymbol{\theta}}^S)^{1/2} \xrightarrow{p} (f_{\boldsymbol{\theta}^*}^S)^{1/2}$ whenever $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$.

In general, assumption 2 is met if assumption 1 is.

Assumption 3

$(f_{\boldsymbol{\theta}})^{1/2}$ is differentiable with respect to $\boldsymbol{\theta}$.

We also need the following assumption for applying Theorem 7.1 given by Newey and McFadden [8] (pages 2185-2186).

Assumption 4

$(f_{\boldsymbol{\theta}}^S)^{1/2}$, with the same seed being used across different values of $\boldsymbol{\theta}$, is differentiable in probability with the same derivatives vector as $\sqrt{f_{\boldsymbol{\theta}}}$, namely

$$\dot{\boldsymbol{s}}_{\boldsymbol{\theta}} = \left(\frac{\partial \sqrt{f_{\boldsymbol{\theta}}}}{\partial \theta_1}, \cdots, \frac{\partial \sqrt{f_{\boldsymbol{\theta}}}}{\partial \theta_m} \right). \tag{17}$$

This assumption appears to be reasonable as $\left(f_{\boldsymbol{\theta}}^S \right)^{1/2} \overset{p}{\longrightarrow} \left(f_{\boldsymbol{\theta}} \right)^{1/2}$ and $\left(f_{\boldsymbol{\theta}}^S \right)^{1/2}$ is continuous in probability. Also, the partial derivatives in probability can be found using definition 2, which involves considering limits which are similar to the deterministic case of real analysis. We can summarize the assumption as follows: by assuming $\left(f_{\boldsymbol{\theta}} \right)^{1/2}$ to be differentiable, we have that $\left(f_{\boldsymbol{\theta}}^S \right)^{1/2}$ is differentiable in probability. This appears to be reasonable.

3.2. Asymptotic Normality

For version S, because the objective function to be minimized is nonsmooth, we will use Theorem 7.1 of Newey and McFadden [8] (pages 2185-2186) to establish the asymptotic normality for SMHD estimators. The ideas behind Theorem 7.1 can be summarized as follows.

The objective function $Q_n^S (\boldsymbol{\theta})$ is nonsmooth and the estimators are given by the vector $\tilde{\boldsymbol{\theta}}$ which is obtained by minimizing $Q_n^S (\boldsymbol{\theta})$. We can consider the vector $\boldsymbol{\theta}^*$ which is obtained by minimizing a smooth function $Q_n^a (\boldsymbol{\theta})$ which approximates $Q_n^S (\boldsymbol{\theta})$. We assume $Q_n^S (\boldsymbol{\theta})$ is differentiable in probability at $\boldsymbol{\theta}_0$, with the derivatives vector given by $\boldsymbol{D}_n (\boldsymbol{\theta}_0)$.

Also, if $Q_n^S (\boldsymbol{\theta}) \overset{p}{\longrightarrow} Q(\boldsymbol{\theta})$ and we assume that $Q(\boldsymbol{\theta})$ is nonrandom and twice differentiable with $H = H(\boldsymbol{\theta}_0) = \dfrac{\partial^2 Q(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$, and $Q(\boldsymbol{\theta})$ attains its minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, then we can define

$$Q_n^a (\boldsymbol{\theta}) = Q_n^S (\boldsymbol{\theta}_0) + \left[\boldsymbol{D}_n (\boldsymbol{\theta}_0) \right]' (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' H (\boldsymbol{\theta} - \boldsymbol{\theta}_0). \tag{18}$$

The vector $\boldsymbol{\theta}^*$ which minimizes $Q_n^a (\boldsymbol{\theta})$ can be obtained explicitly since $Q_n^a (\boldsymbol{\theta})$ is a quadratic function of $\boldsymbol{\theta}$. It is given by $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0 - H^{-1} \boldsymbol{D}_n (\boldsymbol{\theta}_0)$. Using equality in distribution, we have

$$\sqrt{n} \left(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0 \right) =^d -H^{-1} \sqrt{n} \boldsymbol{D}_n (\boldsymbol{\theta}_0). \tag{19}$$

If the remainder of the approximation is small, we also have

$$\sqrt{n} \left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right) =^d \sqrt{n} \left(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0 \right) =^d -H^{-1} \sqrt{n} \boldsymbol{D}_n (\boldsymbol{\theta}_0). \tag{20}$$

Before defining the remainder term $R_n (\boldsymbol{\theta})$, we note that the following approximation $Q_n^b (\boldsymbol{\theta})$,

$$Q_n^b (\boldsymbol{\theta}) = Q_n^S (\boldsymbol{\theta}_0) + \left[\boldsymbol{D}_n (\boldsymbol{\theta}_0) \right]' (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0), \tag{21}$$

can be viewed as equivalent since $Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0) \approx \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' H (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ when we account for the fact that, since $Q(\boldsymbol{\theta})$ is minimized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\dfrac{\partial Q(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = 0$.

The remainder term is defined as

$$R_n(\boldsymbol{\theta}) = \frac{\sqrt{n}\left\{\left[Q_n^S(\boldsymbol{\theta}) - Q_n^S(\boldsymbol{\theta}_0)\right] - \left[\boldsymbol{D}_n(\boldsymbol{\theta}_0)\right]'(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \left[Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0)\right]\right\}}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|}, \quad (22)$$

and, for the approximation to be valid, we require that

$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta_n}|R_n(\boldsymbol{\theta})| \xrightarrow{p} 0$ as $n \to \infty$ and $\delta_n \to 0$.

The following Theorem 2 is essentially Theorem 7.1 given by Newey and McFadden [8], except for the following three differences. First, we restate it for estimators obtained by minimizing, instead of maximizing, an objective function. Second, we require $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta_n}|R_n(\boldsymbol{\theta})| \xrightarrow{p} 0$, which is slightly more stringent than the original condition (v) of their Theorem 7.1. Third, we require compactness of the parameter space Θ.

Theorem 2

Suppose that $Q_n^S(\tilde{\boldsymbol{\theta}}) \le \inf_{\boldsymbol{\theta}\in\Theta} Q_n^S(\boldsymbol{\theta}) + o_p\left(\frac{1}{n}\right)$, and

1) $Q(\boldsymbol{\theta})$ is minimized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$;
2) $\boldsymbol{\theta}_0$ is an interior point of the compact parameter space Θ;
3) $Q(\boldsymbol{\theta})$ is twice differentiable at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ with nonsingular matrix H;
4) $\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0) \xrightarrow{L} N(0, K)$;
5) $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta_n}|R_n(\boldsymbol{\theta})| \xrightarrow{p} 0$ as $n \to \infty$ and $\delta_n \to 0$.

Then, $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} N(0, H^{-1}KH^{-1})$.

Regularity conditions 1 - 3 of Theorem 2 can easily be checked. Condition 4 is a consequence of the results already obtained for version D. The most difficult condition to be verified is condition 5. Because it involves technicalities, its verification will be done toward the end of this section.

Here, assuming all conditions can be validated, we apply Theorem 2 for SMHD estimation with $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^S$.

Clearly, the objective function $Q_n^S(\boldsymbol{\theta})$ is as defined by Equation (3) and

$$Q_n^S(\boldsymbol{\theta}) \xrightarrow{p} Q(\boldsymbol{\theta}) = \int_{-\infty}^{\infty}\left\{\left[f_{\boldsymbol{\theta}_0}(x)\right]^{1/2} - \left[f_{\boldsymbol{\theta}}(x)\right]^{1/2}\right\}^2 dx, \quad (23)$$

with $(f_n)^{1/2} \xrightarrow{p} (f_{\boldsymbol{\theta}_0})^{1/2}$ and $(f_{\boldsymbol{\theta}}^S)^{1/2} \xrightarrow{p} (f_{\boldsymbol{\theta}})^{1/2}$ as $n \to \infty$.

Under the differentiability assumptions for $s_{\boldsymbol{\theta}}$, the second derivatives matrix of $Q(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is given by $H = \frac{\partial^2 Q(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = 2\int_{-\infty}^{\infty}(\dot{s}_{\boldsymbol{\theta}})(\dot{s}_{\boldsymbol{\theta}})' dx = \frac{1}{2}I(\boldsymbol{\theta})$.

Therefore,

$$H^{-1} = 2\left[I(\boldsymbol{\theta}_0)\right]^{-1}. \quad (24)$$

Using assumptions 1 - 4 and by performing limit operations as for finding derivatives in real analysis, we can conclude $\sqrt{n}Q_n^S(\boldsymbol{\theta})$ is differentiable in probability with derivatives vector $\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0)$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and

$$\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0) = -\sqrt{n}2\int_{-\infty}^{\infty}\left\{\left[f_n(x)\right]^{1/2} - \left[f_{\boldsymbol{\theta}_0}^S(x)\right]^{1/2}\right\}\dot{s}_{\boldsymbol{\theta}_0} dx, \quad (25)$$

which can be re-expressed as

$$-\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0) = \sqrt{n}2\int_{-\infty}^{\infty}\left\{\left[f_n(x)\right]^{1/2} - \left[f_{\boldsymbol{\theta}_0}(x)\right]^{1/2}\right\}\dot{\boldsymbol{s}}_{\boldsymbol{\theta}_0}\,\mathrm{d}x$$
$$-\sqrt{n}2\int_{-\infty}^{\infty}\left\{\left[f_{\boldsymbol{\theta}_0}^S(x)\right]^{1/2} - \left[f_{\boldsymbol{\theta}_0}(x)\right]^{1/2}\right\}\dot{\boldsymbol{s}}_{\boldsymbol{\theta}_0}\,\mathrm{d}x \tag{26}$$

Let us denote the first and second terms of Equation (26) by $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ respectively. Then $-\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0) = \boldsymbol{Y}_1 - \boldsymbol{Y}_2$, with $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ being independent since the simulated sample is independent from the original sample. $-\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0)$ will have a limit distribution, hence it is bounded in probability and $-\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta})$ will be continuous in probability for $\boldsymbol{\theta} \in S(\boldsymbol{\theta}_0, \delta_n) = \{\boldsymbol{\theta}\|\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \le \delta_n\}$ for $n \ge n_0$, where n_0 is a positive integer, by invoking the Dominated Convergence Theorem if necessary.

We have $-H^{-1}\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0) =^d H^{-1}\boldsymbol{Y}_1 - H^{-1}\boldsymbol{Y}_2$, and, using Equations (24)-(26), we can conclude

$$-H^{-1}\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0) \xrightarrow{d} N\left(0, \left(1+\frac{1}{\tau}\right)\left[I(\boldsymbol{\theta}_0)\right]^{-1}\right), \tag{27}$$

which implies

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^S - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(0, \left(1+\frac{1}{\tau}\right)\left[I(\boldsymbol{\theta}_0)\right]^{-1}\right). \tag{28}$$

From the results given by Equations (27) and (28), asymptotic properties suggest that the SMHD estimators will have high efficiency in large samples as the lower bound for simulated estimators is attained. We should also keep $\tau \ge 10$ if possible to minimize the loss of efficiency due to simulations and the same seed should be used to generate simulated samples across different values of $\boldsymbol{\theta}$.

To assess the performance of the SMHD estimators in finite samples, we need simulation studies which are based on the parametric family being considered as asymptotic theory, despite being quite general, might not be applicable for finite samples, especially with sample size $n \le 100$.

We now proceed to verify $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta_n}|R_n(\boldsymbol{\theta})| \xrightarrow{p} 0$, where

$$R_n(\boldsymbol{\theta}) = \frac{\sqrt{n}\left\{\left[Q_n^S(\boldsymbol{\theta}) - Q_n^S(\boldsymbol{\theta}_0)\right] - \left[\boldsymbol{D}_n(\boldsymbol{\theta}_0)\right]'(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - \left[Q(\boldsymbol{\theta})-Q(\boldsymbol{\theta}_0)\right]\right\}}{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|}.$$

Once again, $S(\boldsymbol{\theta}_0, \delta_n) = \{\boldsymbol{\theta}\|\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \le \delta_n\}$ is a shrinking compact set: we note that, as $n \to \infty$, $\delta_n \to 0$, and $S(\boldsymbol{\theta}_0, \delta_n) \to \boldsymbol{\theta}_0$.

In order to confirm that Theorem 2 is applicable, we need to study the properties of $R_n(\boldsymbol{\theta})$. Given that $R_n(\boldsymbol{\theta})$ can be defined at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ as $R_n(\boldsymbol{\theta}_0) = 0$, we would like to establish the following for $R_n(\boldsymbol{\theta})$:

1) $R_n(\boldsymbol{\theta})$ is bounded in probability.

2) $R_n(\boldsymbol{\theta})$ is continuous in probability for all $\boldsymbol{\theta} \in S(\boldsymbol{\theta}_0, \delta_n)$, for $n \ge n_0$.

Clearly, if conditions 1 and 2 hold, then $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta_n}|R_n(\boldsymbol{\theta})|$ is attained at a point $\boldsymbol{\theta}_n^* \in S(\boldsymbol{\theta}_0, \delta_n)$ in probability for $n \ge n_0$ as $R_n(\boldsymbol{\theta})$ is continuous in probability. It would then follow that, for $n \ge n_0$, we have the following equality

in probability:

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta_n}\left|R_n(\boldsymbol{\theta})\right|=^p\left|R_n(\boldsymbol{\theta}_n^*)\right|,$$

with $\left|R_n(\boldsymbol{\theta}_n^*)\right|\xrightarrow{p}\left|R_n(\boldsymbol{\theta}_0)\right|=0$ as $\boldsymbol{\theta}_n^*\to\boldsymbol{\theta}_0$. Therefore, Theorem 2 will be justified for SMHD estimators.

We still need to establish the above results.

Define $\sqrt{n}G_n(\boldsymbol{\theta})=\sqrt{n}\left[Q_n^S(\boldsymbol{\theta})-Q(\boldsymbol{\theta})\right]$. Then, clearly, $\sqrt{n}G_n(\boldsymbol{\theta})$ is differentiable in probability at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ with the derivatives vector given by

$$\left[\boldsymbol{D}_n(\boldsymbol{\theta}_0)\right]'\quad\text{since}\quad\frac{\partial Q(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}=0\quad\text{and}\quad Q(\boldsymbol{\theta}_0)=0.$$

We also have $R_n(\boldsymbol{\theta})\xrightarrow{p}0$ as $\boldsymbol{\theta}\to\boldsymbol{\theta}_0$. If we define $R_n(\boldsymbol{\theta}_0)=0$, we can extend $R_n(\boldsymbol{\theta})$ to be continuous in probability for $\boldsymbol{\theta}\in S(\boldsymbol{\theta}_0,\delta_n)$, including the point $\boldsymbol{\theta}=\boldsymbol{\theta}_0$, for $n\ge n_0$.

The expression $\sqrt{n}G_n(\boldsymbol{\theta})-\sqrt{n}G_n(\boldsymbol{\theta}_0)$ can be assumed to be bounded in probability in a neighborhood of $\boldsymbol{\theta}_0$, since it can be approximated by $\sqrt{n}\left[\boldsymbol{D}_n(\boldsymbol{\theta}_0)\right]'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$ and $\sqrt{n}\boldsymbol{D}_n(\boldsymbol{\theta}_0)$ has a limit distribution as discussed earlier; see Equations (26) and (27).

Note that we can also write

$$\sqrt{n}G_n(\boldsymbol{\theta})-\sqrt{n}G_n(\boldsymbol{\theta}_0)-\sqrt{n}\left[\boldsymbol{D}_n(\boldsymbol{\theta}_0)\right]'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$$
$$=\sqrt{n}\left\{\left[Q_n^S(\boldsymbol{\theta})-Q_n^S(\boldsymbol{\theta}_0)\right]-\left[\boldsymbol{D}_n(\boldsymbol{\theta}_0)\right]'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)-\left[Q(\boldsymbol{\theta})-Q(\boldsymbol{\theta}_0)\right]\right\}$$

Therefore, it is not difficult to see that the above expression is continuous in probability. As a result, $R_n(\boldsymbol{\theta})$ is bounded in probability and continuous in probability, and so is $\left|R_n(\boldsymbol{\theta})\right|$.

Hence, the use of Theorem 2 is justified for the SMHD estimators.

Moreover, SMHD estimators are robust as they are obtained by minimizing a distance; see Donoho and Liu [17] and Lindsay [18].

4. Conclusion

Asymptotic properties established in this paper suggest that SMHD estimators are very efficient for large samples for parametric models where all the positive integer moment exists. For the subset of such parametric models that have no closed-form densities, as often are encountered in finance and actuarial science, SMHD estimators appear to be very suitable for large samples based on asymptotic normality results obtained. For any parametric family failing to have finite moments of all positive integer orders, SMHD estimators remain consistent and robust, but large-scale simulation studies seem to be necessary to study the efficiency of the estimators for the specific parametric model being considered.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Luong, A. and Bilodeau, C. (2017) Simulated Minimum Hellinger Distance Estimation for Some Continuous Financial and Actuarial Models. *Open Journal of Statistics*, **7**, 743-759. https://doi.org/10.4236/ojs.2017.74052

[2] 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

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

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

[5] 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

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

[7] Grigoletto, M. and Provasi, C. (2008) Simulation and Estimation of the Meixner Distribution. *Communications in Statistics—Simulation and Computation*, **38**, 58-77. https://doi.org/10.1080/03610910802395679

[8] 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.

[9] Luong, A., Bilodeau, C. and Blier-Wong, C. (2018) Simulated Minimum Hellinger Distance Inference Methods for Count Data. *Open Journal of Statistics*, **8**, 187-219. https://doi.org/10.4236/ojs.2018.81012

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

[11] 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

[12] Davidson, K.R. and Donsig, A.P. (2010) Real Analysis and Applications: Theory in Practice. Springer, New York. https://doi.org/10.1007/978-0-387-98098-0

[13] Rudin, W. (1976) Principles of Mathematical Analysis. 3rd Edition, McGraw-Hill, New York.

[14] Gusak, D., Kukush, A., Kulik, A., Mishura, Y. and Pilipenko, A. (2010) Theory of Stochastic Processes: with Applications to Financial Mathematics and Risk Theory. Springer, New York. https://doi.org/10.1007/978-0-387-87862-1

[15] Keener, R.W. (2010) Theoretical Statistics: Topics for a Core Course. Springer, New York. https://doi.org/10.1007/978-0-387-93839-4

[16] Pollard, D. (1985) New Ways to Prove Central Limit Theorems. *Econometric Theory*, **1**, 295-313. https://doi.org/10.1017/S0266466600011233

[17] Donoho, D.L. and Liu, R.C. (1988) The "Automatic" Robustness of Minimum Distance Functionals. *The Annals of Statistics*, **16**, 552-586.

https://doi.org/10.1214/aos/1176350820

[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