

Simulated Minimum Cramér-Von Mises Distance Estimation for Some Actuarial and Financial Models

Andrew Luong*, Christopher Blier-Wong

École d'Actuariat, Université Laval, Québec, Canada

Email: *andrew.luong@act.ulaval.ca

How to cite this paper: Luong, A. and Blier-Wong, C. (2017) Simulated Minimum Cramér-Von Mises Distance Estimation for Some Actuarial and Financial Models. *Open Journal of Statistics*, 7, 815-833. <https://doi.org/10.4236/ojs.2017.75058>

Received: September 13, 2017

Accepted: October 22, 2017

Published: October 25, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Minimum Cramér-Von Mises distance estimation is extended to a simulated version. The simulated version consists of replacing the model distribution function with a sample distribution constructed using a simulated sample drawn from it. The method does not require an explicit form of the model density functions and can be applied to fitting many useful infinitely divisible distributions or mixture distributions without closed form density functions often encountered in actuarial science and finance. For these models likelihood estimation is difficult to implement and simulated Minimum Cramér-Von Mises (SMCVM) distance estimation can be used. Asymptotic properties of the SCVM estimators are established. The new method appears to be more robust and efficient than methods of moments (MM) for the models being considered which have more than two parameters. The method can be used as an alternative to simulated Hellinger distance (SMHD) estimation with a special feature: it can handle models with a discontinuity point at the origin with probability mass assigned to it such as in the case of the compound Poisson distribution where SMHD method might not be suitable. As the method is based on sample distributions instead of density estimates it is also easier to implement than SMHD method but it might not be as efficient as SMHD methods for continuous models.

Keywords

Compound Poisson Distribution, Double Exponential Jump Diffusion Distribution, Mixture Distribution, Robustness, Influence Functions

1. Introduction

In actuarial science or finance we often model losses or log-returns with distri-

bution functions where neither the distribution function nor its corresponding density function has a closed form expression yet it is not complicated to draw random samples from these distributions. It is clear that likelihood methods are complicated in such a situation.

For statistical inferences using models with these features, we shall assume to have independent and identically distributed (iid) observations X_1, \dots, X_n which have a common distribution as X with model distribution and density given respectively by $F_{\theta}(u)$ and $f_{\theta}(u)$. Neither $F_{\theta}(u)$ nor $f_{\theta}(u)$ has a closed form expression but often its moment generating function (mgf) $M_{\theta}(s)$ has a closed form expression. The vector of parameters of interest is

$$\theta = (\theta_1, \dots, \theta_m)'$$

The compound Poisson distribution used in actuarial sciences and jump diffusion distribution in finance are typical examples for these types of models. Furthermore, in many circumstances distributions derived from the increments of Lévy processes also display these characteristics and it is of interest to make inferences for the vector of parameters. We shall illustrate the situation with example 1 and example 2 below.

Example 1

In this example, we shall consider the compound Poisson gamma distribution which is commonly used in actuarial science and it arises from the compound Poisson processes which also belong to the class of Lévy processes.

The compound gamma distribution is the distribution of a random variable X representable as a random sum, *i.e.*,

$$X = \sum_{i=1}^N Y_i \text{ with the } Y_i \text{'s being iid with a common gamma distribution with}$$

the density function given by $f_Y(y) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}}$, $y > 0$, $\alpha > 1$, $\beta > 0$ and

the moment generating function is given by $M_Y(s) = \frac{1}{(1-\beta s)^\alpha}$. The random

variable N follows a Poisson distribution with parameter $\lambda > 0$ and it is assumed that the Y_i 's and N are independent.

Note that the moment generating function of X is

$$M_{\theta}(s) = e^{\lambda(M_Y(s)-1)} \quad (1)$$

and from the mgf $M_{\theta}(s)$ the first three cumulants can be found and they are given by

$$c_1 = \alpha\beta\lambda, c_2 = \alpha\beta^2\lambda(1+\alpha), c_3 = \alpha\lambda(\alpha+1)(\alpha+2)\beta^3 \quad (2)$$

The vector of parameters is $\theta = (\alpha, \beta, \lambda)'$.

It is not difficult to simulate from the distribution of X but the density function of X has no closed form, see Klugman *et al.* [1] (p. 143) for the series representation of this density and strictly speaking this is not a continuous model but a hybrid model where there is a probability mass assigned to the origin.

Furthermore, continuous distributions created using a mixing mechanism also leads to continuous mixture distributions without closed form density but simulated samples often can be drawn from such distributions. These distributions are commonly used in actuarial science and they are given by Klugman *et al.* [1] (pp. 62-65); see Luong [2] for other distributions with similar features used in actuarial science.

Lévy processes are also used in finance and they can be used as alternative models to the classical Brownian motion. The distributions of the increments of these processes can be more flexible than the normal distribution, they can be asymmetric and have fatter tail than the normal distribution. Consequently, they are more suitable to model log-returns of assets in finance. The following double exponential jump diffusion distribution is an illustration of an alternative distribution to the normal distribution which is the distribution of the increments of a Brownian motion.

Example 2

The double exponential jump diffusion model is a special case of a larger class of jump diffusion models where the distribution for the jumps follows an asymmetric Laplace distribution instead of the classical normal distribution as in the classical jump-diffusion model introduced by Merton [3]. This distribution has six parameters and it has been studied by Kou [4], Kou and Wang [5]. A sub-model which is the double exponential jump diffusion model with only five parameters has been found very useful for modeling log-returns of stocks, see Tsay [1] (pp. 311-319). We shall call this model, the KWT model. Exact pricing for European call option for this model is also possible with the use of some special functions. The distribution can be represented as the distribution of X with

$$X = Z + \sum_{i=1}^N Y_i$$

The Y_i 's are iid with a common distribution and mgf given respectively by

$$f_Y(y; \omega, \eta) = \frac{1}{2\eta} e^{-\frac{|x-\omega|}{\eta}}, -\infty < x < \infty, -\infty < \omega < \infty, \eta > 0, \quad (3)$$

$$M_Y(s) = \frac{e^{\omega s}}{1 - \eta^2 s^2}. \quad (4)$$

The distribution function of the double exponential distribution is

$$F_Y(y) = \frac{1}{2} e^{-\frac{x-\omega}{\eta}} \text{ for } x \leq \omega \text{ and } F_Y(y) = \frac{1}{2} + \frac{1}{2} \left(1 - e^{-\frac{x-\omega}{\eta}} \right) \text{ for } x > \omega. \quad (5)$$

Since this distribution has an explicit expression, simulated samples drawn from the double exponential distribution can be based on the inverse method.

Tsay [6] (p. 312) also gives additional properties of the distribution of Y , *i.e.*, the mean and variance are given respectively by

$$E(Y) = \omega, V(Y) = 2\eta^2 \quad (6)$$

It is assumed that the Y_i 's, Z and N are independent, N follows a Pois-

son distribution with parameter λ and Z has a normal distribution $N(\mu, \sigma^2)$.

It is easy to see that the mgf of X is given by

$$M_{\theta}(s) = e^{\mu s + \frac{1}{2}\sigma^2 s^2} e^{\lambda \left(\frac{e^{\omega s}}{1 - \eta^2 s^2} - 1 \right)}. \quad (7)$$

From $M_{\theta}(s)$, the first five cumulants can be found and they are given by

$$c_1 = \mu + \lambda\omega, \quad c_2 = \sigma^2 + \lambda(2\eta^2 + \omega^2), \quad c_3 = \lambda(6\eta^2\omega + \omega^3) \quad (8)$$

and

$$c_4 = \lambda(24\eta^4 + 12\eta^2\omega^2 + \omega^4), \quad c_5 = \lambda(120\eta^4\omega + 20\eta^2\omega^3 + \omega^5). \quad (9)$$

The vector of parameters is $\theta = (\mu, \sigma^2, \lambda, \omega, \eta)'$.

For models introduced by these examples if we use methods of moments (MM) to estimate the parameters, the MM estimators will lack of robustness properties and they might not even be efficient as for models with more than two parameters, MM estimators will depend on polynomials of degree higher or equal to three hence will be unstable in the presence of outliers. Estimators based on empirical characteristic functions procedures such as the KL procedures of Feuerverger and McDunnough [7] involves an arbitrariness of choice of points to match the empirical characteristic function with its model counterpart which motivates us in this paper to extend Cramer-von Mises estimation to a simulated version (version S). The classical Minimum Cramer Von-Mises estimators (version D) are given by the vector $\hat{\theta}$ which minimize the objective function

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n (F_n(x_i) - F_{\theta}(x_i))^2 \quad \text{or equivalently} \quad (10)$$

$$Q_n(\theta) = \int_{-\infty}^{\infty} (F_n(u) - F_{\theta}(u))^2 dF_n(u) \quad (11)$$

as given by Duchesne *et al.* [8] with $F_n(u)$ and $F_{\theta}(u)$ are respectively the sample and model distribution. The MCVN estimators are known to be robust,

$F_n(u) = \frac{1}{n} \sum_{i=1}^n I[x_i \leq u]$ is the commonly used sample distribution and $I[\cdot]$ is

the indicator function. Note that if it is easy to draw samples from $F_{\theta}(u)$, we can construct the simulated sample distribution function $F_{\theta}^S(u)$ using S observations drawn from $F_{\theta}(u)$ similarly and minimize instead the following objective function

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n (F_n(x_i) - F_{\theta}^S(x_i))^2 \quad (12)$$

to obtain estimators. We shall call these estimators simulated MCVN (SMCVN) estimators and denoted them by the vector $\hat{\theta}^S$ and we shall call this version, version S. The method is numerically relatively simple to implement using simplex direct search methods which are derivative free. Packages like R already have built in function to minimize a function using the Nelder-Mead simplex method. The SMCVM method does not require a proxy model like other simu-

lated methods such as methods based on indirect inference, see Garcia *et al.* [9], also see Smith [10]. Therefore, the method appears to be useful for actuarial science and finance where there are needs to analyze data using these types of distributions. It can also be viewed as a natural extension of the classical MCV methods proposed by Hogg and Klugman [11] (p. 83) where the asymptotic properties of the estimators have been established by Duchesne *et al.* [8]. Like Simulated minimum Hellinger (SMHD) method proposed by Luong and Bilo-deau [12], the new method is robust and it is even easier to implement than SMHD method as it makes use of sample distribution functions instead of density estimates. Furthermore, it can handle models like the compound Poisson model which displays a probability mass at the origin where SMHD method might not be suitable but comparing to SMHD estimators, the SCVM estimators might not be as efficient as the SMHD estimators for continuous models.

The paper is organized as follows. Following the approach in section 3 by Pakes and Pollard [13] (pp. 1037-1043) who make use of the Euclidean space and Euclidean norm to establish asymptotic properties of estimators, the Hilbert space l^2 is used in this paper with a natural norm extending respectively the Euclidean space and the commonly used Euclidean norm. Asymptotic properties for both the CVM estimators and the SCVM estimators can be established using a unified approach by considering minimizing the norm of a random function to obtain estimators and they are given in section 2. This approach also facilitates the use of the available results of their Theorems given in section 3 by Pakes and Pollard [13] as most of the results of their Theorems continue to hold in l^2 . The SMCVM estimators are shown to be consistent and have an asymptotic normal distribution. Their asymptotic covariance can be estimated using the influence function approach which were used by Duchesne *et al.* [8]. An estimate for the covariance matrix is also given in section 2 and by having such an estimate, it will make hypotheses testing for parameters easier to handle. Section 3 displays results of a limited simulation study using the compound Poisson gamma model and the double exponential jump distribution where we compare the SMCVM estimators with methods of moment (MM) estimators. For both models, it appears that the SCVM estimators are much more efficient than MM estimators using the overall relative efficiency criterion.

2. Asymptotic Properties of the SCVM Estimators

2.1. The Space l^2 and Its Norm

We can make use elegant results of Theorem 3.1 and 3.2 in section 3 of the paper by Pakes and Pollard [13] (pp. 1037-1043) to investigate asymptotic properties of CVM estimators and simulated CVM estimators (SCVM). For asymptotic results of estimators using simulations in their seminal works, Pakes and Pollard [13] consider estimators obtained by minimizing the Euclidean norm of a vector of random functions. The vector of random functions in their set up belong to the Euclidean space. If we used their results to investigate CVM estimation it is

more convenient to consider the Hilbert space l^2 with infinite dimension which generalize the Euclidean space and the following norm $\|\cdot\|$ defined below which generalizes the Euclidean norm.

For an element $x_l = (x_{l,1}, x_{l,2}, \dots)'$ which belongs to l^2 , define $\|x_l\| = \left(\sum_{i=1}^{\infty} x_{l,i}^2\right)^{\frac{1}{2}}$ assumed to be finite. Clearly, $\|\cdot\|$ is a norm for l^2 and it generalizes naturally the Euclidean norm. Also, a vector $u = (u_1, \dots, u_p)'$ is of finite dimension p , hence belongs to the Euclidean space then it belongs to l^2 with $u = (u_1, \dots, u_p, 0, 0, \dots)'$. The space l^2 and the norm $\|\cdot\|$ have been studied in functional analysis or real analysis, see Davidson and Donsig [14] (pp. 137-141) for example.

For a matrix $A = (a_{ij}), i = 1, 2, \dots, j = 1, 2, \dots$ in l^2 , define $\|A\| = \left(\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij}^2\right)^{\frac{1}{2}}$. With the space l^2 , most of the results of their Theorems in section 3 are valid and only some minor changes are needed.

For estimation, we assume that we have a random sample which consist of n iid observations X_1, \dots, X_n from a continuous parametric family with distribution $F_{\theta}(u)$. We also assumed that F_{θ} has no closed form expression but simulated samples can be drawn from $F_{\theta}(u)$. The commonly used sample distribution function is denoted by $F_n(u)$. The vector of parameters is denoted by

$$\theta = (\theta_1, \dots, \theta_m)'$$

Define the following vectors of random functions

$$G_n(\theta) = \left(\frac{F_n(x_1) - F_{\theta}(x_1)}{\sqrt{n}}, \dots, \frac{F_n(x_n) - F_{\theta}(x_n)}{\sqrt{n}}, 0, 0, \dots \right)' \tag{13}$$

for version D and it is easy to see that

$$\|G_n(\theta)\|^2 = \frac{1}{n} \sum_{i=1}^n (F_n(x_i) - F_{\theta}(x_i))^2 \tag{14}$$

Equivalently,

$$\|G_n(\theta)\|^2 = \int_{-\infty}^{\infty} (F_n(u) - F_{\theta}(u))^2 dF_n, \tag{15}$$

if $F_{\theta}(u)$ has support on the real line and

$\|G_n(\theta)\|^2 = \int_0^{\infty} (F_n(u) - F_{\theta}(u))^2 dF_n$, if $F_{\theta}(u)$ is the distribution of a non-negative random variable. Using the set up given by section 3 in Pollard and Pakes [13], the classical MCVm estimators can be viewed as the vector of values which minimize $\|G_n(\theta)\|^2$ or $\|G_n(\theta)\|$ as defined by expression (15).

For the simulated version of MCVm estimation, i.e., version S, define

$$G_n(\theta) = \left(\frac{F_n(x_1) - F_{\theta}^S(x_1)}{\sqrt{n}}, \dots, \frac{F_n(x_n) - F_{\theta}^S(x_n)}{\sqrt{n}}, 0, 0, \dots \right)' \tag{16}$$

with $F_{\theta}^S(u)$ being the sample distribution function based on the simulated sample of observations of size S drawn out $F_{\theta}(u)$. Then the SCVM estimators

given by the vector $\hat{\theta}^S$ is obtained by minimizing

$$\|G_n(\theta)\|^2 = \frac{1}{n} \sum_{i=1}^n (F_n(x_i) - F_{\theta}^S(x_i))^2. \tag{17}$$

Clearly, both versions of MCVm estimation can be treated in a unified way using this set up, we also have $\|G(\theta)\| < \infty$ in probability. For both versions, let

$$G(\theta) = \left(\frac{F_{\theta_0}(x_1) - F_{\theta}(x_1)}{\sqrt{n}}, \dots, \frac{F_{\theta_0}(x_n) - F_{\theta}(x_n)}{\sqrt{n}}, 0, 0, \dots \right)',$$

$\|G(\theta)\|^2 = \int_{-\infty}^{\infty} (F_{\theta_0}(u) - F_{\theta}(u))^2 dF_n(u)$ and we have $F_n(u) \xrightarrow{p} F_{\theta_0}(u)$, $\|G_n(\theta)\| - \|G(\theta)\| = o_p(1)$, with $o_p(1)$ being an expression which converges to 0 in probability.

We shall restate Theorem 3.1 given by Pakes and Pollard [13] (p. 1038) assuming the space l^2 and its norm as defined earlier are used so that it is more suitable for CVM estimation. The condition ii) which requires $G_n(\theta_0) = o_p(1)$ in their Theorem 3.1 can be replaced by $\|G_n(\theta_0)\| = o_p(1)$ as only this condition is used in their proof. Note that the set up for their Theorem is very general, we only need to verify their conditions for estimators obtained by minimizing the objective function of the form

$$Q_n(\theta) = \|G_n(\theta)\|^2 \text{ or } \|G_n(\theta)\| \tag{18}$$

Theorem 1: Under the following conditions, the estimators given by the vector $\tilde{\theta}$ converges in probability to θ_0 , the vector of the true parameters, *i.e.*, $\tilde{\theta} \xrightarrow{p} \theta_0$.

- 1) $\|G_n(\tilde{\theta})\| \leq o_p(1) + \inf_{\theta \in \Omega} \|G_n(\theta)\|$, Ω is the parameter space assumed to be compact.
- 2) $\|G_n(\theta_0)\| = o_p(1)$.
- 3) $\sup_{\|\theta - \theta_0\| > \delta} \|G_n(\theta)\|^{-1} = O_p(1)$ for each $\delta > 0$, $O_p(1)$ is an expression bounded in probability.

Clearly for the SMCVM estimators given by the vector $\hat{\theta}^S$ which minimizes $Q_n(\theta) = \|G_n(\theta)\|^2$ will satisfy condition 1) and 2) of Theorem 1 as $\|G_n(\theta)\|^2 \xrightarrow{p} 0$ only at $\theta = \theta_0$ if the parametric family is well parameterized which is the case in general. Note that the integrand of the integral defined by expression (10) is nonnegative and smaller or equal to one. Therefore, in probability,

$$0 < Q_n(\theta) \leq 1 \text{ for } \|\theta - \theta_0\| > \delta, \delta > 0$$

The condition 3) is satisfied in general which implies consistency for the SCVM estimators, we then have $\hat{\theta}^S \xrightarrow{p} \theta_0$. Note that since $Q_n(\theta)$ is always bounded it is not surprising that it generates robust estimators. For more on robustness in the sense of bounded influence functions for the SMCVM estimators see section (2.2.2). Also, observe that $\hat{\theta}^S$ remains consistent even the parame-

tric models are only hybrid, *i.e.*, with some discontinuity points such as in the case of the compound Poisson models. Now we turn our attention to the question of asymptotic normality for $\hat{\theta}^S$ and discuss informally the arguments used to establish asymptotic normality for $\hat{\theta}^S$ first and the formal arguments will follow subsequently from the proofs of Theorem 3.3 by Pakes and Pollard [13] (pp. 1040-1043). A version of their Theorem 3.3 is restated as Theorem 2 below.

Since $Q_n(\theta) = (\|G_n(\theta)\|)^2$ is not differentiable, the traditional Taylor expansion argument cannot be used to establish asymptotic normality of estimators obtained by minimizing $(\|G_n(\theta)\|)^2$. Here, we assume $G(\theta)$ is differentiable with derivative matrix $\Gamma(\theta)$, it means Fréchet differentiable with respect to the norm $\|\cdot\|$ for l^2 ; see Luenberger [15] for the notion of Fréchet differentiability and see chapter 3 of the book by Luenberger [15] for the notion of Hilbert space.

If the property of differentiability holds then we can define the random function $Q_n^a(\theta)$ to approximate $Q_n(\theta)$ with

$$Q_n^a(\theta) = (\|L_n(\theta)\|)^2, L_n(\theta) = G_n(\theta_0) + \Gamma(\theta_0)(\theta - \theta_0) \tag{19}$$

Let $\hat{\theta}^S$ and θ^* be the vectors which minimize $Q_n(\theta)$ and $Q_n^a(\theta)$ respectively. The ideas behind the proofs for asymptotic normality of Theorem (3.3) of Pakes and Pollard are if the approximation of the original objective function $Q_n(\theta)$ which is not differentiable by a differentiable one namely $Q_n^a(\theta)$ is of the right order then the vector $\hat{\theta}^S$ which minimizes $Q_n(\theta)$ and θ^* , the vector which minimizes $Q_n^a(\theta)$ are asymptotically equivalent, *i.e.*, we have:

- 1) $\sqrt{n}(\hat{\theta}^S - \theta_0) = \sqrt{n}(\theta^* - \theta_0) + o_p(1)$ or using equality in distribution, $\sqrt{n}(\hat{\theta}^S - \theta_0) \stackrel{d}{=} \sqrt{n}(\theta^* - \theta_0)$ and it is easy to see that θ^* can be expressed explicitly as $\theta^* = -(\Gamma'\Gamma)^{-1}\Gamma'G_n(\theta_0)$, $\Gamma = \Gamma(\theta_0)$ since $L_n(\theta)$ is an affine transformation.
- 2) $Q_n(\hat{\theta}^S) = Q_n^a(\theta^*) + o_p(n^{-1})$, $o_p(n^{-1})$ is an expression converging to 0 in probability at a faster rate than n^{-1} .

Note that the matrix $\Gamma(\theta_0)$ is of rank m with m columns but infinite number of rows given by

$$\Gamma = \Gamma(\theta_0) = \frac{1}{\sqrt{n}}(b_{ij}) \text{ with } b_{ij} = -\frac{\partial F_{\theta_0}(x_i)}{\partial \theta_j}, i = 1, \dots, n, j = 1, \dots, m$$

and

$$b_{ij} = 0, i = n + 1, \dots, j = 1, \dots, m$$

An estimate of this matrix $\Gamma = \Gamma(\theta_0)$ is $\hat{\Gamma}_n$ and is defined by expression (33) in section (2.2), consequently we can estimate $\frac{\partial F_{\theta_0}(x_i)}{\partial \theta_j}$ by its estimate $-\hat{b}_{ij}, i = 1, \dots, n, j = 1, \dots, m$ using the corresponding elements $\hat{\Gamma}_n(i, j)$ extracted from $\hat{\Gamma}_n$,

$$-\widehat{b}_{ij} = -\sqrt{n}\widehat{\Gamma}_n(i, j), i = 1, \dots, n, j = 1, \dots, m \quad (20)$$

Under these conditions, it suffices to work with $\boldsymbol{\theta}^*$ and $Q_n^a(\boldsymbol{\theta}^*)$ to derive asymptotic distribution for of $\widehat{\boldsymbol{\theta}}^S$. A regularity condition for the approximation is of the right order given by their Theorem 3.3 which is the most difficult to check is given as

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta_n} \frac{\|G_n(\boldsymbol{\theta}) - G(\boldsymbol{\theta}) - G_n(\boldsymbol{\theta}_0)\|}{n^{\frac{1}{2}} + \|G_n(\boldsymbol{\theta})\| + \|G_n(\boldsymbol{\theta}_0)\|} = o_p(1)$$

by Pakes and Pollard [13] (p. 1040).

A slightly more stringent condition which obviously implies the above regularity condition is

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta_n} \sqrt{n} \|G_n(\boldsymbol{\theta}) - G(\boldsymbol{\theta}) - G_n(\boldsymbol{\theta}_0)\| = o_p(1). \quad (21)$$

For simulated methods for this condition to hold, in general independent samples for each $\boldsymbol{\theta}$ cannot be used, see Pakes and Pollard [13] (p. 1048). Otherwise, only consistency can be guaranteed for estimators using version S, see section 2.2.2 for the same seed issue. For version S, the simulated samples are assumed to have size $U = \tau n$ and the same seed is used across different values of $\boldsymbol{\theta}$ to draw samples of size U . These two assumptions are quite standard for simulated methods of inferences, see section 9.6 for method of simulated moments (MSM) given by Davidson and McKinnon [16] (p. 384), also see Smith [10] (p. S66) for this assumption for his simulated quasi-likelihood estimators. For numerical optimization to find the minimum of $Q_n(\boldsymbol{\theta})$, we rely on direct search simplex methods which are derivative free. Chong and Zak [17] (pp. 273-278) provides a good overview of derivative free simplex algorithm.

2.2. Asymptotic Normality

In this section, we shall state Theorem 2 which is essentially Theorem (3.3) given by Pakes and Pollard [13] and comment on the conditions needed to verify asymptotic normality for the MCVm estimators for version D and S.

Theorem 2

Let $\tilde{\boldsymbol{\theta}}$ be a vector of consistent estimators for $\boldsymbol{\theta}_0$, the unique vector which satisfies $G(\boldsymbol{\theta}_0) = 0$.

Under the following conditions:

- 1) The parameter space Ω is compact.
- 2) $\|G_n(\tilde{\boldsymbol{\theta}})\| \leq o_p\left(n^{\frac{1}{2}}\right) + \inf_{\boldsymbol{\theta} \in \Omega} \|G_n(\boldsymbol{\theta})\|$
- 3) $G(\cdot)$ is differentiable at $\boldsymbol{\theta}_0$ with a derivative matrix $\Gamma = \Gamma(\boldsymbol{\theta}_0)$ of full rank
- 4) $\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta_n} \sqrt{n} \|G_n(\boldsymbol{\theta}) - G(\boldsymbol{\theta}) - G_n(\boldsymbol{\theta}_0)\| = o_p(1)$ for every sequence $\{\delta_n\}$ of positive numbers which converge to zero.
- 5) $\sqrt{n} \|G_n(\boldsymbol{\theta}_0)\| = O_p(1)$, $O_p(1)$ is an expression bounded in probability.

6) θ_0 is an interior point of the parameter space Ω , assumed to be compact.

Then, we have the following representation which will give the asymptotic distribution of $\tilde{\theta}$ in Corollary 1, *i.e.*,

$$\sqrt{n}(\tilde{\theta} - \theta_0) = -(\Gamma\Gamma)^{-1} \sqrt{n}\Gamma'G_n(\theta_0) + o_p(1), \tag{22}$$

or equivalently, using equality in distribution,

$$\sqrt{n}(\tilde{\theta} - \theta_0) \stackrel{d}{=} -(\Gamma\Gamma)^{-1} \sqrt{n}\Gamma'G_n(\theta_0). \tag{23}$$

The proofs of these results follow from the results used to prove Theorem 3.3 given by Pakes and Pollard [13]. For expression (22) or expression (23) to hold only condition 5) of Theorem 2 is needed and used in their proofs of Theorem 3.3 and there is no need to assume that $\sqrt{n}G_n(\theta_0)$ has an asymptotic distribution. Clearly MCVM estimators or SCVM estimators are obtained by minimizing $\|G_n(\theta)\|$ hence they will satisfy the condition 2) of Theorem 2 with $\|G_n(\theta)\|$ as defined by expression (14) or expression (17) depending it is version D or version S being considered.

Therefore, for version D,

$$\sqrt{n}(\tilde{\theta} - \theta_0) \stackrel{d}{=} -(\Gamma\Gamma)^{-1} \sqrt{n}\Gamma'G_n(\theta_0), \quad G_n(\theta) \text{ as defined by expression (13)}$$

And for version S,

$$\sqrt{n}(\hat{\theta}^S - \theta_0) \stackrel{d}{=} -(\Gamma\Gamma)^{-1} \sqrt{n}\Gamma'G_n(\theta_0), \quad G_n(\theta) \text{ as defined by expression (16)}.$$

From the result of the Theorem, it is easy to see that we can obtain the main result of the following corollary which gives the asymptotic covariance matrix of the estimators.

Corollary 1.

Let $Y_n = \sqrt{n}\Gamma'G_n(\theta_0)$, if $Y_n \xrightarrow{L} N(0, V)$ and $(\Gamma\Gamma) \xrightarrow{p} A$, A is full rank and symmetric then $\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{L} N(0, D)$ with

$$D = (A)^{-1} V (A)^{-1} \tag{24}$$

The matrices D and V depend on θ_0 , and we adopt the notations $D = D(\theta_0), V = V(\theta_0)$.

These results are proved by Pakes and Pollard [13], see the proofs of their Theorem (3.3). We just need to verify these conditions are met for SMCVM estimation. Before verifying these conditions for both versions of MCVM estimation, the following assumptions are needed to verify the condition 4 of Theorem 2 which is the most difficult condition to verify. We need to define the following sequence of functions, $\{g_n(\theta)\}$ as it will be used later,

$$g_n(\theta) = n\|G_n(\theta) - G(\theta) - G_n(\theta_0)\|^2, n = 1, 2, \dots$$

Assumption 1

1) As $n \rightarrow \infty$ and $\theta \rightarrow \theta_0$, for version S of CVM estimation

$$E_x \left\{ (F_{\theta_0}^S(x) - F_{\theta_0}(x))(F_{\theta}^S(x) - F_{\theta}(x)) \right\} \rightarrow E_x \left\{ (F_{\theta_0}^S(x) - F_{\theta_0}(x))^2 \right\}, \tag{25}$$

$E_x \{ \cdot \}$ Is the conditional expectation on x of the expression inside the

bracket.

2) The sequence of functions $g_n^e(\boldsymbol{\theta})$ is differentiable with continuous partial derivatives, $g_n^e(\boldsymbol{\theta}) = E(g_n(\boldsymbol{\theta}))$, the expectation is under $\boldsymbol{\theta}_0$ and using the usual conditioning argument, it can also be expressed

$$g_n^e(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \left(nE_{|x} \left\{ \left(F_{\boldsymbol{\theta}_0}^S(x) - F_{\boldsymbol{\theta}_0}(x) \right)^2 \right\} + nE_{|x} \left\{ \left(F_{\boldsymbol{\theta}}^S(x) - F_{\boldsymbol{\theta}}(x) \right)^2 \right\} \right. \\ \left. - 2nE_{|x} \left\{ \left(F_{\boldsymbol{\theta}_0}^S(x) - F_{\boldsymbol{\theta}_0}(x) \right) \left(F_{\boldsymbol{\theta}}^S(x) - F_{\boldsymbol{\theta}}(x) \right) \right\} \right) dF_{\boldsymbol{\theta}_0}(x) \quad (26)$$

For the condition 1) of Assumption 1 to hold we cannot use independent samples for different values of $\boldsymbol{\theta}$ to draw simulated samples for version S of CVM estimation, otherwise

$E_{|x} \left\{ \left(F_{\boldsymbol{\theta}_0}^S(x) - F_{\boldsymbol{\theta}_0}(x) \right) \left(F_{\boldsymbol{\theta}}^S(x) - F_{\boldsymbol{\theta}}(x) \right) \right\} = 0$ and $g_n^e(\boldsymbol{\theta})$ cannot converge to 0 in probability. This justifies the same seed must be used to generate random samples for different values of $\boldsymbol{\theta}$.

We shall proceed to check the regularity conditions for both versions of MCVM estimation and note that $\Gamma(\boldsymbol{\theta})$ is the derivative of $G(\boldsymbol{\theta})$ in l^2 means that $\Gamma = \Gamma(\boldsymbol{\theta}_0)$ is the Fréchet derivative at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ with the property

$$\|G_n(\boldsymbol{\theta}) - G_n(\boldsymbol{\theta}_0) - \Gamma(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\| = o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)$$

As for the Euclidean space, the sufficient condition for differentiability here only requires the partial derivatives $\frac{\partial F_{\boldsymbol{\theta}}(x)}{\partial \theta_j}$ being continuous with respect to

$\boldsymbol{\theta}$. For the notion of derivative in Hilbert space, see the notion of Fréchet derivative in Luenberger [15] (pp. 171-177) which generalizes the notion of derivative of Euclidean space. The conditions (1-3) of Theorem 2 can be verified easily. The condition (4) of Theorem 2 will be met in general if Assumption 1 holds, see Appendix for details and justifications.

We proceed to find the asymptotic distribution for $\sqrt{n}\Gamma'G_n(\boldsymbol{\theta}_0)$. Using expression (22) and expression (23), we shall obtain the asymptotic covariance matrix for the MCVM estimators for both versions. For version D, the asymptotic covariance matrix has been obtained by Duchesne et al. [8] (p. 407), using the influence function approach with the statistical functional $R(F_n)$ being defined as

$$R(F_n) = -\Gamma'G_n(\boldsymbol{\theta}_0) = \int_{-\infty}^{\infty} \left(F_n(u) - F_{\boldsymbol{\theta}_0}(u) \right) \frac{\partial F_{\boldsymbol{\theta}_0}(u)}{\partial \boldsymbol{\theta}} dF_n(u), \frac{\partial F_{\boldsymbol{\theta}_0}(u)}{\partial \boldsymbol{\theta}} = \frac{\partial F_{\boldsymbol{\theta}}(u)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

and consider the vector of influence function

$$IC_1(x) = \frac{\partial R(F_{\epsilon})}{\partial \epsilon} \Big|_{\epsilon=0}, F_{\epsilon} = F + \epsilon(\delta_x - F) \quad (27)$$

δ_x is the degenerate distribution at the point x , $F = F_{\boldsymbol{\theta}_0}$, $0 \leq \epsilon \leq 1$. The influence function $IC_1(x)$ is bounded provided that $\frac{\partial F_{\boldsymbol{\theta}_0}(u)}{\partial \boldsymbol{\theta}}$ as a vector of functions of u

is bounded which implies the MCVm estimators are robust for version D. We shall assume this property of bounded influence functions holds implicitly; we shall see this also makes version S robust. Furthermore, based on standard results of robust estimation theory, the representations given by expressions (28) and (31) using influence functions are valid for the statistical functionals being considered. Now since $R(F_{\theta_0}) = 0$,

$$\sqrt{n}R(F_n) = -\sqrt{n}\Gamma'G_n(\theta_0) = \sqrt{n}(R(F_n) - R(F_{\theta_0})) = \frac{1}{\sqrt{n}} \sum_{i=1}^n IC_1(x_i) + o_p(1) \tag{28}$$

This is the influence function representation of $R(F_n) = \Gamma'G_n(\theta_0)$ for version D and we have $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{L} N(0, D_1)$ with $D_1 = (A)^{-1}V_1(A)^{-1}$ for version D, V_1 is the covariance matrix of $IC_1(x)$, $IC_1(x)$ is given by expression (2.15) in Duchesne *et al.* [8] (p. 407), with

$$IC_1(x) = \int_{-\infty}^{\infty} (\delta_x(u) - F_{\theta_0}(u)) \frac{\partial F_{\theta_0}(u)}{\partial \theta} dF_{\theta_0}(u) \tag{29}$$

$$E(IC_1(x)) = 0, \text{ since } E_{\theta_0}(\delta_x(u)) = F_{\theta_0}(u).$$

Replacing $\frac{\partial F_{\theta_0}(u)}{\partial \theta}$ by $\frac{\partial F_{\hat{\theta}}(u)}{\partial \theta}$ and $F_{\theta_0}(u)$ by $F_n(u)$ in the above expression leads to approximate the vector

$IC_1(x)$ by $\widehat{IC}_1(x)$ with its elements given by

$$\widehat{IC}_{1l}(x) = \frac{1}{n} \sum_{j=1}^n (I[x_j \geq x] - F_n(x_j)) \frac{\partial F_{\hat{\theta}}(x_j)}{\partial \theta_l}, l = 1, \dots, m$$

An estimate for the covariance matrix V_1 can be defined as

$$\widetilde{V}_1 = \frac{1}{n} \sum_{i=1}^n (\widehat{IC}_1(x_i)) (\widehat{IC}_1(x_i))' \tag{30}$$

Using \widetilde{V}_1 , an estimate for the asymptotic covariance matrix of $\hat{\theta}$ can be constructed, see expression (2.15) and expression (2.13) given by Duchesne *et al.* [8] (pp. 406-407). Clearly, the results for version D as given by Duchesne *et al.* [8] can be reobtained using this unified approach.

Note that the property of asymptotic normality continues to hold even the parametric model fails to be continuous and is only hybrid as in the compound Poisson gamma case. Using the arguments of the next paragraph to establish asymptotic normality, the same conclusion can be reached for version S. The derivation of the asymptotic covariance matrix D_2 for the SCVM estimators is similar. We shall make use of the notion of bivariate statistical functional introduced by expression (1.6) given by Reid [18] (pp. 80-81). This leads to define the bivariate statistical functional $B(F_n, F_{\theta_0}^S)$,

$$B(F_n, F_{\theta_0}^S) = -\Gamma'G_n(\theta_0) = \int_{-\infty}^{\infty} (F_n(u) - F_{\theta_0}^S(u)) \frac{\partial F_{\theta_0}(u)}{\partial \theta} dF_n(u)$$

We have a representation which is similar to the representation given by expression (28) but using both $IC_1(x)$ and $IC_2(y)$ with

$IC_2(y) = \left. \frac{\partial B(F_\epsilon, F_\tau)}{\partial \tau} \right|_{\epsilon=0, \tau=0}$, F_ϵ is as defined by expression (27) and F_τ is similarly defined with $F_\tau = F + \tau(\delta_y - F)$, δ_y is the degenerate distribution at y and $0 \leq \tau \leq 1$. Note that $IC_1(x)$ as given by expression (29) can also be reobtained using the bivariate statistical functional with

$$IC_1(y) = \left. \frac{\partial B(F_\epsilon, F_\tau)}{\partial \epsilon} \right|_{\epsilon=0, \tau=0}.$$

Based on the expression defining $B(F_n, F_{\theta_0}^S)$, we have $IC_2(y) = -IC_1(y)$ and $IC_1(x)$ is identical for version D and S. Therefore, for version S, we have the representation

$$\begin{aligned} & \sqrt{n}B(F_n, F_{\theta_0}^S) \\ &= -\sqrt{n}\Gamma'G_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n IC_1(x_i) + \frac{\sqrt{n}}{U} \sum_{i=1}^U IC_2(y_i) + o_p(1). \end{aligned} \tag{31}$$

Note that the size of the random sample drawn from the model distribution is $U = \tau n$ and the y_i 's are iid and have the same distribution as the x_i 's but the y_i 's are independent of the x_i 's as the simulated sample is independent from the original sample represented by the data. Therefore,

$$\sqrt{n}\Gamma'G_n(\theta_0) \xrightarrow{L} N(0, V), \quad V = \left(1 + \frac{1}{\tau}\right)V_1. \tag{32}$$

It is also clear that the elements of $\Gamma'\Gamma$ are given by $a_{ij} = \int_{-\infty}^{\infty} \frac{\partial F_{\theta_0}(u)}{\partial \theta_i} \frac{\partial F_{\theta_0}(u)}{\partial \theta_j} dF_n(u), i = 1, \dots, m, j = 1, \dots, m$ which converge in probability to the corresponding elements \bar{a}_{ij} of the matrix A with

$$\bar{a}_{ij} = \int_{-\infty}^{\infty} \frac{\partial F_{\theta_0}(u)}{\partial \theta_i} \frac{\partial F_{\theta_0}(u)}{\partial \theta_j} dF_{\theta_0}(u), i = 1, \dots, m, j = 1, \dots, m, \text{ i.e.} \tag{33}$$

$$A = (\bar{a}_{ij}), i = 1, \dots, m, j = 1, \dots, m.$$

2.3. An Estimate for the Covariance Matrix for SCVM Estimators

The asymptotic covariance matrix of $\hat{\theta}^S$ can be estimated if we can estimate $\Gamma = \Gamma(\theta_0)$. Using a result given by Pakes and Pollard (p. 1043), an estimate for Γ is the matrix

$$\hat{\Gamma}_n = \left[\frac{G_n(\hat{\theta}_G^S + \epsilon_n e_1) - G_n(\hat{\theta}_G^S)}{\epsilon_n}, \dots, \frac{G_n(\hat{\theta}_G^S + \epsilon_n e_m) - G_n(\hat{\theta}_G^S)}{\epsilon_n} \right] \tag{34}$$

$e_i = (0, 0, \dots, 1, 0, \dots, 0)'$ with 1 occurring at the i^{th} entry of the vector e_i and $\epsilon_n = n^{-\delta}$, $\delta \leq \frac{1}{2}$ and in general we can let $\delta = \frac{1}{2}$. Note that the columns of $\hat{\Gamma}_n$ estimate the corresponding columns of $\Gamma(\theta_0)$ with elements depend on

$$\frac{\partial F_{\theta_0}(x_i)}{\partial \theta_j}, i=1, \dots, n, j=1, \dots, m \text{ as mentioned in section (2.2).}$$

Therefore, using results of Corollary 1 we have the asymptotic for version S

$$\sqrt{n}(\hat{\theta}^S - \theta_0) \xrightarrow{L} N(0, D_2) \text{ with } D_2 = \left(1 + \frac{1}{\tau}\right) (A)^{-1} V_1 (A)^{-1}. \quad (35)$$

The factor $1 + \frac{1}{\tau}$ represents the loss of overall efficiency due to simulations and can be controlled if we let $\tau \geq 10$. This factor is identical to the one for simulated unweighted minimum chi-square method or the one for simulated quasi-likelihood method, see Pakes and Pollard [13] (p. 1049), also see Smith [10] (p. S69). It suffices to estimate V_1 then we can have an estimate for the asymptotic covariance matrix of the SCVM estimators as clearly A can be estimated by $\hat{\Gamma}'_n \hat{\Gamma}_n$.

Define $\widehat{IC}_1(x)$ with its elements given by

$$\widehat{IC}_{1l}(x) = \frac{1}{n} \sum_{j=1}^n (I[x_j \geq x] - F_n(x_j)) (-\widehat{b}_{jl}), l=1, \dots, m,$$

$$\widehat{b}_{jl} = -\frac{\partial F_{\theta_0}(x_j)}{\partial \theta_l}, j=1, \dots, n, l=1, \dots, m \text{ are as given by expression (20).}$$

An estimate for V_1 for version S can then be defined as

$$\widehat{V}_1 = \frac{1}{n} \sum_{i=1}^n (\widehat{IC}_1(x_i)) (\widehat{IC}_1(x_i))'. \quad (36)$$

Consequently, an estimate \widehat{D}_2 for D_2 can be defined as

$$\widehat{D}_2 = \left(1 + \frac{1}{\tau}\right) (\hat{\Gamma}'_n \hat{\Gamma}_n)^{-1} \widehat{V}_1 (\hat{\Gamma}'_n \hat{\Gamma}_n)^{-1} \quad (37)$$

Clearly with \widehat{D}_2 available, it will facilitate hypothesis testing for the parameters of the model.

3. Numerical Study

3.1. MM Estimation for the Compound Poisson Gamma Model

The MM method consists of matching the empirical cumulants with its model counterpart to form estimating equations and solutions will give the moment estimators. For the compound gamma model of example 1 this leads to the system of equations given by

$$c_{1n} = \bar{X} = \lambda\alpha\beta, s^2 = \lambda\alpha\beta^2(\alpha + 1), c_{3n} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3 = \lambda\alpha\beta^3(\alpha + 1)(\alpha + 2).$$

The sample mean and variance are given respectively by \bar{X} and s^2 , the moment estimators can be obtained explicitly. Note from these equations let

$$r_{3n} = \frac{c_{3n}}{s^2} = \beta(\alpha + 2) \text{ and } r_{2n} = \frac{s^2}{\bar{X}} = \beta(\alpha + 1) \text{ which implies } \frac{r_{3n}}{r_{2n}} = \frac{\alpha + 2}{\alpha + 1} \text{ and}$$

from the last equation, we can solve for α which gives $\widehat{\alpha}_M$ the MM estimator

for α with $\widehat{\alpha}_M = \frac{2r_{2n} - r_{3n}}{r_{3n} - r_{2n}}$. Since the parameter $\alpha \geq 1$, we might want to define the moment estimator as $\widetilde{\alpha}_M = \min(\widehat{\alpha}_M, 1)$. It is not difficult to obtain $\widehat{\beta}_M = \frac{r_{2n}}{\widehat{\alpha}_M + 1}$ and $\widehat{\lambda}_M = \frac{\bar{X}}{\widehat{\alpha}_M \widehat{\beta}_M}$ the corresponding MM estimators for β and λ and when we also consider the constraints imposed on β and λ , this leads to define

$$\widetilde{\beta}_M = \min(\widehat{\beta}_M, 0) \quad \text{and} \quad \widetilde{\lambda}_M = \min(\widehat{\lambda}_M, 0).$$

3.2. MM Estimation for the KWT Model

For the KWT model, there are five parameters so beside the first three empirical cumulants as defined above we also need the fourth and fifth empirical cumulants with

$$c_{4n} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4 - 3s^4, \quad c_{5n} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^5 - 10c_{3n}s^2 \quad \text{and matching}$$

$c_{1n} = c_1, c_{2n} = c_2, c_{3n} = c_3, c_{4n} = c_4, c_{5n} = c_5$ will give the moment estimators as in the previous example. It might be easier to let $\delta = \eta^2$ and from these estimating equations, it is not difficult to see that the following two equations

$$\frac{c_{3n}}{c_{4n}} = \frac{c_3}{c_4}$$

and $\frac{c_{5n}}{c_{4n}} = \frac{c_5}{c_4}$ depend only on δ and ω and can be solved numerically to

obtain the MM estimators for δ and ω which are given respectively by $\widehat{\delta}_M$ and $\widehat{\omega}_M$. Also, using the first three equations we obtain

$$\widehat{\lambda}_M = \frac{c_{3n}}{\widehat{\omega}_M^3 + 6\widehat{\delta}_M \widehat{\omega}_M}, \quad \widehat{\sigma}_M^2 = c_{2n} - \widehat{\lambda}_M (2\widehat{\delta}_M + \widehat{\omega}_M^2), \quad \widehat{\mu}_M = c_{1n} - \widehat{\lambda}_M \widehat{\omega}_M.$$

We might want to redefine these MM estimators by imposing $\widehat{\lambda}_M \geq 0, \widehat{\sigma}_M^2 \geq 0$.

In the limited simulation study, we draw $M = 100$ samples of size $n=1000$ for each sample and use $U = 10000, \tau = 10$.

For the overall asymptotic relative efficiency (ARE) for the compound gamma model we use

$$ARE = \frac{MSE(\widehat{\lambda}^s) + MSE(\widehat{\alpha}^s) + MSE(\widehat{\beta}^s)}{MSE(\widetilde{\lambda}) + MSE(\widetilde{\alpha}) + MSE(\widetilde{\beta})}, \quad \text{the mean square errors (MSE) are}$$

estimated using random samples and displayed in **Table 1**. The mean square error of an estimator $\widehat{\pi}$ for π_0 is defined as

$$MSE(\widehat{\pi}) = E(\widehat{\pi} - \pi_0)^2.$$

The range of the parameters being considered is given by

$$2 \leq \alpha \leq 10, 1 \leq \lambda \leq 10, 1 \leq \beta \leq 10.$$

We find that the SCVM method is more efficient than MM method, the order of ARE gained by using SCVM method is illustrated with results displayed in **Table 1**. We also test for various parameters outside the range and we also have

Table 1. Compound Poisson gamma model with $\beta = 10$ asymptotic overall relative efficiency between SCVM estimators and MM estimators.

$\alpha \cdot \lambda$	1.00	5.00	6.00	7.00	8.00	9.00	10.00
2.00	0.4726	0.0105	0.4970	0.6418	0.7768	0.0030	0.2702
4.00	1.1277	0.1560	0.0348	0.0929	0.0393	0.0000	0.2973
6.00	1.0468	0.0396	0.0906	0.0070	0.0449	0.0592	0.02834
8.00	0.9032	0.0196	0.0351	0.0124	0.0553	0.0068	0.0032
10.00	0.8560	0.0352	0.3730	0.0896	0.0010	0.0179	0.0180

The overall efficiency used for comparisons used is $ARE = \frac{MSE(\hat{\lambda}^s) + MSE(\hat{\alpha}^s) + MSE(\hat{\beta}^s)}{MSE(\tilde{\lambda}) + MSE(\tilde{\alpha}) + MSE(\tilde{\beta})}$.

Table 2. Model KWT ($\mu = 0.001, \sigma = 0.001, \eta = 0.02$) asymptotic overall relative efficiency between SCVM estimators and MM estimators.

$\lambda \cdot \omega$	0.005	0.006	0.007	0.008	0.009	0.010
0.002	0.00000	0.00123	0.00099	0.00069	0.00045	0.00029
0.004	0.00070	0.00041	0.00036	0.00022	0.00012	0.00010
0.006	0.00038	0.00021	0.00015	0.00007	0.00004	0.00001
0.008	0.00019	0.00016	0.00005	0.00001	0.00000	0.00000
0.010	0.00018	0.00008	0.00001	0.00000	0.00000	0.00000

The overall efficiency used for comparisons used is

$$ARE = \frac{MSE(\hat{\mu}^s) + MSE(\hat{\sigma}^s) + MSE(\hat{\lambda}^s) + MSE(\hat{\omega}^s) + MSE(\hat{\eta}^s)}{MSE(\tilde{\mu}) + MSE(\tilde{\sigma}) + MSE(\tilde{\lambda}) + MSE(\tilde{\omega}) + MSE(\tilde{\eta})}$$

similar findings.

For the KWT model we use the corresponding asymptotic relative efficiency (*ARE*) and it is defined as

$$ARE = \frac{MSE(\hat{\mu}^s) + MSE(\hat{\sigma}^s) + MSE(\hat{\lambda}^s) + MSE(\hat{\omega}^s) + MSE(\hat{\eta}^s)}{MSE(\tilde{\mu}) + MSE(\tilde{\sigma}) + MSE(\tilde{\lambda}) + MSE(\tilde{\omega}) + MSE(\tilde{\eta})}$$

The mean square errors (*MSE*) are similarly defined as in the case of the compound gamma model and again estimated using simulated samples. The ARE is a ratio with the total of mean square errors for the SCVM estimators appearing in the numerator and the total of mean square errors of MM estimators appearing in the denominator.

The key findings are illustrated using **Table 2** and again SCVM method seems to perform much better than MM method for the common range of parameters used for modeling daily returns of stocks with

$0 \leq \lambda \leq 0.010, 0.005 \leq \omega \leq 0.010$ and $0 \leq \mu \leq 0.001, 0 \leq \sigma \leq 0.008$. With the results displayed in **Table 2** which give an idea of the order of the overall efficiency gained by using SCVM method, we can see that overall SCVM method is at least 100 time better than MM method for the range of parameters being considered. Clearly, more numerical studies are needed but we do not have the computer

resources to conduct larger scale of study being in a small department equipped with only laptop personal computers. Despite the limited nature of the study it does point to better efficiency when using SCVM methods for models having at least three parameters, in general.

4. Conclusion

It appears that SCVM method has the potential to generate more efficient estimators than MM method especially for models with more than two parameters. Like SMHD method, it is also robust and easier to implement than SMHD method as it is based on sample distribution function instead of density estimates. It can handle continuous models with a few discontinuity points with probability masses attached to them where the SMHD method might not be suitable but it might be less efficient than SMHD method for continuous model, in general.

Acknowledgements

The helpful comments of an anonymous referee and the support of the staffs of OJS which lead to an improvement of the presentation of the paper are gratefully acknowledged.

References

- [1] Klugman, S.A., Panjer, H.H. and Willmott, G.E. (2012) *Loss Models: From Data to Decisions*. New York, Wiley.
- [2] 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>
- [3] Merton, R.C. (1976) Option Pricing When the Underlying Stock Returns Are Discontinuous. *Journal of Financial Economics*, **3**, 125-144. [https://doi.org/10.1016/0304-405X\(76\)90022-2](https://doi.org/10.1016/0304-405X(76)90022-2)
- [4] Kou, S.G. (2002) A Jump Diffusion Model for Option Pricing. *Management Science*, **48**, 1086-1101. <https://doi.org/10.1287/mnsc.48.8.1086.166>
- [5] Kou, S.G. and Wang, H. (2004) Option Pricing under a Double Exponential Jump Diffusion Model. *Management Science*, **50**, 1178-1192. <https://doi.org/10.1287/mnsc.1030.0163>
- [6] Tsay, R.S. (2016) *The Analysis of Financial Time Series*. 2nd Edition, Wiley, New York.
- [7] 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.
- [8] Duchesne, T., Rioux, J. and Luong, A. (1997) Minimum Cramer-Von Mises Methods for Complete and Grouped Data. *Communications in Statistics, Theory and Methods*, **26**, 401-420. <https://doi.org/10.1080/03610929708831923>
- [9] Garcia, R., Renault, E. and Veredas, D. (2004) Estimation of Stable Distribution by Indirect Inference. *Journal of Econometrics*, **161**, 325-337. <https://doi.org/10.1016/j.jeconom.2010.12.007>
- [10] 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>

- [11] Hogg, R.V. and Klugman, S.A. (1984) Loss Distributions. Wiley, New York.
<https://doi.org/10.1002/9780470316634>
- [12] Luong, A. and Bilodeau, C. (2017) Simulated Minimum Hellinger Distance for Some Continuous Financial and Actuarial Models. *Open Journal of Statistics*, **7**, 743-759. <https://doi.org/10.4236/ojs.2017.74052>
- [13] Pakes, A. and Pollard, D. (1989) Simulation Asymptotic of Optimization Estimators. *Econometrica*, **57**, 1027-1057. <https://doi.org/10.2307/1913622>
- [14] Davidson, K.R. and Donsig, A.P. (2009) Real Analysis and Applications: Theory and Practice. Springer, New York.
- [15] Luenberger, D.G. (1968) Optimization by Vector Space Methods. Wiley, New York.
- [16] Davidson, R. and McKinnon, J.G. (2004) Econometric Theory and Methods. Oxford.
- [17] Chong, E.K.P. and Zak, S.H. (2013) An Introduction to Optimization. 4th Edition, Wiley, New York.
- [18] Reid, N. (1981) Influence Functions for Censored Data. *Annals of Statistics*, **9**, 78-92. <https://doi.org/10.1214/aos/1176345334>
- [19] Rudin, W. (1976) Principles of Mathematical Analysis. McGraw-Hill, New York.
- [20] Newey, W.K. and McFadden, D. (1994) Large Sample Estimation and Hypothesis Testing. In: Engle, R.F. and McFadden, D., Eds., *Handbook of Econometrics*, Vol. 4, North Holland, Amsterdam.

Appendix

In this technical appendix, we shall prove that with the conditions of Assumption 1, the condition 4 of Theorem 2 will hold, *i.e.*,

$$\sup_{\|\theta - \theta_0\| \leq \delta_n} \sqrt{n} \|G_n(\theta) - G(\theta) - G_n(\theta_0)\| = o_p(1), \quad \text{i.e.},$$

$$\sqrt{n} \|G_n(\theta) - G(\theta) - G_n(\theta_0)\| \xrightarrow{p} 0 \quad \text{uniformly as } \theta \rightarrow \theta_0 \text{ and } n \rightarrow \infty$$

Now define the sequence of functions $g_n(\theta) = n \|G_n(\theta) - G(\theta) - G_n(\theta_0)\|^2$, it suffices to show

$$g_n(\theta) \xrightarrow{p} 0 \quad \text{uniformly as } \theta \rightarrow \theta_0 \text{ and } n \rightarrow \infty.$$

Using Markov's type inequality, for any $\epsilon > 0$, we have the following inequality

$$P(g_n(\theta) \geq \epsilon) \leq \frac{g_n^e(\theta)}{\epsilon} \quad \text{with } g_n^e(\theta) = E(g_n(\theta)) \quad \text{as given by expression (26).}$$

Consequently, it suffices to have $g_n^e(\theta) \rightarrow 0$ uniformly as $\theta \rightarrow \theta_0$ and $n \rightarrow \infty$. Clearly under Assumption 1 we have $g_n^e(\theta) \rightarrow 0$ pointwise but we need to strengthen it to uniform convergence for $\{g_n^e(\theta)\}$. Therefore, it suffices to have equicontinuity for the sequence $\{g_n^e(\theta)\}$ as the domain of the sequence of functions is compact, see Rudin [19] (1974, p. 168). A sufficient condition for this property is the Lipschitz property which is related to the property of differentiability of the sequence of functions, see Davidson and Donsig [14] (2009, p. 88). Since the parameter space is compact and if the sequence $\{g_n^e(\theta)\}$ is differentiable hence Lipschitz then with Assumption 1, these properties implies equicontinuity for the sequences of functions $\{g_n^e(\theta)\}$.

For the notion of stochastic equicontinuity a stochastic version of equicontinuity, see Newey and McFadden [20] (1994, pp. 2136-2138) equicontinuity which extend the notion of equicontinuity of deterministic functions of real analysis and section 7 in Newey and McFadden [20] (1994, pp. 2184-2193) on asymptotic normality with non-smooth objective function.