Paper accepted for publication in Metrika. doi: 10.1007/s00184-020-00800-3 manuscript No.
(will be inserted by the editor)

Convergence and inference for mixed Poisson random sums

Gabriela Oliveira [1] · **Wagner Barreto-Souza** [2 1] ·
Roger W.C. Silva [1]

Abstract We study the limit distribution of partial sums with a random number of terms following a class of mixed Poisson distributions. The resulting weak limit is a mixture between a normal distribution and an exponential family, which we call by normal exponential family (NEF) laws. A new stability concept is introduced and a relationship between α-stable distributions and NEF laws is established. We propose the estimation of the NEF model parameters through the method of moments and also by the maximum likelihood method via an Expectation-Maximization algorithm. Monte Carlo simulation studies are addressed to check the performance of the proposed estimators, and an empirical illustration of the financial market is presented.

Keywords EM-algorithm · Mixed Poisson distribution · Stability · Weak convergence

Mathematics Subject Classification (2010) 60F05 · 62E20 · 62Fxx

1 Introduction

One of the most important and beautiful theorems in probability theory is the Central Limit Theorem, which lays down the convergence in distribution of the partial sum (properly normalized) of *i.i.d.* random variables with finite second moment to a normal distribution. This can be seen as a characterization of the normal distribution as the weak limit of such sums. A natural variant of this problem is placed when the number of terms in the sum is random. For instance, counting processes are of fundamental importance in the theory of probability and statistics. A comprehensive account for this topic is given in Gnedenko and Korolev (1996). One of the earliest counting models is the compound Poisson process $\{C_t\}_{t\geq 0}$ defined as

$$C_t = \sum_{n=1}^{N_t} X_n, \quad t \geq 0, \tag{1}$$

where $\{N_t\}_{t\geq 0}$ is a Poisson process with rate λt, $\lambda > 0$, and $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of *i.i.d.* random variables independent of $\{N_t\}_{t\geq 0}$. Applications of the random summation (1) include risk theory, biology, queuing theory and finance; for instance, see Embrechts et al. (2003), Paulsen (2008) and Puig and Barquinero (2010). For fixed t, it can be shown that the random summation given in (1), when properly normalized, converges weakly to the standard normal distribution as $\lambda \to \infty$.

G. Oliveira
[1] Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
E-mail: gabriela.oliveira.mat@gmail.com

W. Barreto-Souza
[2] Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[1] Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
Corresponding author. E-mail: wagner.barretosouza@kaust.edu.sa. Orcid: 0000-0003-0831-7881

R.W.C. Silva
[1] Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
E-mail: rogerwcs@est.ufmg.br. Orcid: 0000-0002-6365-9211

Another important quantity is the geometric random summation defined as

$$S_p = \sum_{n=1}^{\nu_p} X_n,$$

where ν_p is a geometric random variable with probability function $P(\nu_p = k) = (1-p)^{k-1}p, \;\; k = 1, 2, \ldots,$ and $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of *i.i.d.* random variables independent of ν_p, for $p \in (0,1)$. Geometric summation has a wide range of applications such as risk theory, modeling financial asset returns, insurance mathematics and others, as discussed in Kalashnikov (1997).

In Rényi (1956) it is shown that if X_n is a positive random variable with $E(X_n) < \infty$ for all n, then pS_p converges weakly to an exponential distribution as $p \to 0$. If the $X_n's$ are symmetric with $E(X_1) = 0$ and finite second moment, then there exists a_p such that $a_p S_p$ converges weakly to a Laplace distribution when $p \to 0$. If X_n has an asymmetric distribution, it is possible to show that the geometric summation, when properly normalized, converges in distribution to the asymmetric Laplace distribution. These last two results and their proofs can be found in Kotz et al. (2001).

The purpose of the present paper is to study the random summation with mixed Poisson number of terms. For a review about mixed Poisson distributions see Karlis and Xekalaki (2005). In Gavrilenko and Korolev (2006) it is shown that, if $\{X_n\}_{n\in\mathbb{N}}$ is *i.i.d.* (with $E(X_1) = 0$ and $\mathrm{Var}(X_1) = 1$) and there exists $\delta > 0$ such that $E(|X_1|^{2+\delta}) < \infty$, then the mixed Poisson (MP) random sum converges weakly to a scale mixture of normal distributions (see West (1987) for a definition of such mixture). This last assumption is necessary since the main interest in that paper is to find a Berry-Esseen type bound for the weak convergence. The study of accuracy for the convergence of MP random sums is also considered in Korolev and Shevtsova (2012), Korolev and Dorofeeva (2017) and Shevtsova (2018). Limit theorems for random summations with a negative binomial or generalized negative binomial (which are MP distributions) number of terms, with applications to real practical situations, are addressed in Bening and Korolev (2005), Schluter and Trede (2016) and Korolev and Zeifman (2019).

Our chief goal in this work is to explore mixed Poisson random summations under different assumptions compared to those in previous works in the literature since our aims here are also different. We assume that the number of terms follows the MP class of distributions proposed in Barreto-Souza (2015) and Barreto-Souza and Simas (2016). This class contains the negative binomial and Poisson inverse-Gaussian distributions as particular cases. Further, we assume that the sequence $\{X_n\}_{n\in\mathbb{N}}$ is *i.i.d.* with non-null mean and finite second moment. We do not require more than the finite second moment, in contrast to the work in Gavrilenko and Korolev (2006). Under these conditions, we show that the weak limit of an MP random sum belongs to a class of normal variance-mean mixtures (see Barndorff-Nielsen (1997) for a definition of this type of distribution) driven by a latent exponential family. We call this new class of distributions by normal exponential family (in short NEF). In particular, this class contains the normal inverse-Gaussian (NIG) distribution introduced in Barndorff-Nielsen (1997) as a special case. Therefore, this provides a new characterization for the NIG law.

Another contribution of this paper is the introduction of a new mixed Poisson stability concept, which includes geometric stability (Kozubowski and Rachev, 1994; Mittnik and Rachev, 1991) as a particular case. We also provide a theorem establishing a relationship between our proposed MP stability and the α-stable distributions.

The statistical contribution of our paper is the inferential study of the limiting class of distributions, which is of practical interest. We propose the estimation of the NEF model parameters through the method of moments and also by the maximum likelihood method, which is performed via an Expectation-Maximization (EM) algorithm (Dempster et al., 1977).

The paper is organized in the following manner. In Section 2 we show that the mixed Poisson random sums converge weakly, under some mild conditions, to a normal variance-mean mixture. Further, we define a new concept called mixed Poisson stability, which generalizes the well-known geometric stability. Properties of the limiting class of NEF distributions are explored in Section 3. Inferential aspects of the NEF models are addressed in Section 4. In Section 5 we present Monte Carlo simulations to check the finite-sample behavior of the proposed estimators. A real data application is presented in Section 6. All the proofs of propositions and theorems are provided in the Appendix.

2 Weak convergence and stability

In this section, we derive the main probabilistic results of the paper. To do this, we first present some basic concepts about the mixed Poisson distributions considered here. Then we establish the weak convergence for mixed Poisson summations, and based on this we introduce a new stability concept.

2.1 Weak limit of mixed Poisson random sums

A mixed Poisson distribution is a generalization of the Poisson distribution which is constructed as follows.

Definition 1 Let W_ϕ be a strictly positive random variable with distribution function $M_\phi(\cdot)$, where ϕ denotes a parameter associated to M. We will later assume W_ϕ belongs to a particular exponential family of distributions. Let $N|W_\phi = w \sim \text{Poisson}(\lambda w)$, $\lambda > 0$. In this case we say that N follows a mixed Poisson distribution. Its probability function takes the form

$$P(N = n) = \int_0^\infty \frac{e^{-\lambda w}(\lambda w)^n}{n!} dM_\phi(w), \qquad n \in \{0, 1, 2, \ldots\}.$$

For instance, if W_ϕ is assumed to be gamma or inverse-Gaussian (IG) distributed, then N is negative binomial or Poisson inverse-Gaussian distributed, respectively.

We consider the class of mixed Poisson distributions introduced in Barreto-Souza and Simas (2016), which is defined by assuming that W_ϕ is a continuous positive random variable belonging to the exponential family of distributions. This family was also considered in a survival analysis context in Barreto-Souza (2015). We assume that the probability density function (pdf) of W_ϕ assumes the form

$$f_{W_\phi}(w) = \exp\{\phi[w\xi_0 - b(\xi_0)] + c(w; \phi)\}, \quad w > 0, \quad \phi > 0, \tag{2}$$

where $b(\cdot)$ is a continuous and four times differentiable function, ξ_0 is such that $b'(\xi_0) = 1$, and $c(\cdot, \cdot) : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$. In this case, $E(W) = b'(\xi_0) = 1$ and $\text{Var}(W) = \phi^{-1} b''(\xi_0)$. For more details about this class of MP distributions, please see Barreto-Souza and Simas (2016).

From now on we adopt the following notation: for any random variable X we write $\Psi_X(t)$ for its characteristic function (*ch.f.*). We write $W_\phi \sim \text{EF}(\phi)$ when W_ϕ belongs to the exponential family and $N_\lambda \sim \text{MP}(\lambda, W_\phi)$, highlighting the mixture distribution involved. Let $S_\lambda \equiv X_1 + X_2 + \cdots + X_{N_\lambda}$, where $N_\lambda \sim \text{MP}(\lambda, W_\phi)$ as before and $S_\lambda \equiv 0$ when $N_\lambda = 0$. Throughout the text $\{X_n\}_{n \in \mathbb{N}}$ will always be a sequence of *i.i.d.* random variables independent of N_λ.

Before we can state our main result we need an additional observation. In Sampson (1975) the author provides a characterization of the exponential family with a single natural parameter θ in terms of its characteristic function. In that paper, T_θ belongs to this family if there exists a σ-finite measure ν such that the pdf of T_θ with respect to ν is of the form

$$f_{T_\theta}(y) = \exp\{\theta y + B(\theta) + R(y)\}, \quad y \in \mathbb{S}, \tag{3}$$

where $\mathbb{S}$ is the support of the distribution, which does not depend on θ, $B(\cdot)$ is a two times differentiable function, and $R(\cdot)$ is a real function. The following theorem appears in Sampson (1975) and plays an important role in this paper.

Theorem 1 (Sampson, 1975) *Let $\{T_\theta, \theta \in \Theta\}$ be a family of random variables such that (3) holds and $E(T_\theta) \equiv \Upsilon(\theta)$, where Θ is the parameter space. Then, the characteristic function of T_θ is given by*

$$\Psi_{T_\theta}(t) = \exp\{D(\theta + it) - D(\theta)\}, \quad t \in \mathbb{R},$$

where $D(z)$ is the analytic extension to the complex plane of $\int \Upsilon(w)dw$.

We are ready to state the main result of this section.

Theorem 2 *Let $N_\lambda \sim MP(\lambda, W_\phi)$, $\{X_n\}_{n\in\mathbb{N}}$ a sequence of i.i.d. random variables with $E(X_1) = \mu \in \mathbb{R}$ and $0 < Var(X_1) = \sigma^2 < \infty$. There exist numbers $a_\lambda = \frac{1}{\sqrt{\lambda}}$ and $d_\lambda = \mu\left(\frac{1}{\sqrt{\lambda}} - 1\right)$ such that*

$$\widetilde{S}_\lambda = a_\lambda \sum_{i=1}^{N_\lambda} (X_i + d_\lambda) \xrightarrow[\lambda\to\infty]{d} Y,$$

where Y is a random variable with ch.f.

$$\Psi_Y(t) = \exp\left\{-\phi\left[b(\xi_0) - b\left(\xi_0 + \phi^{-1}\left(it\mu - \frac{t^2\sigma^2}{2}\right)\right)\right]\right\}, \tag{4}$$

with $b(\cdot)$ and ξ_0 as in (2).

Remark 1 It should be emphasized that a more general and stronger result can be obtained by combining the results given in Korolev (1999), Korolev and Zeifman (2016a), and Korolev and Zeifman (2016b). On the other hand, such a result does not provide a specific form for the limiting distribution. In our case, specifying the class of mixed Poisson distribution allows us to obtain an explicit form for the limit distribution. The resulting class unifies existing models in the literature in a novel way. Moreover, our proof is simpler than the one mentioned above.

A special case of Theorem 2 is obtained when the sequence of random variables has null-mean.

Corollary 1 *Let $N_\lambda \sim MP(\lambda, W_\phi)$, $\{X_n\}_{n\in\mathbb{N}}$ a sequence of i.i.d. random variables with $E(X_1) = 0$ and $Var(X_1) = 1$. Then, $\lim_{\lambda\to\infty} \Psi_{\widetilde{S}_\lambda}(t) = \Psi_{W_\phi}\left(-\frac{t^2}{2}\right)$, for all $t \in \mathbb{R}$.*

Before we move on to more theoretical results, let us present a few examples.

Example 1 Let N_λ with negative binomial distribution with parameters $\lambda > 0$ and $\phi = 1$, which we denote by $N_\lambda \sim \text{NB}(\lambda, \phi)$. In this case $b(\theta) = -\log(-\theta), \xi_0 = -1$ and $c(w; 1) = 0$. From Theorem 2, it follows that

$$\Psi_Y(t) = \frac{1}{1 - \frac{t^2\sigma^2}{2} - t\mu}, \quad t \in \mathbb{R}.$$

This is the *ch.f.* of an asymmetric Laplace distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, denoted here by $\text{AL}(\mu, \sigma^2)$. In other words, $\widetilde{S}_\lambda \xrightarrow{d} \text{AL}(\mu, \sigma^2)$ as $\lambda \to \infty$.

In Example 1, we have that the density function of Y can be expressed in terms of the parameterization given in Kotz et al. (2001), *i.e.*,

$$f_Y(y) = \frac{\sqrt{2}}{\sigma}\frac{\upsilon}{1+\upsilon^2}\begin{cases}\exp\left(-\frac{\sqrt{2}\upsilon}{\sigma}|y|\right), \text{ for } & y \geq 0,\\ \exp\left(-\frac{\sqrt{2}}{\sigma\upsilon}|y|\right), \text{ for } & y < 0,\end{cases}$$

where $\upsilon = \frac{\sqrt{2\sigma^2+\mu^2}-\mu}{\sqrt{2}\sigma}$ is the skewness parameter.

Example 2 We say a random variable Z has normal inverse-Gaussian distribution with parameters a, b, c and d, and write $X \sim \text{NIG}(p, q, c, d)$, if its *ch.f.* is given by

$$\Psi_Z(t) = \exp\left\{d\left[\sqrt{p^2 - q^2} - \sqrt{p^2 - (b + it)^2}\right] + ict\right\}, \ t \in \mathbb{R}.$$

See Barndorff-Nielsen (1997) for more details on this distribution. Now, if N_λ has Poisson inverse-Gaussian distribution with parameters λ and ϕ (which we denote by $N_\lambda \sim \text{PIG}(\lambda, \phi)$), then $b(\theta) = -\sqrt{(-2\theta)}$ and $\xi_0 = -\frac{1}{2}$. Using again Theorem 2, we get

$$\Psi_Y(t) = \exp\left\{\phi\left(1 - \sqrt{1 - \phi^{-1}\left(t^2\sigma^2 + 2t\mu\right)}\right)\right\}.$$

This is the *ch.f.* of a random variable with normal inverse-Gaussian distribution with parameters $p = \sqrt{\frac{\phi}{\sigma^2} + \frac{\mu^2}{\sigma^4}}$, $q = \frac{\mu}{\sigma^2}$, $c = 0$ and $d = \sqrt{\phi}\sigma$. Therefore, $\widetilde{S}_\lambda \xrightarrow{d} \text{NIG}\left(\sqrt{\frac{\phi}{\sigma^2} + \frac{\mu^2}{\sigma^4}}, \frac{\mu}{\sigma^2}, 0, \sqrt{\phi}\sigma\right)$ as $\lambda \to \infty$.

The above examples provide characterizations for the Laplace and NIG distributions as weak limits of properly normalized mixed Poisson random sums.

2.2 Mixed Poisson-stability

Kozubowski and Rachev (1994) introduced the notion of a geometric stable distribution with respect to some summation scheme (see also Kruglov and Korolev (1990); Mittnik and Rachev (1989)). In this section we mimic this concept and introduce the notion of a stable mixed Poisson distribution. Our aim is to characterize such a distribution in terms of its *ch.f.*. We start with the following definition.

Definition 2 A random variable Y is said to be mixed Poisson stable (MP-stable) with respect to the summation scheme, if there exist a sequence of *i.i.d.* random variables $\{X_n\}_{n\in\mathbb{N}}$, a mixed Poisson random variable N_λ independent of all X_i, and constants $a_\lambda > 0$, $d_\lambda \in \mathbb{R}$ such that

$$a_\lambda \sum_{i=1}^{N_\lambda} (X_i + d_\lambda) \stackrel{d}{\to} Y, \tag{5}$$

when $\lambda \to \infty$. In particular, if N_λ follows a negative binomial or a Poisson inverse-Gaussian distribution, we say that Y is NB-stable or PIG-stable, respectively.

One of the most important objects in the theory of stable laws is the description of domains of attraction. This object is defined as follows.

Definition 3 The distribution H belongs to the domain of attraction of a MP-stable distribution G, if (5) holds with $X_i \sim H$ and $Y \sim G$. We denote this by $H \in \mathcal{D}(G)$.

The following theorem gives a characterization of MP-stable distributions in terms of its *ch.f.*. To prove this result, we follow the strategy considered in Mittnik and Rachev (1991).

Theorem 3 *A random variable Y with distribution function G is MP-stable if and only if its ch.f. Ψ_Y is of the form*

$$\Psi_Y(t) = \exp\left\{-\phi\left[b(\xi_0) - b\left(\xi_0 + \frac{1}{\phi}\log\Psi_A(t)\right)\right]\right\}, \quad t \in \mathbb{R}, \tag{6}$$

where $\Psi_A(t)$ is the ch.f. of some α-stable random variable A.

We emphasize that Theorem 3 generalizes Proposition 1 in Mittnik and Rachev (1991). To obtain their result it is enough to take $\phi = 1$ in the next example.

Example 3 Take $N_\lambda \sim \mathrm{MP}(\lambda, W_\phi)$ with $W_\phi \sim \mathrm{Gamma}(\phi)$. In this case $N_\lambda \sim \mathrm{NB}(\lambda, \phi)$ with probability function

$$P(N_\lambda = n) = \frac{\Gamma(n+\phi)}{n!\Gamma(\phi)}\left(\frac{\lambda}{\lambda+\phi}\right)^n\left(\frac{\phi}{\lambda+\phi}\right)^\phi, \quad n = 0, 1, \cdots .$$

Also, we have $b(\theta) = -\log(-\theta)$ and $\xi_0 = -1$. Apply Theorem 3 to deduce that a random variable Y is NB-stable if and only if $\Psi_Y(t) = \left\{1 - \phi^{-1}\log\Psi_A(t)\right\}^{-\phi}$, where $\Psi_A(\cdot)$ is the *ch.f.* of some α-stable random variable A.

Example 4 Consider $N_\lambda \sim \mathrm{MP}(\lambda, W_\phi)$ with $W_\phi \sim \mathrm{IG}(\phi)$. In this case $N_\lambda \sim \mathrm{PIG}(\lambda, \phi)$ with probability function

$$P(N_\lambda = n) = \sqrt{\frac{2}{\pi}}\,[\phi(\phi+2\lambda)]^{-(n-\frac{1}{2})}\,\frac{e^\phi(\lambda\phi)^n}{n!}\mathcal{L}_{n-\frac{1}{2}}\left(\sqrt{\phi(\phi+2\lambda)}\right), \ n = 0, 1, \cdots$$

where $\mathcal{L}_\nu(z) = \frac{1}{2}\int_0^\infty u^{\nu-1}\exp\left\{-\frac{1}{2}z(u+u^{-1})\right\}du$ is the modified Bessel function of the third kind; see Abramowitz and Stegun (1965). Also, we have $b(\theta) = -\sqrt{-2\theta}$, $\xi_0 = -\frac{1}{2}$. Applying Theorem 3 we obtain that Y is PIG-stable if and only if $\Psi_Y(t) = \exp\left\{\phi\left(1 - \sqrt{1 - 2\phi^{-1}\log\Psi_A(t)}\right)\right\}$, with $\Psi_A(\cdot)$ being the *ch.f.* of some α-stable distribution. For instance, by taking the *ch.f.* of a normal distribution with mean μ and variance σ^2, that is $\Psi(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$, it follows that $\Psi_Y(t) = \exp\left\{\phi\left(1 - \sqrt{1 - \phi^{-1}\left(2i\mu t - \sigma^2 t^2\right)}\right)\right\}$, which is the *ch.f.* of the NIG $\left(\sqrt{\frac{\phi}{\sigma^2} + \frac{\mu^2}{\sigma^4}}, \frac{\mu}{\sigma^2}, 0, \sqrt{\phi\sigma^2}\right)$ distribution. In other words, the normal inverse-Gaussian distribution is PIG-stable.

3 Properties of the limiting distribution

In this section we obtain statistical properties of the limiting class of distributions with *ch.f.* (4) arising from Theorem 2. The main result here is the stochastic representation of these distributions as a normal mean-variance mixture (see Barndorff-Nielsen et al. (1982)) with a latent effect belonging to an exponential family. We emphasize that this class of normal exponential family (NEF) mixture distributions is new in the literature.

Proposition 1 *Let Y be a random variable with ch.f. (4). Then Y satisfies the following stochastic representation:*

$$Y \stackrel{d}{=} \mu W_\phi + \sigma\sqrt{W_\phi} Z,$$

where $W_\phi \sim EF(\phi)$ and $Z \sim N(0,1)$ are independent and '$\stackrel{d}{=}$' stands for equality in distribution.

Since we rely on an Expectation-Maximization algorithm to estimate the parameters of the class of normal mean-variance mixture distributions, the stochastic representation given in the previous proposition plays an important role in this paper. Furthermore, this representation enables us to find explicit forms for the corresponding density function as stated in the following proposition (whose proof follows directly and therefore it is omitted) and examples.

Proposition 2 *Let $Y = \mu W_\phi + \sigma\sqrt{W_\phi} Z$ with $Z \sim N(0,1)$ and W_ϕ independent, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ and $\phi > 0$. Assume that the function $c(\cdot,\cdot)$ in (2) can be expressed as $c(w;\phi) = d(\phi) + \phi g(w) + h(w)$. Then, the density function of Y is given by*

$$f_Y(y) = \frac{e^{\frac{y\mu}{\sigma^2} - \phi b(\xi_0) + d(\phi)}}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{\phi g(w) + h(w)} w^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\left(\frac{\mu^2}{\sigma^2} - 2\phi\xi_0\right) w + \frac{y^2}{\sigma^2}\frac{1}{w}\right]\right\} dw, \qquad (7)$$

for $y \in \mathbb{R}$.

Example 5 Let $Y = \mu W_\phi + \sigma\sqrt{W_\phi} Z$ with $W_\phi \sim \text{Gamma}(\phi)$ independent of $Z \sim N(0,1)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ and $\phi > 0$. In this case, we have $d(\phi) = \phi \log\phi - \log\Gamma(\phi)$, $g(w) = \log w$, and $h(w) = -\log w$. Using (7), we obtain

$$f_Y(y) = \sqrt{\frac{2}{\pi\sigma^2}} \frac{\phi^\phi}{\Gamma(\phi)} e^{y\mu/\sigma^2} \mathcal{K}_{\phi-\frac{1}{2}}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2} + 2\phi\right)\frac{y^2}{\sigma^2}}\right)\left(\frac{y^2}{\mu^2 + 2\phi\sigma^2}\right)^{\frac{\phi}{2}-\frac{1}{4}}, \; y \in \mathbb{R}, \qquad (8)$$

where $\mathcal{K}_\cdot(\cdot)$ is the modified Bessel function of the third kind. This function satisfies the property $\mathcal{K}_{\frac{1}{2}}(z) = \left(\frac{\pi}{2z}\right)^{\frac{1}{2}} e^{-z}$, for $z \in \mathbb{R}$. Using this fact and replacing $\phi = 1$ in Equation (8), we obtain the probability density function of the asymmetric Laplace distribution.

Example 6 If we assume $W_\phi \sim \text{IG}(\phi)$, then $d(\phi) = \frac{1}{2}\log\phi$, $g(w) = -(2w)^{-1}$ and $h(w) = -\frac{1}{2}\log(2\pi w^3)$. Hence, it follows from Equation (7) that

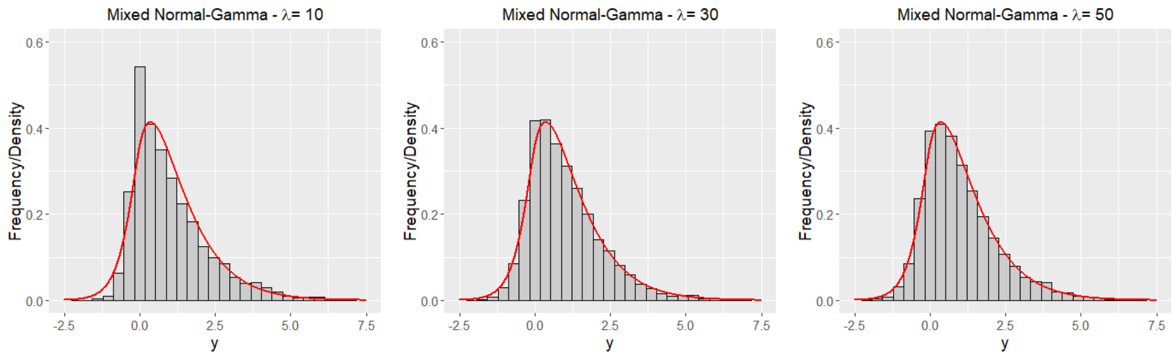
$$f_Y(y) = \frac{1}{\pi}\sqrt{\frac{\phi}{\sigma^2}} \exp\left(\frac{y\mu}{\sigma^2} + \phi\right) \mathcal{K}_{-1}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2} + \phi\right)\left(\frac{y^2}{\sigma^2} + \phi\right)}\right)\left(\frac{\mu^2 + \sigma^2\phi}{y^2 + \sigma^2\phi}\right)^{\frac{1}{2}}, \quad y \in \mathbb{R}. \qquad (9)$$

The density function (9) corresponds to a NIG distribution.

Example 7 Consider W_ϕ following a generalized hyperbolic secant distribution with dispersion parameter $\phi > 0$; for more details on this distribution see Barreto-Souza and Simas (2016). In this case, we cannot use Proposition (2) since the function $c(\cdot,\cdot)$ cannot be decomposed as required. Anyway, we can express the density function of Y by

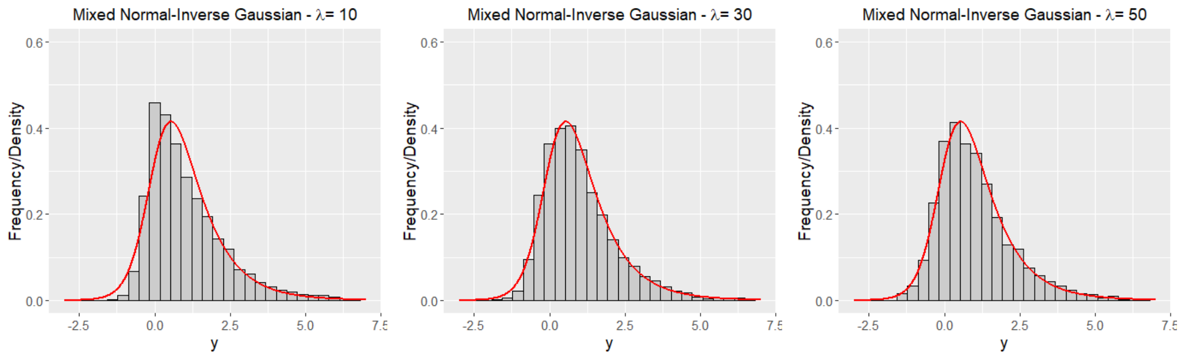
$$f_Y(y) = \frac{2^{\frac{\phi-5}{2}}}{\sqrt{\pi^3\sigma^2}} \frac{\phi}{\Gamma(\phi)} \exp(\mu y/\sigma^2) E\left(\Gamma\left(|\phi + iV|^2/4\right)\right), \; y \in \mathbb{R},$$

where $V \sim \text{GIG}\left(\frac{3\pi\phi}{2} + \frac{\mu^2}{\sigma^2}, \frac{y^2}{\sigma^2}, \frac{1}{2}\right)$.



(a) $\lambda = 10$ (b) $\lambda = 30$ (c) $\lambda = 50$

Fig. 1 Histograms for the generated random sample from $\widetilde{S}_\lambda$ with $N_\lambda \sim \mathrm{NB}(\lambda, 2)$ for $\lambda = 10, 30, 50$ and normal gamma density curve.



(a) $\lambda = 10$ (b) $\lambda = 30$ (c) $\lambda = 50$

Fig. 2 Histograms for the generated random sample from $\widetilde{S}_\lambda$ with $N_\lambda \sim \mathrm{PIG}(\lambda, 2)$ for $\lambda = 10, 30, 50$ and normal inverse-Gaussian density curve.

We conclude this section with a numerical illustration of the weak convergence obtained in Theorem 2 through a small Monte Carlo simulation. We generate random samples (5000 replicas) from the partial sums $\widetilde{S}_\lambda$ with $\mathrm{NB}(\lambda, \phi)$ and $\mathrm{PIG}(\lambda, \phi)$ number of terms; we set $\phi = 2$ and $\lambda = 10, 30, 50$. The sequence $\{X_n\}_{n\in\mathbb{N}}$ is generated from the exponential distribution with mean equal to 1 (in this case $\mu = \sigma^2 = 1$). Figures 1 and 2 show the histograms of the generated random samples with the curve of the corresponding density functions (NG and NIG) for the negative binomial and Poisson inverse-Gaussian cases, respectively. As expected, we observe a good agreement between the histograms and the theoretical densities as λ increases, which is according to Theorem 2.

4 Inference for NEF laws

In this section, we discuss the estimation of the parameters of the limiting class of normal-exponential family laws obtained in Section 2. We consider the method of moments and maximum likelihood estimation via the Expectation-Maximization algorithm. Throughout this section, $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ denotes a random sample (*i.i.d.*) from the NEF distribution and n stands for the sample size.

4.1 Method of moments

Let $Y \sim \text{NEF}(\mu, \sigma^2, \phi)$. By using the characteristic function of the NEF distributions given in Proposition 2, we see that the first four cumulants of Y are given by

$$\begin{aligned}
\kappa_1 &\equiv E(Y) = \mu, \\
\kappa_2 &\equiv \text{Var}(Y) = \frac{\mu^2 b''(\xi_0) + \phi\sigma^2}{\phi}, \\
\kappa_3 &\equiv E\left((Y-\mu)^3\right) = \frac{\mu^3 b^{(3)}(\xi_0) + 3\phi\sigma^2\mu b''(\xi_0)}{\phi^2}, \\
\kappa_4 &\equiv E\left((Y-\mu)^4\right) = \frac{\mu^4 b^{(4)}(\xi_0) + 6\phi\sigma^2\mu^2 b^{(3)}(\xi_0) + 3\phi^2\sigma^4 b''(\xi_0)}{\phi^3},
\end{aligned} \tag{10}$$

where $b^{(3)}(\cdot)$ and $b^{(4)}(\cdot)$ are the third and forth derivatives of the function $b(\cdot)$. The skewness coefficient and the excess of kurtosis, denoted respectively by β_1 and β_2, can be obtained from the well-known relationships

$$\beta_1 = \frac{\kappa_3}{\kappa_2^{3/2}} \quad \text{and} \quad \beta_2 = \frac{\kappa_4}{\kappa_2^2} - 3. \tag{11}$$

The following examples give explicit expressions for the first four cumulants for some special cases of the NEF class of distributions.

Example 8 Consider Y following a NEF distribution with $W_\phi \sim \text{Gamma}(\phi)$. In this case Y follows a normal-gamma distribution. Also, $b(\theta) = -\log(-\theta)$ for $\theta < 0$ and $\xi_0 = -1$. By using these quantities and taking the derivatives of $b(\cdot)$ in (10) and (11) we obtain that

$$\left\{\begin{aligned}
\kappa_1 &= \mu, \\
\kappa_2 &= \frac{\mu^2 + \phi\sigma^2}{\phi}, \\
\kappa_3 &= \frac{2\mu^3 + 3\mu\phi\sigma^2}{\phi^2}, \\
\kappa_4 &= \frac{6\mu^4 + 12\mu^2\phi\sigma^2 + 3\phi^2\sigma^4}{\phi^3},
\end{aligned}\right. \qquad
\left\{\begin{aligned}
\beta_1 &= \frac{2\mu^3 + 3\mu\phi\sigma^2}{\sqrt{\phi}(\mu^2 + \phi\sigma^2)^{\frac{3}{2}}}, \\
\beta_2 &= \frac{6\mu^4 + 12\mu^2\phi\sigma^2 + 3\phi^2\sigma^4}{\phi(\mu^2 + \phi\sigma^2)^2}.
\end{aligned}\right.$$

Example 9 Now consider $W_\phi \sim \text{IG}(\phi)$. Then, $b(\theta) = -\sqrt{-2\theta}$ for $\theta < 0$ and $\xi_0 = -1/2$. We have that Y follows a NIG distribution with parameters μ, σ^2 and ϕ. Its central moments and cumulants are given by

$$\left\{\begin{aligned}
\kappa_1 &= \mu, \\
\kappa_2 &= \frac{\mu^2 + \phi\sigma^2}{\phi}, \\
\kappa_3 &= \frac{3\mu^3 + 3\mu\phi\sigma^2}{\phi^2}, \\
\kappa_4 &= \frac{15\mu^4 + 18\mu^2\phi\sigma^2 + 3\phi^2\sigma^4}{\phi^3},
\end{aligned}\right. \qquad
\left\{\begin{aligned}
\beta_1 &= \frac{3\mu^3 + 3\mu\phi\sigma^2}{\sqrt{\phi}(\mu^2 + \phi\sigma^2)^{\frac{3}{2}}}, \\
\beta_2 &= \frac{15\mu^4 + 18\mu^2\phi\sigma^2 + 3\phi^2\sigma^4}{\phi(\mu^2 + \phi\sigma^2)^2}.
\end{aligned}\right.$$

Example 10 For the case $W_\phi \sim \text{GHS}(\phi)$ (generalized hyperbolic secant), it follows that $b(\theta) = \frac{1}{2}\log\left(1 + \tan^2\theta\right)$ for $\theta \in \mathbb{R}$ and $\xi_0 = -3\pi/4$. We have that Y follows a normal-generalized hyperbolic secant distribution.

Its central moments and cumulants are given by

$$\begin{cases} \kappa_1 = \mu, \\ \kappa_2 = \dfrac{2\mu^2 + \sigma^2\phi}{\phi}, \\ \kappa_3 = \dfrac{4\mu^3 + 6\mu\sigma^2\phi}{\phi^2}, \\ \kappa_4 = \dfrac{16\mu^4 + 24\sigma^2\phi + 6\sigma^4\phi^2}{\phi^3}, \end{cases} \qquad \begin{cases} \beta_1 = \dfrac{4\mu^3 + 6\mu\sigma^2\phi}{\sqrt{\phi}(2\mu^2 + \sigma^2\phi)^{\frac{3}{2}}}, \\ \beta_2 = \dfrac{16\mu^4 + 24\sigma^2\phi + 6\sigma^4\phi^2}{\phi(2\mu^2 + \sigma^2\phi)^2}. \end{cases}$$

Let us discuss the estimation procedure. Since we have three parameters, we need three equations to estimate them. We use the three first moments $\mu_k \equiv E(Y^k)$ and its respective empirical quantities $M_k \equiv \dfrac{1}{n}\sum_{i=1}^{n} Y_i^k$ to do this job, where $k = 1, 2, 3$ and $\mu_1 \equiv \mu$. The theoretical moments can be obtained from the cumulants by using the relationships $\mu_1 = \kappa_1$, $\mu_2 = \kappa_2 + \mu_1^2$ and $\mu_3 = \kappa_3 + 3\mu_1\mu_2 - 2\mu_1^3$.

By equating theoretical moments with their empirical quantities, the method of moments (MM) estimators are obtained as the solution of the following system of non-linear equations:

$$\begin{cases} \widetilde{\mu}_1 = M_1 \\ \widetilde{\mu}_2 = M_2 \\ \widetilde{\mu}_3 = M_3 \end{cases} \Longrightarrow \begin{cases} \widetilde{\mu} = M_1 \\ \dfrac{\widetilde{\mu}^2 b''(\xi_0) + \widetilde{\phi}\widetilde{\sigma}^2}{\widetilde{\phi}} + \mu_1^2 = M_2 \\ \dfrac{\widetilde{\mu}^3 b^{(3)}(\xi_0) + 3\widetilde{\phi}\widetilde{\sigma}^2\widetilde{\mu} b''(\xi_0)}{\widetilde{\phi}^2} + 3\widetilde{\mu}_1\widetilde{\mu}_2 - 2\widetilde{\mu}_1^3 = M_3 \end{cases}$$

The solution of the above system of equations, denoted by $\widetilde{\mu}$, $\widetilde{\sigma}^2$ and $\widetilde{\phi}$, is the MM estimator and is given explicitly by

$$\begin{cases} \widetilde{\mu} = M_1, \\ \widetilde{\sigma}^2 = M_2 - M_1^2 \left(1 + \dfrac{b''(\xi_0)}{\widetilde{\phi}}\right), \end{cases}$$

where $\widetilde{\phi}$ is the admissible solution of the quadratic equation

$$\left(3M_1M_2 - 2M_1^3 - M_3\right)\widetilde{\phi}^2 + b''(\xi_0)(3M_1M_2 - 3M_1^3)\widetilde{\phi} + M_1^3(b^{(3)}(\xi_0) - 3b''(\xi_0)^2) = 0.$$

A potential problem of the MM estimators is that estimates can lie outside of the parameter space, especially under small sample sizes. When admissible MM estimates are available, they also can be used as initial guesses for the EM-algorithm, as discussed in the sequel.

4.2 Expectation-Maximization algorithm

In this section, we obtain the Expectation-Maximization (EM) algorithm to find the maximum likelihood estimators for the parameters of the model. From the stochastic representation of the NEF laws, we can use W_ϕ as the latent variable to construct such an estimation algorithm.

Consider the complete data $(Y_1, W_{\phi,1}), \cdots, (Y_n, W_{\phi,n})$, where $Y_1, \ldots, Y_n$ are observable variables with respective latent effects $W_{\phi,1}, \ldots, W_{\phi,n}$. Let $\Omega = (\mu, \sigma^2, \phi)^\top$ be the parameter vector. The complete log-likelihood function is $\ell_c(\Omega) = \sum_{i=1}^{n} \log\{P(Y_i = y_i | W_{\phi,i} = w_i) f_{W_\phi}(w_i)\}$, where $f_{W_\phi}(\cdot)$ is the density function of the exponential family given in Equation (2). From now on, we assume that the function $c(\cdot,\cdot)$ can be expressed as $c(w;\phi) = d(\phi) + \phi g(w) + h(w)$, with $d(\cdot)$ a three times differentiable function (see Barreto-Souza and Simas (2016)). The functions $d(\cdot)$, $g(\cdot)$, and $h(\cdot)$ for the gamma and inverse-Gaussian cases are given explicitly in Examples 5 and 6, respectively.

More explicitly, we obtain that the complete log-likelihood function takes the form

$$\ell_c(\Omega) \propto \sum_{i=1}^{n} \left\{ -\frac{1}{2}\log\sigma^2 - \frac{y_i^2}{2\sigma^2}{w_i}^{-1} + \frac{\mu}{\sigma^2}y_i - \frac{\mu^2}{2\sigma^2}w_i + d(\phi) + \phi\left[w_i\xi_0 - b(\xi_0) + g(w_i)\right] \right\}.$$

We now obtain the E-step and M-step of the EM algorithm with details. We denote by $\Omega^{(r)}$ the estimate of the parameter vector Ω in the rth loop of the EM-algorithm.

E-step. Here, we need to find the conditional expectation of $\ell_c(\Omega)$ given the observable random variables $Y = (Y_1, \ldots, Y_n)^\top$. We denote this conditional expectation by Q, which assumes the form

$$\begin{aligned} &Q(\Omega; \Omega^{(r)}) \equiv E(\ell_c(\Omega)|Y = y; \Omega^{(r)}) \\ &\propto \sum_{i=1}^{n} \left\{ -\frac{1}{2}\log\sigma^2 - \frac{y_i^2}{2\sigma^2}\gamma_i^{(r)} + \frac{\mu}{\sigma^2}y_i - \frac{\mu^2}{2\sigma^2}\alpha_i^{(r)} + d(\phi) + \phi\left[\xi_0\alpha_i^{(r)} - b(\xi_0) + \delta_i^{(r)}\right] \right\}, \end{aligned}$$

where $\gamma_i^{(r)} \equiv E\left(W_{\phi,i}^{-1}|Y_i = y_i; \Omega^{(r)}\right)$, $\alpha_i^{(r)} \equiv E\left(W_{\phi,i}|Y_i = y_i; \Omega^{(r)}\right)$ and $\delta_i^{(r)} \equiv E\left(g(W_{\phi,i})|Y_i = y_i; \Omega^{(r)}\right)$, for $i = 1, \ldots, n$.

In the following, we obtain the conditional expectations above for the gamma and inverse-Gaussian cases. For simplicity of notation, the index i is omitted.

Proposition 3 *Assume $W_\phi \sim$ Gamma(ϕ). Then, for $K, L \in \mathbb{Z}$, we have that*

$$E\left(W_\phi^K g(W_\phi)^L \middle| Y = y\right) = \frac{\mathcal{K}_{\phi+K-\frac{1}{2}}(\sqrt{ab})}{\mathcal{K}_{\phi-\frac{1}{2}}(\sqrt{ab})}\left(\frac{b}{a}\right)^{\frac{K}{2}} E\left(g(U)^L\right),$$

where $U \sim GIG(a, b, p)$, $a = \frac{\mu^2}{\sigma^2} + 2\phi$, $b = \frac{y^2}{\sigma^2}$, $p = \phi + K - \frac{1}{2}$, and $\mathcal{K}_\cdot(\cdot)$ is the modified Bessel function of the third kind.

Example 11 Let $Y = \mu W_\phi + \sigma\sqrt{W_\phi}Z$ with $Z \sim N(0, 1)$ and $W_\phi \sim$ Gamma(ϕ) independent of each other, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ and $\phi > 0$. Replacing $(K, L) = (1, 0)$, $(K, L) = (-1, 0)$ and $(K, L) = (0, 1)$ in the previous proposition, we get

$$\alpha \equiv E(W_\phi|Y = y) = \frac{\mathcal{K}_{\phi+\frac{1}{2}}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2} + 2\phi\right)\left(\frac{y^2}{\sigma^2}\right)}\right)}{\mathcal{K}_{\phi-\frac{1}{2}}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2} + 2\phi\right)\left(\frac{y^2}{\sigma^2}\right)}\right)}\left(\frac{y^2}{\mu^2 + 2\phi\sigma^2}\right)^{\frac{1}{2}},$$

$$\gamma \equiv E\left({W_\phi}^{-1}|Y = y\right) = \frac{\mathcal{K}_{\phi-\frac{3}{2}}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2} + 2\phi\right)\left(\frac{y^2}{\sigma^2}\right)}\right)}{\mathcal{K}_{\phi-\frac{1}{2}}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2} + 2\phi\right)\left(\frac{y^2}{\sigma^2}\right)}\right)}\left(\frac{y^2}{\mu^2 + 2\phi\sigma^2}\right)^{-\frac{1}{2}}$$

and

$$\delta = E(\log W_\phi|Y = y) = \frac{1}{2}\log\left(\frac{\mu^2 + 2\phi\sigma^2}{y^2}\right) + \frac{\mathcal{K}'_{\phi-\frac{1}{2}}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2} + 2\phi\right)\left(\frac{y^2}{\sigma^2}\right)}\right)}{\mathcal{K}_{\phi-\frac{1}{2}}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2} + 2\phi\right)\left(\frac{y^2}{\sigma^2}\right)}\right)}.$$

We now present explicit expressions for the conditional expectation for the NIG case.

Proposition 4 *Consider $W_\phi \sim IG(\phi)$. Then, for $K, L \in \mathbb{Z}$, we obtain that*

$$E\left(W_\phi^K g(W_\phi)^L \middle| Y = y\right) = \frac{\mathcal{K}_{K-1}(\sqrt{ab})}{\mathcal{K}_{-1}(\sqrt{ab})}\left(\frac{b}{a}\right)^{\frac{K}{2}} E\left(g(U)^L\right),$$

where $U \sim GIG(a, b, p)$, $a = \frac{\mu^2}{\sigma^2} + \phi$, $b = \frac{y^2}{\sigma^2} + \phi$.

Example 12 By considering the NIG case and applying Proposition 4 with $(K, L) = (1, 0)$, $(K, L) = (-1, 0)$ and $(K, L) = (0, 1)$, we obtain

$$\alpha \equiv E(W_\phi | Y = y) = \frac{\mathcal{K}_0\left(\sqrt{\left(\frac{\mu^2}{\sigma^2}+\phi\right)\left(\frac{y^2}{\sigma^2}+\phi\right)}\right)}{\mathcal{K}_{-1}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2}+\phi\right)\left(\frac{y^2}{\sigma^2}+\phi\right)}\right)}\left(\frac{y^2+\phi\sigma^2}{\mu^2+\phi\sigma^2}\right)^{\frac{1}{2}},$$

$$\gamma \equiv E\left({W_\phi}^{-1} | Y = y\right) = \frac{\mathcal{K}_{-2}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2}+\phi\right)\left(\frac{y^2}{\sigma^2}+\phi\right)}\right)}{\mathcal{K}_{-1}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2}+\phi\right)\left(\frac{y^2}{\sigma^2}+\phi\right)}\right)}\left(\frac{y^2+\phi\sigma^2}{\mu^2+\phi\sigma^2}\right)^{-\frac{1}{2}} \quad \text{and}$$

$$\delta \equiv E\left(-\frac{1}{2W_\phi} | Y = y\right) = -\frac{1}{2}\gamma.$$

M-step. This step of the EM-algorithm consists in maximizing the function $Q \equiv Q(\Omega; \Omega^{(r)})$. The score function associated to this function is

$$\begin{aligned}
\frac{\partial Q}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left\{y_i - \mu\alpha_i^{(r)}\right\}, \\
\frac{\partial Q}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}\left\{y_i^2\gamma_i^{(r)} - 2\mu y_i + \mu^2\alpha_i^{(r)}\right\}, \\
\frac{\partial Q}{\partial \phi} &= n(d'(\phi) - b(\xi_0)) + \sum_{i=1}^{n}\left\{\xi_0\alpha_i^{(r)} + \delta_i^{(r)}\right\}.
\end{aligned}$$

The estimate of Ω in the $(r+1)$th loop of the EM-algorithm is obtained as the solution of the system of equations $\partial Q(\Omega; \Omega^{(r)})/\partial\Omega = 0$. After some algebra, we get

$$\mu^{(r+1)} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n}\alpha_i^{(r)}}, \quad \sigma^{2\,(r+1)} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^2\gamma_i^{(r)} - 2\mu^{(r+1)}y_i + {\mu^{(r+1)}}^2\alpha_i^{(r)}\right) \quad \text{and}$$

$$\phi^{(r+1)} = v\left(b(\xi_0) - \frac{\xi_0}{n}\sum_{i=1}^{n}\alpha_i^{(r)} - \frac{1}{n}\sum_{i=1}^{n}\delta_i^{(r)}\right),$$

where $v(\cdot)$ is the inverse function of $d'(\cdot)$.

We now describe briefly how the EM-algorithm works. As an initial guess for $\Omega^{(0)}$ we can take the MM estimates. Update the conditional expectations with the previous EM-estimates, denoted by $\Omega^{(r)}$, as well as the Q-function. The next step is to find the maximum global point of the Q-function, say $\Omega^{(r+1)}$, which is provided in the closed-form above. Check if some convergence criterion is satisfied, for instance $||\Omega^{(r+1)} - \Omega^{(r)}||/||\Omega^{(r)}|| < \epsilon$, for some small $\epsilon > 0$. If this criterion is satisfied, the current EM-estimate is returned. Otherwise, update the previous EM-estimate by the current one and perform the above algorithm again until convergence is achieved.

The standard error of the parameter estimates can be obtained through the observed information matrix in Louis (1982), which is given by

$$I(\Omega) = E\left(-\frac{\partial^2 \ell_c(\Omega)}{\partial\Omega\partial\Omega^\top}\Big|Y\right) - E\left(\frac{\partial \ell_c(\Omega)}{\partial\Omega}\frac{\partial \ell_c(\Omega)}{\partial\Omega}^\top\Big|Y\right). \tag{12}$$

Explicit expressions for the elements of the information matrix (12) can be obtained from the authors upon request.

5 Simulation

We present a small Monte Carlo study for comparing the performance of the EM-algorithm and the method of moments for estimating the parameters of the NEF laws. We also check the estimation of the standard errors obtained from the observed information matrix via the EM-algorithm.

We consider the cases where data are generated from the normal-gamma and NIG distributions. To generate from these distributions, we use the stochastic representation $Y \stackrel{d}{=} \mu W_\phi + \sigma\sqrt{W_\phi}Z$, where $Z \sim N(0,1)$ independent of W_ϕ, which is Gamma(ϕ) or IG(ϕ) distributed, respectively. We set the true parameter vector $\Omega = (\mu, \sigma^2, \phi) = (3, 4, 2)$ and sample sizes $n = 30, 50, 100, 150, 200, 500, 1000$. We run a Monte Carlo simulation with 5000 replicas. Further, we use the MM estimates as initial guesses for the EM-algorithm and consider its convergence criterion to be the one proposed in Subsection 4.2 with $\epsilon = 10^{-4}$.

Figures 3 and 4 present boxplots of the estimates of the parameters based on the EM-algorithm and method of moments for some sample sizes under the normal gamma and NIG distributions, respectively. Overall, the bias and variance of the estimates go to 0 as the sample size increases, as expected. Let us now discuss each case in more detail.

Concerning the parameter μ, both methods yield similar results under normal gamma and NIG assumptions. On the other hand, regarding the estimation of the parameters σ^2 and ϕ, the EM-algorithm has a superior performance over the method of moments in all cases considered for both normal gamma and NIG distributions. We observe that the method of moments yields a considerable bias, even for sample sizes $n = 200, 500, 1000$, in contrast with EM-approach which produces unbiased estimates even for sample sizes $n = 50, 100$.

Another problem of the MM estimator is that it can produce estimates out of the parameter space. Under the normal gamma distribution, the percentages of negative estimates for σ^2 and/or ϕ with sample sizes $n = 30, 50, 100, 150, 200, 500, 1000$ were respectively 6.34%, 4.72%, 3.18%, 2.5%, 0.24%, 0.18% and 0.24%. The respective percentages for the NIG case were respectively 3.12%, 1.34%, 0.74%, 0.66%, 0.62%, 0.24% and 0.04%. In these cases, the Monte Carlo replicas were discarded and new values were generated. It is worth to mention that some huge outlier estimates were yielded by the method of moments (for small sample sizes). They cannot be seen from the plots due to the scale of the boxplots, which were chosen to give a clear view of the big picture.

We finish this section by presenting the estimation of the standard errors of the EM-estimates based on the information matrix given in (12). Tables 1 and 2 show the standard error of the estimates of the parameters (empirical) and the mean of the standard errors obtained from the information matrix (theoretical) for normal gamma and NIG cases, respectively, for some sample sizes. From these tables, we observe a good agreement between the empirical and the estimated theoretical standard errors, mainly for sample sizes $n \geq 100$, for both normal gamma and NIG distributions.

6 Real data application

We now apply the NEF laws and the proposed EM-algorithm in real data to illustrate their usefulness in practical situations. We consider daily log-returns of Petrobras stock from Jan 1st, 2010 to Dec 31th 2018, which consists of 2263 observations. These data can be obtained through the website `https://finance.yahoo.com/`. Denote P_t being the stock price at time t and $Y_t = \log(P_t/P_{t-1})$ the log-return, for $t = 1, \ldots, n$, where n denotes the sample size; in this application, $n = 2263$. According to Schluter and Trede (2016), if N_λ is the number of market transactions in an interval of time (one day, for example), where λ denotes the mean number of transactions, each of these transactions has an associated return, here denoted by X_j, which are a sequence of *i.i.d.* random variables with finite variance. Therefore, under these assumptions, we have that $Y_t = \log(P_t/P_{t-1}) = \sum_{j=1}^{N_\lambda} X_j$. The number of daily transactions of the Petrobras stock is high due to its liquidity, in other words, the mean number of transactions λ is high. This justifies the modeling of these stocks through a NEF class of distributions due to Theorem 2.

Table 3 shows the maximum likelihood estimates of the parameters based on the normal distribution and the EM-estimates for the NG and NIG models. The standard errors are also provided in this table. All models provide similar estimates for the mean and scale parameters as expected. We emphasize that the parameter ϕ controls the departure from the normal distribution. By taking $\phi \to \infty$, we obtain the

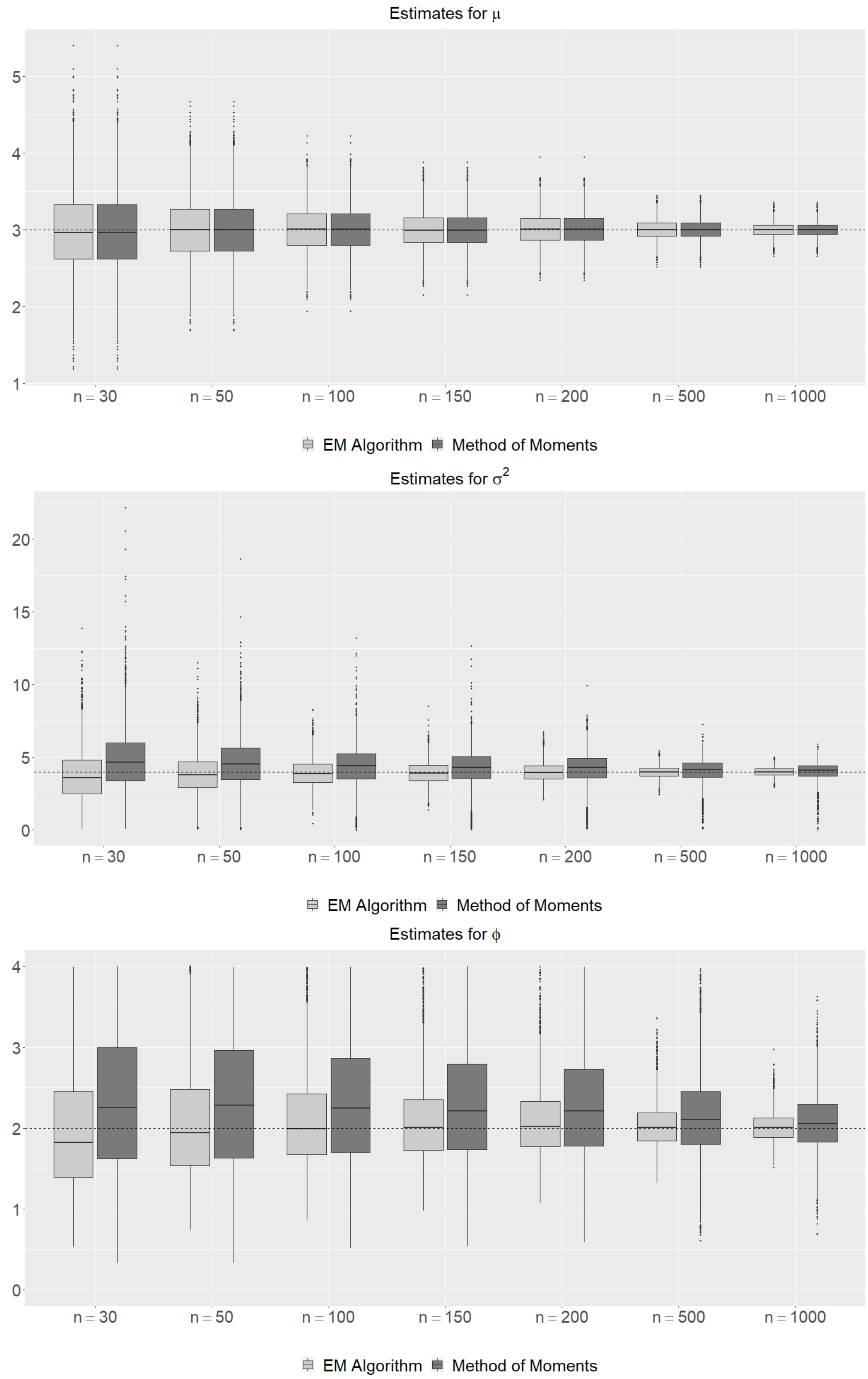


Fig. 3 Boxplots with the estimates of μ, σ^2 and ϕ obtained based on the EM-algorithm and method of moments under normal gamma distribution. Dotted horizontal lines indicate the true value of the parameter.

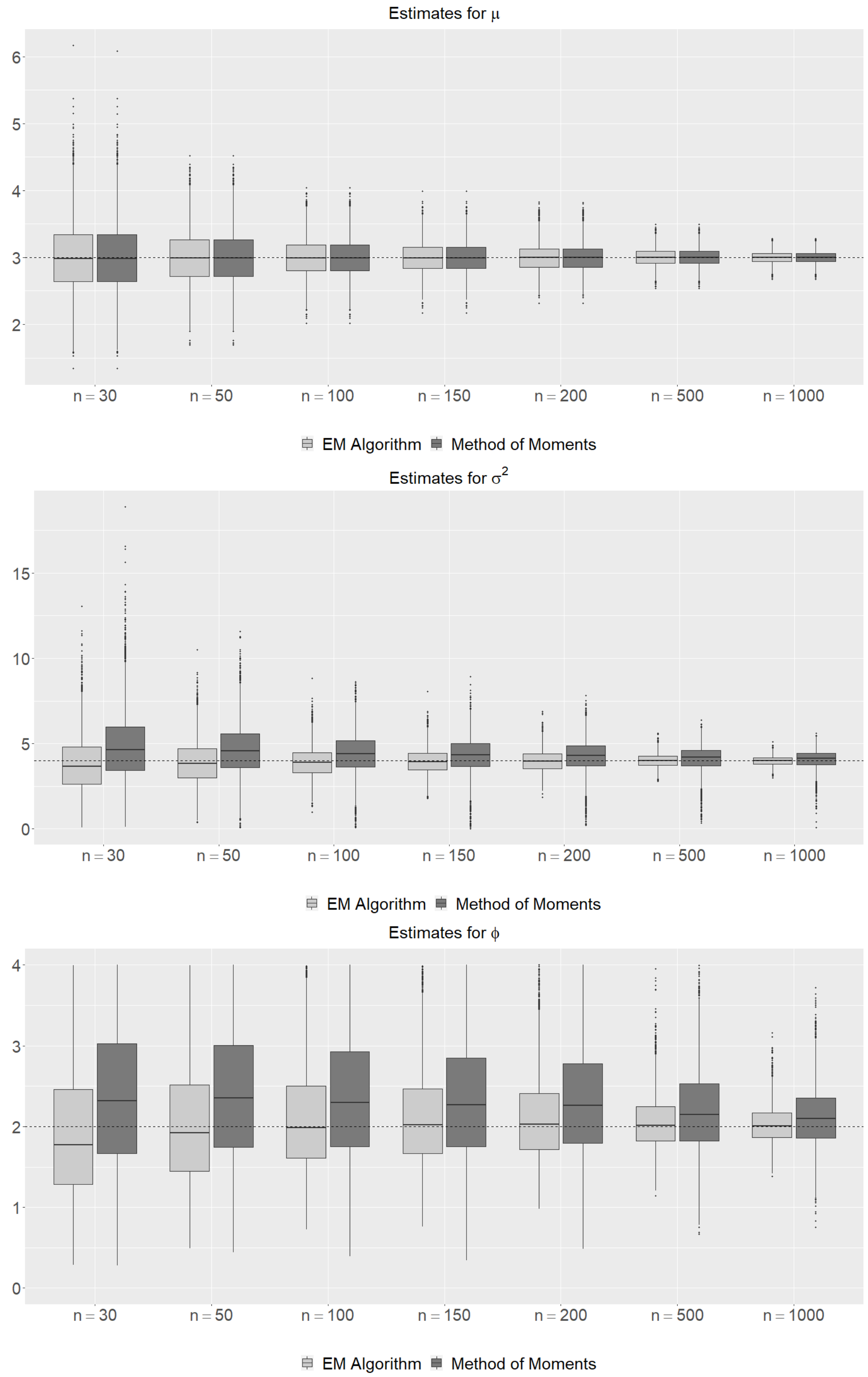


Fig. 4 Boxplots with the estimates of μ, σ^2 and ϕ obtained based on the EM-algorithm and method of moments under normal inverse-Gaussian distribution. Dotted horizontal lines indicate the true value of the parameter.

Table 1 Empirical and theoretical standard errors of parameter estimates under normal gamma assumption.

		μ	σ^2	ϕ
$n = 30$	Empirical	0.5251	1.7598	6.3974
	Theoretical	0.5227	1.6458	4.9593
$n = 50$	Empirical	0.4144	1.3670	3.6329
	Theoretical	0.4073	1.2927	2.1497
$n = 100$	Empirical	0.2937	0.9669	0.8815
	Theoretical	0.2899	0.9242	0.7585
$n = 150$	Empirical	0.2376	0.7768	0.6010
	Theoretical	0.2368	0.7533	0.5414
$n = 200$	Empirical	0.2096	0.6723	0.4754
	Theoretical	0.2056	0.6558	0.4529
$n = 500$	Empirical	0.1313	0.4195	0.2727
	Theoretical	0.1302	0.4159	0.2658
$n = 1000$	Empirical	0.0910	0.2957	0.1851
	Theoretical	0.0922	0.2948	0.1846

Table 2 Empirical and theoretical standard errors of parameter estimates under normal inverse-Gaussian assumption.

		μ	σ^2	ϕ
$n = 30$	Empirical	0.5349	1.6942	21.5352
	Theoretical	0.5287	1.6094	21.2078
$n = 50$	Empirical	0.4067	1.3139	11.3969
	Theoretical	0.4099	1.2545	6.1186
$n = 100$	Empirical	0.2890	0.8872	1.2192
	Theoretical	0.2901	0.8849	0.9936
$n = 150$	Empirical	0.2376	0.7379	0.7910
	Theoretical	0.2374	0.7280	0.6852
$n = 200$	Empirical	0.2083	0.6385	0.6106
	Theoretical	0.2057	0.6309	0.5596
$n = 500$	Empirical	0.1325	0.4013	0.3391
	Theoretical	0.1303	0.4000	0.3266
$n = 1000$	Empirical	0.0903	0.2765	0.2295
	Theoretical	0.0921	0.2827	0.2254

Table 3 Estimates of the parameters with their respective standard errors (in parentheses) for the daily log-return of Petrobras stock prices under NIG, NG, and normal models.

Model	Estimates		
	μ	σ^2	ϕ
Normal	-0.0006 (0.0007)	0.0010 (0.0001)	∞
NG	-0.0006 (0.0006)	0.0009 (0.0001)	1.3105 (0.1105)
NIG	-0.0006 (0.0006)	0.0010 (0.0001)	0.8201 (0.1161)

normal law as a limiting case of the NEF class. The estimates for this parameter under both NG and NIG models indicate some departure from the normal distribution.

For comparison purposes, we also consider the generalized extreme-value (GEV), Student-t (St), and skew-normal (SN) (Azzalini, 1985) distributions with parameter vector (μ, σ, χ), where μ and σ denotes respectively the location and scale. For the GEV, St, and SN distributions, χ denotes the shape ($\mathbb{R}$-valued), degrees of freedom ($\mathbb{R}^+$-valued), and skewness ($\mathbb{R}$-valued) parameter, respectively. The maximum likelihood estimates and their respective standard errors are presented in Table 4. The GEV parameter estimates were obtained by using the `R` package `evd`. For the St and SN fits, we used the package `GAMLSS`.

Table 4 Estimates of the parameters with their respective standard errors (in parentheses) for the daily log-return of Petrobras stock prices under GEV, St, and SN models.

Model	Estimates		
	μ (location)	σ (scale)	χ
GEV	-0.0138 (0.0007)	0.0296 ($2 \cdot 10^{-6}$)	-0.1700 (0.0021)
St	-0.0005 (0.0006)	0.0224 (0.0006)	3.8434 (0.3445)
SN	-0.0155 (0.0035)	0.0346 (0.0016)	0.6459 (0.1733)

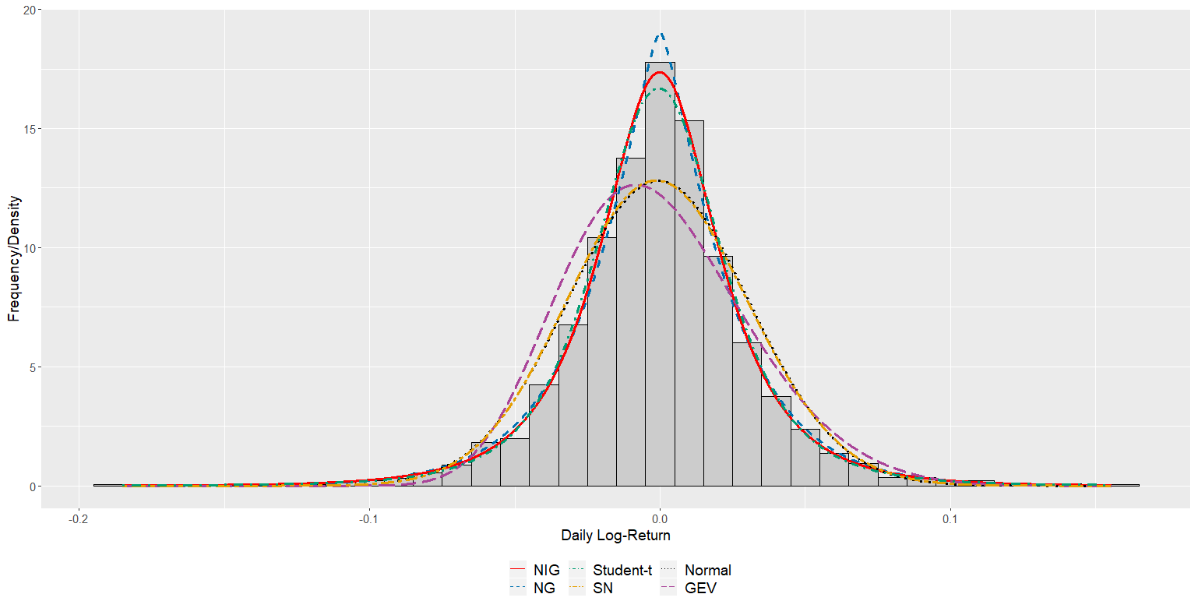


Fig. 5 Histogram of daily log-returns of Petrobras stock price with fitted fitted NIG, NG, St, SN, normal, and GEV densities.

Figure 5 provides the histogram of the data with the estimated densities of NIG, NG, St, SN, GEV, and normal laws. We can observe that the NIG, NG, and St densities capture well the peak, in contrast with the other densities considered. Among the best fits, we note a slight advantage for the NIG model.

To check the goodness-of-fit of the models, we consider qq-plots, which consist of plotting the empirical quantiles against the fitted ones. A well-fitted model provides a qq-plot looking like a linear function $y = x$. Figure 6 exhibits the qq-plots based on the NIG, NG, St, SN, GEV, and normal fitted models. From this figure, we observe that the normal distribution is not suitable for modeling the log-returns, which was already expected. This figure also shows that the SN and GEV models do not perform well. We notice that NIG, NG, and St models provide satisfactory fits, being the first one capturing better the tails. We stress that the modeling of tails is an important task in the study of financial data. For the particular dataset presented in this section, we recommend the use of the NIG law.

Conflicts of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgements

We would like to express our gratitude to the referee for the insightful comments, suggestions, and careful reading which lead to an improvement in the presentation of our paper. G. Oliveira thanks

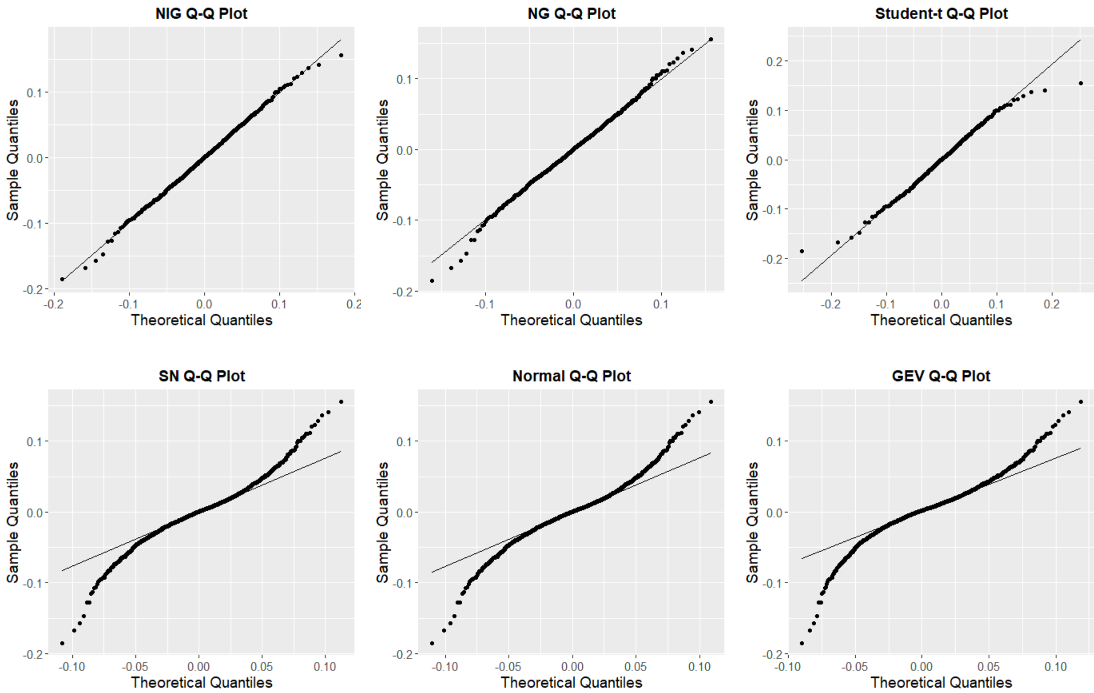


Fig. 6 QQ plots of the daily log-returns of Petrobras stock for the fitted NIG, NG, St, SN, normal, and GEV laws.

the financial support from *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES-Brazil). W. Barreto-Souza acknowledges support for his research from the KAUST Research Fund, NIH 1R01EB028753-01, and the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq-Brazil, grant number 305543/2018-0). This work is part of the Ph.D. thesis of Gabriela Oliveira realized at the Department of Statistics of the Universidade Federal de Minas Gerais, Brazil.

References

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*. **12**, 171-178.

Abramowitz, M. and Stegun, I.A. (1965) Handbook of Mathematical Functions. Dover Publications, New York.

Barndorff-Nielsen, O.E. (1997) Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*. **24**, 1-13.

Barndorff-Nielsen, O.E., Kent, J. Sørensen, M. (1982) Normal variance-mean mixtures and z distributions. *International Statistical Review*. **50**, 145-159.

Barreto-Souza, W. (2015) Long-term survival models with overdispersed number of competing causes. *Computational Statistics and Data Analysis*. **91**, 51-63.

Barreto-Souza, W. and Simas, A.B. (2016) General mixed Poisson regression models with varying dispersion. *Statistics and Computing*. **26**, 1263-1280.

Bening, V.E. and Korolev, V.Yu. (2005) On an application of the Student distribution in the theory of probability and mathematical statistics. *Theory of Probability and its Applications*. **49**, 377-391.

Dempster, A.P. Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society - Series B*. **39**, 1-38.

Embrechts, P., Kluppelberg, C. and Mikosch, T. (2003) Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Feller, W. (1971) An Introduction to Probability Theory and its Applications vol. 2, 2nd edition. Wiley, New York.

Gavrilenko, S.V. and Korolev, V.Yu. (2006) Convergence rate estimates for mixed Poisson random sums. *Sistemy i Sredstva Informatiki*. **special issue**, 248-257.

Gnedenko, B.V. and Korolev, V.Yu. (1996) Random Summation: Limit Theorems and Applications. Boca Raton, FL: CRC Press.

Gut, A. (2013) Probability: A Graduate Course. New York, Springer.

Kalashnikov, V. (1997) Geometric Sums: Bounds for Rare Events with Applications, Kluwer Acad. Publ., Dordrecht.

Karlis, D. and Xekalaki, E. (2005) Mixed Poisson distributions. *International Statistical Review*. **73**, 35-58.

Korolev, V.Yu. (1999) On convergence of distributions of compound Cox processes to stable laws. *Theory of Probability and its Applications.* **43**, 644-650.
Korolev, V.Yu. and Dorofeeva, A. (2017) Bounds of the accuracy of the normal approximation to the distributions of random sums under relaxed moment conditions. *Lithuanian Mathematical Journal.* **57**, 38-58.
Korolev, V.Yu. and Shevtsova, I.G. (2012) An improvement of the Berry-Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scandinavian Actuarial Journal.* **2012**, 81-105.
Korolev, V.Yu. and Zeifman, A. I. (2016) On convergence of the distributions of random sequences with independent random indexes to variance-mean mixtures. *Stochastic Models.* **32**, 414-432.
Korolev, V.Yu. and Zeifman, A. I. (2016) On normal variance–mean mixtures as limit laws for statistics with random sample sizes. *Journal of Statistical Planning and Inference.* **169**, 34-42.
Korolev, V.Yu. and Zeifman, A. (2019) Generalized negative binomial distributions as mixed geometric laws and related limit theorems. *Lithuanian Mathematical Journal.* **59**, 366-388.
Kotz, S., Kozubowski, T. and Podgorski, K (2001) The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance. Springer.
Kozubowski, T.J. and Rachev, S.T. (1994) The theory of geometric stable distributions and its use in modeling financial data. *European Journal of Operational Research.* **74**, 310-324.
Kruglov, V.M., and Korolev, V. Yu. (1990) Limit Theorems for Random Sums, Izdatelstvo Moskovskogo Universiteta, Moscow (in Russian).
Louis, T.A. (1982) Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society - Series B.* **44**, 226-233.
Mittnik, S. and Rachev, S.T. (1989) Stable distributions for asset returns. *Applied Mathematics Letters*, **2**, 301-304.
Mittnik, S. and Rachev, S.T. (1991) Alternative multivariate stable distributions and their applications to financial modeling. *Stable Processes and Related Topics*, 107-119.
Paulsen, J. (2008) Ruin models with investment income. *Probability Surveys.* **5**, 416-434.
Puig, P. and Barquinero, J.F. (2010) An application of compound Poisson modelling to biological dosimetry. *Proceedings of the Royal Society A.* **467**, 897-910.
Rényi, A. (1956) A characterization of the Poisson process. Translated in: P. Turan, ed., Selected Papers of Alfréd Rényi, Vol. I (Akad. Kiado, Budapest, 1976), 622-628.
Sampson, A.R. (1975) Characterizing exponential family distributions by moment generating functions. *Annals of Statistics.* **3**, 747-753.
Shevtsova, I.G. (2018) Convergence rate estimates in the global CLT for compound mixed Poisson distributions. *Theory of Probability and its Applications.* **63**, 72–93.
Schluter, C. and Trede, M. (2016) Weak convergence to the Student and Laplace distributions. *Journal of Applied Probability.* **53**, 121-129.
West, M. (1987) On scale mixtures of normal distributions. *Biometrika.* **74**, 646-648.

Appendix

Proof of Theorem 2

Consider the change of variable $W_\phi^* = \phi W_\phi$. With this, we obtain that the density function of W_ϕ^* assumes the form (3) with $\theta = \xi_0$, $B(\theta) = -\phi b(\theta)$, $R(y) = -\log\phi + c(y;\phi)$, and support $\mathbb{S} = \mathbb{R}^+$. Then, we use Theorem 1 above to obtain the *ch.f.* of W_ϕ^*. This immediately gives us the *ch.f.* of W_ϕ as follows:

$$\Psi_{W_\phi}(t) = \exp\left\{-\phi\left[b(\xi_0) - b\left(\xi_0 + \frac{it}{\phi}\right)\right]\right\}, \ \ t\in\mathbb{R}. \tag{13}$$

From (13), we obtain that

$$\Psi_{N_\lambda}(t) = \Psi_{W_\phi}\left(\lambda(e^{it}-1)\right) = \exp\left\{-\phi\left[b(\xi_0) - b\left(\xi_0 + \frac{i}{\phi}\lambda\left(e^{it}-1\right)\right)\right]\right\}, \quad t\in\mathbb{R}. \tag{14}$$

The Tower Property of conditional expectations gives

$$\Psi_{\widetilde{S}_\lambda}(t) = E\left[E\left(\exp\left\{it\left(a_\lambda\sum_{i=1}^{N_\lambda}(X_i + d_\lambda)\right)\right\}\Bigg|N_\lambda\right)\right].$$

Let $G_{N_\lambda}(\cdot)$ denote the probability generating function of N_λ. Then, we use (14) to obtain

$$\begin{aligned}\Psi_{\widetilde{S}_\lambda}(t) &= G_{N_\lambda}\left(\Psi_{X_1-\mu}\left(\frac{t}{\sqrt{\lambda}}\right)e^{i\frac{\mu}{\lambda}t}\right) = \Psi_{N_\lambda}\left(\frac{1}{i}\log\left\{\Psi_{X_1-\mu}\left(\frac{t}{\sqrt{\lambda}}\right)e^{i\frac{\mu}{\lambda}t}\right\}\right)\\ &= \exp\left\{-\phi\left[b(\xi_0) - b\left(\xi_0 + \frac{i\lambda}{\phi}\left(\Psi_{X_1-\mu}\left(\frac{t}{\sqrt{\lambda}}\right)e^{i\frac{\mu}{\lambda}t} - 1\right)\right)\right]\right\}.\end{aligned} \tag{15}$$

Taking $\lambda \to \infty$ and applying L'Hôpital's rule twice (in the second-order derivative we are using the assumption of finite variance of X_1) we obtain

$$\lim_{\lambda\to\infty} \Psi_{\widetilde{S}_\lambda}(t) = \exp\left\{-\phi\left[b(\xi_0) - b\left(\xi_0 + \phi^{-1}\left(it\mu - \frac{t^2\sigma^2}{2}\right)\right)\right]\right\} \equiv \Psi_Y(t),\ \forall\, t \in \mathbb{R}.$$

This completes the proof. □

Proof of Theorem 3

If Y is MP-stable, then $\mathcal{D}(G) \neq \emptyset$. By Lévy's Continuity Theorem and (15), the convergence in Expression (5) holds if and only if

$$\exp\left\{-\phi\left[b(\xi_0) - b\left(\xi_0 + \frac{\lambda}{\phi}(\varphi_\lambda(t) - 1)\right)\right]\right\} \xrightarrow[\lambda\to\infty]{} \Psi_Y(t), \tag{16}$$

where $\varphi_\lambda(t) = \Psi_{X_1}(a_\lambda t)\exp(ita_\lambda d_\lambda)$.

The first derivative of the function $b(\cdot)$ is positive since it is equal to the mean of the exponential family (its support here is assumed to be $\mathbb{R}^+$, so that the mean is positive). This implies that $b(\cdot)$ is strictly monotone increasing and therefore it is invertible. Using this invertibility, it follows that (16) is equivalent to

$$\lambda(\varphi_\lambda(t) - 1) \xrightarrow[\lambda\to\infty]{} \phi\left\{b^{-1}\left(b(\xi_0) + \frac{1}{\phi}\log\Psi_Y(t)\right) - \xi_0\right\},\ \ \forall t \in \mathbb{R}.$$

We can take the limit above with $\lambda = \lambda_n = n \in \mathbb{N}$ instead of $\lambda \in \mathbb{R}$; $\{\lambda_n\}$ can be seen as a subsequence. In this case, letting $a_\lambda = a_n$ and $d_\lambda = d_n$, it follows that

$$n(\varphi_n(t) - 1) \xrightarrow[n\to\infty]{} \phi\left\{b^{-1}\left(b(\xi_0) + \frac{1}{\phi}\log\Psi_Y(t)\right) - \xi_0\right\},\ \ \forall t \in \mathbb{R}.$$

From Theorem 1 in Chapter XVII in Feller (1971), the above limit implies that

$$(\varphi_n(t))^n \xrightarrow[n\to\infty]{} \exp\left\{\phi\left\{b^{-1}\left[b(\xi_0) + \frac{1}{\phi}\log\Psi_Y(t)\right] - \xi_0\right\}\right\},\ \ \forall t \in \mathbb{R}. \tag{17}$$

Since the left side of (17) is the *ch.f.* of

$$a_n\sum_{i=1}^{n}(X_i + d_n),$$

it follows that the right side of that equation is the *ch.f.* of some α-stable random variable A (see Theorem 3.1, Chapter 9 in Gut (2013)), say $\Psi_A(t)$. Hence, we obtain

$$\Psi_Y(t) = \exp\left\{-\phi\left[b(\xi_0) - b\left(\xi_0 + \frac{1}{\phi}\log\Psi_A(t)\right)\right]\right\},$$

which gives us the sufficiency part of the theorem.

Conversely, if (6) holds and $\Psi_A(t)$ is the *ch.f.* of some α-stable random variable A, then there exist an *i.i.d.* sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables and real sequences $\{a_n\}_{n\in\mathbb{N}}$ and $\{d_n\}_{n\in\mathbb{N}}$ such that

$$a_n\sum_{i=1}^{n}(X_i + d_n) \xrightarrow[n\to\infty]{d} U. \tag{18}$$

Denote by $\Psi_{X_1}(t)$ the *ch.f.* of X_1 and let $\gamma_n(t) = \Psi_{X_1}(a_n t)\exp(ita_n d_n)$, $n \in \mathbb{N}$. Then, the weak limit in (18) is equivalent to

$$(\gamma_n(t))^n \xrightarrow[n\to\infty]{} \Psi_A(t),\ \ \forall t \in \mathbb{R}.$$

From Feller (1971), we have that the above limit implies that

$$n(\gamma_n(t) - 1) \xrightarrow[n\to\infty]{} \log\Psi_A(t).$$

Since by hypothesis

$$\Psi_A(t) = \exp\left\{\phi\left\{b^{-1}\left(b(\xi_0) + \frac{1}{\phi}\log\Psi_Y(t)\right) - \xi_0\right\}\right\},$$

we obtain that

$$n(\gamma_n(t) - 1) \xrightarrow[n\to\infty]{} \phi\left\{b^{-1}\left(b(\xi_0) + \frac{1}{\phi}\log\Psi_Y(t)\right) - \xi_0\right\},$$

which is equivalent to

$$\lambda(\gamma_\lambda(t) - 1) \xrightarrow[\lambda\to\infty]{} \phi\left\{b^{-1}\left(b(\xi_0) + \frac{1}{\phi}\log\Psi_Y(t)\right) - \xi_0\right\}.$$

The above limit gives Equation (16) with $\gamma_\lambda(t)$ instead of $\varphi_\lambda(t)$. This shows that $\mathcal{D}(G) \neq \emptyset$ and the proof is complete. □

Proof of Proposition 1

By standard properties of conditional expectation and using the *ch.f.* of W_ϕ given in (13), we obtain that

$$\begin{aligned}\Psi_{\mu W_\phi + \sigma\sqrt{W_\phi}Z}(t) &= E\left(\exp\left\{it(\mu W_\phi + \sigma\sqrt{W_\phi}Z)\right\}\right) = E\left[E\left(\exp\{it(\mu W_\phi + \sigma\sqrt{W_\phi}Z)\}|W_\phi\right)\right] \\ &= \int_0^\infty e^{it\mu w}E\left(e^{it\sigma\sqrt{w}Z}\right)f_{W_\phi}(w)dw = \int_0^\infty \exp\left\{w\left(it\mu - \frac{1}{2}t^2\sigma^2\right)\right\}f_{W_\phi}(w)dw \\ &= \Psi_{W_\phi}\left(t\mu + \frac{it^2\sigma^2}{2}\right) = \exp\left\{-\phi\left[b(\xi_0) - b\left(\xi_0 + \frac{1}{\phi}\left(it\mu - \frac{1}{2}t^2\sigma^2\right)\right)\right]\right\},\end{aligned}$$

which is the characteristic function given in Theorem 2. □

Proof of Proposition 3

We have that

$$E\left(W_\phi^K g(W_\phi)^L\middle|Y\right) = \int_0^\infty w^K g(w)^L \frac{f_{Y|W}(y|w)f_W(w)}{f_Y(y)}dw =$$
$$\frac{1}{f_Y(y)}\int_0^\infty \frac{w^K g(w)^L}{\sqrt{2\pi\sigma^2 w}}\exp\left\{-\frac{(y-\mu w)^2}{2\sigma^2 w} + \phi[-w + \log(\phi) + \log(w)] - \log\Gamma(\phi) - \log w\right\}dw.$$

By using the explicit form of the normal-gamma distribution density given in Expression (8), we get

$$\begin{aligned}E\left(W_\phi^K g^L(W_\phi)\middle|Y\right) &= \frac{\left(\frac{\mu^2+2\phi\sigma^2}{y^2}\right)^{\frac{\phi}{2}-\frac{1}{4}}}{2\mathcal{K}_{\phi-\frac{1}{2}}\left(\sqrt{\left[\frac{\mu^2}{\sigma^2}+2\phi\right]\frac{y^2}{\sigma^2}}\right)} \\ &\times \int_0^\infty g(w)^L \underbrace{w^{\left(\phi+K-\frac{1}{2}\right)-1}\exp\left\{-\frac{1}{2}\left[\left(\frac{\mu^2}{\sigma^2}+2\phi\right)w + \left(\frac{y^2}{\sigma^2}\right)\frac{1}{w}\right]\right\}}_{\text{Kernel GIG}\left(\frac{\mu^2}{\sigma^2}+2\phi,\frac{y^2}{\sigma^2},\phi+K-\frac{1}{2}\right)}dw.\end{aligned}$$

Denoting $U \sim \text{GIG}(a, b, p)$ with $a = \frac{\mu^2}{\sigma^2} + 2\phi$, $b = \frac{y^2}{\sigma^2}$ and $p = \phi + K - \frac{1}{2}$ and noting that the integrand above is the kernel of a GIG density function, we obtain the desired result. □

Proof of Proposition 4

We have that

$$
\begin{aligned}
E\left(W_\phi^K g(W_\phi)^L \middle| Y=y\right) &= \int_0^\infty w^K g(w)^L \frac{f_{Y|W}(y|w) f_W(w)}{f_Y(y)} dw \\
&= \frac{1}{f_Y(y)} \int_0^\infty w^K g(w)^L \frac{1}{\sqrt{2\pi\sigma^2 w}} e^{-\frac{(y-\mu w)^2}{2\sigma^2 w}} e^{-\frac{\phi}{2}\left(w+\frac{1}{w}\right)+\frac{1}{2}\left(\log\phi - \log(2\pi w^3)\right)} dw.
\end{aligned}
$$

Let $f_Y(y)$ be the NIG density function as given in (9) and denote $U \sim \mathrm{GIG}\,(a,b,p)$ with $a = \frac{\mu^2}{\sigma^2}+\phi$, $b = \frac{y^2}{\sigma^2}+\phi$ and $p = K-1$. It follows that

$$
\begin{aligned}
E\left(W_\phi^K g(W_\phi)^L \middle| Y=y\right) &= \frac{\left(\frac{y^2+\phi\sigma^2}{\mu^2+\phi\sigma^2}\right)^{\frac{1}{2}}}{\mathcal{K}_{-1}\left(\sqrt{\left(\frac{\mu^2}{\sigma^2}+\phi\right)\left(\frac{y^2}{\sigma^2}+\phi\right)}\right)} \\
&\quad \times \int_0^\infty g(w)^L \underbrace{w^{(K-1)-1} \exp\left\{-\frac{1}{2}\left[\left(\frac{\mu^2}{\sigma^2}+\phi\right) w + \left(\frac{y^2}{\sigma^2}+\phi\right)\frac{1}{w}\right]\right\}}_{\text{Kernel GIG}\left(\frac{\mu^2}{\sigma^2}+\phi,\, \frac{y^2}{\sigma^2}+\phi,\, K-1\right)} dw \\
&= \frac{\mathcal{K}_{K-1}(\sqrt{ab})}{\mathcal{K}_{-1}(\sqrt{ab})} \left(\frac{b}{a}\right)^{\frac{K}{2}} E\left(g(U)^L\right).
\end{aligned}
$$

This completes the proof. □