Goodness-of-fit tests for compound distributions with applications in insurance

P.-O. Goffard[1], S. Rao Jammalamadaka[2], and S. G. Meintanis[3,4]

[1]Univ Lyon, Université Lyon 1, LSAF EA2429, Institut de Science Financière et d'Assurances, Lyon, France
[2]Department of Statistics and Applied Probability, University of California, Santa Barbara, USA
[3]Department of Economics, National and Kapodistrian University of Athens, Athens, Greece
[4]Unit for Business Mathematics and Informatics, North-West University Potchefstroom, South Africa*

May 3, 2019

Abstract

Goodness–of–fit procedures are provided to test the validity of compound models for the total claims, involving specific laws for the constituent components, namely the claim frequency distribution and the distribution of individual claim sizes. This is done without the need for observations on these two component variables. Goodness-of-fit tests that utilize the Laplace transform as well as classical tools based on the distribution function, are proposed and compared. These methods are validated by simulations and then applied to insurance data.

MSC 2010: 60G55, 60G40, 12E10.
Keywords: Compound distributions; goodness-of-fit tests; Katz family; Laplace transform.

1 Introduction

Consider the random variable,

$$X = \sum_{k=1}^{N} U_k, \tag{1}$$

where N is a count random variable with probability mass function p^N and the U_k's form a sequence of independently and identically distributed (**iid**) non-negative random variables having a distribution function (**DF**) F^U, and independent of N.

The "compound" random variable in (1) has several interpretations and has many practical applications. We focus on the risk–related interpretation of model (1), as the total amount of claim associated with a non-life insurance portfolio over a fixed time

*On sabbatical leave from the University of Athens period. The random variable N represents the claim frequency while the U_k's represent the individual claim sizes. Our goal is to identify the components of model (1), namely the distribution of the claim frequency N and of the individual claim sizes U, based only on observations of X. Such incomplete data situations arise in practice when an insurance company only keeps track of monthly, quarterly, or yearly aggregated data. The methodology can also be useful to a reinsurance company, which has only access to partial information on X, and would like to better understand the underlying risk and improve the rate-making. Model (1) is also useful in the banking industry which has only access to data on annualized operational risk. As pointed out in Chaudhury [4], the data available to assess operational risk are typically incomplete as banks often report aggregated losses and discard small losses. Such loss of information also occurs when merging the data after acquiring another banking operation.

Assume we observe aggregated claim size $X_1, \ldots, X_n$, and we wish to assess the consistency with particular parametric models specified for the claim frequency N and the individual claim amount U. We are interested in the composite situation whereby the **DF**'s involved depends on unknown parameters, say $p^N := p^N(.\,;\vartheta_N)$ and $F^U := F^U(.\,;\vartheta_U)$, with the parameter vector $\vartheta = (\vartheta_N, \vartheta_U)$ treated as unknown. If F^X denotes the DF of the compound r.v. X, we wish to test the null hypothesis

$$H_0 : \text{The DF of } X \text{ in (1.1) is } \; F^X \equiv F^X(.\,;\vartheta), \text{ for some } \vartheta \in \Theta, \tag{2}$$

where Θ denotes an appropriate parameter space.

Two types of non-parametric goodness-of-fit (**GOF**) tests are considered. The first type is based on a dissimilarity measure between the population **DF** and the empirical **DF** of the available data; see e.g. D'Agostino and Stephens [7], or Thas [40] for reviews on the subject of **DF**–based **GOF** tests. Since the random variable X typically has a point mass at 0 corresponding to $N = 0$, the standard Kolmogorov-Smirnov (**KS**) and Cramèr-von Mises (**CvM**) **GOF** tests need some corresponding modifications. Although these procedures are originally meant to handle continuous data, extensions have been proposed to assess the adequacy for discrete, grouped, or mixed data. See for instance Schmid [31], Walsh [41], Noether [27], Slakter [32], Conover [6], Gleser [13], and Dimitrova et al. [8] for the **KS GOF** test. Regarding the **CvM** criterion, the reader is referred to the works of Choulakian et al. [5], Henze [18], Spinelli and Stephens [34], Spinelli [33] and Lockhart et al. [24]. We propose modified estimators of the **KS** and **CvM** test statistics that take care of the point mass at 0 and also address the lack of closed form expression for the **DF** of X.

A second group of procedures we employ here, measure the model discrepancy in terms of the distance between the population Laplace transform (**LT**) and the empirical **LT**. Statistical tests involving this approach work directly with transform–based statistics, thus avoiding **LT** inversion which is often complex and costly. These methods are quite convenient in cases where the **DF** is complicated while the **LT** is readily available. Such methods are relatively new but since their introduction they have been used in various estimation and testing problems; see for instance Laurence and Morgan [23], Henze [17], Yao and Moran [42], Henze and Klar [19], Henze and Meintanis [20], Meintanis and Iliopoulos [25], Besbeas and Morgan [2], and Ghosh and Beran [11].

The test statistics involved in these procedures lead to non-standard asymptotic distributions for which finding critical values requires sophisticated numerical methods. Moreover, due to the fact that the parameters of the null distribution have to be estimated a priori, the tests are not distribution free. We overcome these difficulties using a parametric bootstrap approach, which has gained popularity in approximating the null distribution in goodness-of-fit testing.

The paper is organized as follows. Section 2 provides a brief background on compound distributions and reviews moment-based estimation of the parameters. Section 3 details the goodness of fit testing procedures tailored to the distribution of aggregated claim sizes. Section 4 reports the results of a Monte-Carlo simulation study conducted to compare of the **GOF** tests in terms of power. Finally, Section 5 presents the application of our **GOF** procedures to real data sets from the insurance industry.

2 Preliminaries

2.1 Compound Distribution

Recall that $X = \sum_{k=1}^{N} U_i$, where N is a counting **rv** with probability mass function p^N and the U_k's are **iid** non-negative random variables with **DF** F^U, and independent of N. Given the fact that N can take the value zero with probability $p^N(0)$, the **DF** of X is given by

$$F^X(x) = p^N(0) + \left[1 - p^N(0)\right] F^{X|N>0}(x),\ x \geq 0, \tag{3}$$

where $F^{X|N>0}$ denotes the **DF** of X provided that $N > 0$. Note that the conditional probability distribution of $X|N > 0$ is continuous. The Laplace transform of X, defined as $L^X(t) := \mathbb{E}(e^{-tX})$, may be expressed as

$$L^X(t) = G^N\left[L^U(t)\right],\ t \geq 0, \tag{4}$$

where $G^N(t) := \mathbb{E}(t^N)$ denotes the probability generating function of N and L^U is the **LT** of the claim size distribution.

We let the distribution of the claim sizes be quite general (besides its parametric form), but model the claim frequency via a counting distribution from the Katz family. Recall that the distribution of N belongs to the Katz family [22], written $N \sim \mathrm{KF}(a, b)$, if the probability mass function satisfies the recursive equation

$$p^N(k) = \left(a + \frac{b}{k}\right) p^N(k-1),\ \text{for } k \geq 1. \tag{5}$$

A characterization is given in Sundt and Jewell [39]. Prominent members of this family (5) include

1. The binomial $N \sim \mathrm{Bin}(\alpha, p)$ ($\alpha \in \mathbb{N}$, $0 < p < 1$) with probability mass function

   $$p^N(k) = \binom{\alpha}{k} p^k (1-p)^{\alpha-k},\ \text{for } k = 0, \ldots, \alpha, \tag{6}$$

   which satisfies (5) with $\alpha = -(a+b)/a$ and $p = -a/(1-a)$.

2. The Poisson $N \sim \text{Pois}(\lambda)$ ($\lambda > 0$) with probability mass function

$$p^N(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \text{ for } k \geq 0, \tag{7}$$

which satisfies (5) with $a = 0$ and $b = \lambda$.

3. The negative binomial $N \sim \text{Neg-Bin}(\alpha, p)$ ($\alpha > 0$, $0 < p < 1$) with probability mass function

$$p^N(k) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\Gamma(k+1)} p^{\alpha}(1-p)^k, \text{ for } k \geq 0, \tag{8}$$

which satisfies (5) with $\alpha = (a+b)/a$ and $p = 1 - a$.

These discrete distributions are commonly used to model claim frequencies and this choice seems quite general and justified.

Throughout this paper, we assume that we hold a sample $X_1, \ldots, X_n$ of n **iid** observations distributed as X. Among these observations, we will denote by $n_0 \leq n$ the number of zeros, while the remaining $n - n_0$ non-zeros are denoted by $X_1^+, \ldots, X_{n-n_0}^+$.

2.2 Moments based estimation for aggregate claims

The method of moment estimator is obtained by matching the empirical moments with the theoretical moments of the parametric model. If $N \sim \text{KF}(a, b)$ then the moments of U may be expressed in terms of the moments of X by inverting the recursive relations provided in De Pril [29, Equation 3], namely it holds that

$$(1-a)\mathbb{E}\left(X^{k+1}\right) = \sum_{i=0}^{k} \binom{k}{i} \left(\frac{k+1}{i+1}a + b\right) \mathbb{E}\left(U^{i+1}\right) \mathbb{E}\left(X^{k-i}\right), \text{ for } k \geq 0. \tag{9}$$

Solving the system (9) for $\vartheta_1 = (a, b)$ and ϑ_2 yields the Method of Moments Estimators (MMEs). A few limitations are outlined in the following remark.

Remark 2.1. *First, since the system of equations* (9) *is not necessarily linear, they may not yield an explicit analytical solution. Second, the proposed method only applies if the method of moment estimators for the claim size distribution are valid. Such is not the case if the moments of the claim size distribution do not exist, as in the case of distributions having Pareto–like tails, or if the distribution is not characterized by its moments, as in the log-normal case.*

Denote by $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $m_k = n^{-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^k$ the sample mean and the sample centered moments of order $k \geq 2$, respectively. The following examples provide expressions for the MMEs in the Poisson-exponential, geometric-exponential and Poisson-gamma cases:

Example 1. *1. (Poisson-exponential): Assume a Poisson frequency for $N \sim Pois(\lambda)$, with claim size $U \sim Exp(\theta)$ following an exponential distribution, with density*

$$f_U(x) = \frac{e^{-x/\theta}}{\theta}, \textit{ for } x \geq 0. \tag{10}$$

The Poisson distribution with parameter λ corresponds to $a = 0$, $b = \lambda$ in the Katz family parametrization in (5), while for the exponential distribution with parameter θ we have, $\mathbb{E}(U) = \theta$, and $\mathbb{E}\left(U^2\right) = 2\theta^2$. Substituting in (9) *yields*

$$\begin{cases} \mathbb{E}(X) = \lambda\theta, \\ \mathbb{E}\left(X^2\right) = 2\lambda\theta^2 + \lambda\theta\mathbb{E}(X). \end{cases}$$

The parameters are then estimated by

$$\widehat{\theta} = \frac{m_2}{2\bar{X}} \text{ and } \widehat{\lambda} = \frac{2\bar{X}^2}{m_2}. \tag{11}$$

2. *(geometric-exponential): Assume that the claim sizes follow an exponential distribution, $U \sim Exp(\theta)$, and that the claim frequency is geometric $N \sim Neg\text{-}Bin(1, p)$. Thus we have that $a = 1 - p$, $b = 0$, in the Katz family parametrization in (5). Substituting in* (9) *yields*

$$\begin{cases} p\mathbb{E}(X) = (1-p)\theta, \\ p\mathbb{E}\left(X^2\right) = 2(1-p)\theta^2 + 2(1-p)\theta\mathbb{E}(X). \end{cases}$$

The parameters are then estimated via

$$\widehat{\theta} = \frac{m_2 - \bar{X}^2}{2\bar{X}} \text{ and } \widehat{p} = \frac{\bar{X}}{\widehat{\theta} + \bar{X}}. \tag{12}$$

3. *(Poisson-gamma): Assume that the claim sizes follow a gamma distribution, $U \sim gamma(r, \theta)$, with density*

$$f_U(x) = \frac{e^{-x/\theta} x^{r-1}}{\theta^r \Gamma(r)}, \text{ for } x \geq 0. \tag{13}$$

Let the claim frequency be Poisson distributed $N \sim Pois(\lambda)$. We have that $a = 0$, $b = \lambda$, $\mathbb{E}(U) = r\theta$, $\mathbb{E}\left(U^2\right) = r(r+1)\theta^2$, and $\mathbb{E}\left(U^3\right) = r(r+1)(r+2)\theta^3$. Substituting in (9) *yields*

$$\begin{cases} \mathbb{E}(X) = \lambda r\theta, \\ \mathbb{E}\left(X^2\right) = r(r+1)\lambda\theta^2 + \lambda r\theta\mathbb{E}(X), \\ \mathbb{E}\left(X^3\right) = \lambda r\theta\mathbb{E}(X^2) + 2\lambda r(r+1)\theta^2\mathbb{E}(X) + \lambda r(r+1)(r+2)\theta^3. \end{cases}$$

The parameters are then estimated by

$$\widehat{r} = \frac{2m_2^2 - m_3\bar{X}}{m_3\bar{X} - m_2^2}\ , \widehat{\theta} = \frac{m_2}{\bar{X}(\widehat{r}+1)}, \quad \text{and } \widehat{\lambda} = \frac{\bar{X}}{\widehat{r}\widehat{\theta}}. \tag{14}$$

Remark 2.2. *The estimates of the parameters in Examples* (12) *and* (14) *may turn out to be negative due to the lack of fit of the model. The partial Method of Moments presented below often resolves this difficulty.*

The "partial Method of Moments" idea is the following: whenever the data consists of several X_i that take the value zero i.e. $n_0 > 0$, consider adding to the system of equations (9), an additional estimation equation corresponding to the probability of this event. If $N \sim \text{Bin}(\alpha, p)$ or $N \sim \text{Neg-Bin}(\alpha, p)$, it is given by

$$p_N(0) = p^\alpha, \tag{15}$$

and when $N \sim \text{Pois}(\lambda)$, it is

$$p_N(0) = e^{-\lambda}. \tag{16}$$

This probability that $N = 0$ is estimated by $\widehat{p}_N(0) = n_0/n$. The parameters of the claim sizes distribution follow from the other MME equations. The resulting estimates are referred to as partial Method of Moments Estimators (*partial*-MMEs) in the remainder. The following example provides the expressions of the *partial*-MMEs in the geometric-exponential, Poisson-gamma and Poisson-inverse Gaussian cases.

Example 2. 1. *Assume that the claim frequency is geometric $N \sim$ Neg-Bin$(1, p)$, then $b = 0$ and $a = p$ is estimated via*

$$\widehat{p} = \frac{n_0}{n}.$$

Let the claim sizes be exponentially distributed $U \sim$ Exp(θ). The parameter is estimated via

$$\widehat{\theta} = \frac{(1-\widehat{p})\bar{X}}{\widehat{p}}.$$

Hence, the partial-MME cannot be negative in the geometric-exponential case.

2. *Assume that the claim frequency is Poisson distributed $N \sim$ Pois(λ), then $a = 0$ and $b = \lambda$ is estimated via*

$$\widehat{\lambda} = -\log\left(\frac{n_0}{n}\right).$$

Let the claim sizes be gamma distributed $U \sim$ Gamma(r, θ). Substituting in (9) *yields*

$$\begin{cases} \mathbb{E}(X) = \widehat{\lambda} r\theta, \\ \mathbb{E}\left(X^2\right) = r(r+1)\widehat{\lambda}\theta^2 + \widehat{\lambda} r\theta\mathbb{E}(X). \end{cases}$$

The parameters are then estimated via

$$\widehat{r} = \frac{\bar{X}^2}{\widehat{\lambda} m_2 - \bar{X}^2} \text{ and } \widehat{\theta} = \frac{\bar{X}}{\widehat{\lambda}\widehat{r}}. \tag{17}$$

These estimators do not involve the third order moment anymore, and if λ is in a reasonable range ($\lambda > \frac{\bar{X}^2}{m_2}$) , their values will be non-negative.

3. *Assume that the claim frequency is Poisson distributed $N \sim$ Pois(λ), then $a = 0$ and $b = \lambda$ is estimated via*

$$\widehat{\lambda} = -\log\left(\frac{n_0}{n}\right).$$

Let the claim sizes be inverse-Gaussian distributed $U \sim IG(\mu, \phi)$, with density

$$f_U(x) = \left(\frac{1}{2\pi x^3\phi}\right) \exp\left[-\frac{(x-\mu)^2}{2\mu^2\phi x}\right], \; x > 0. \tag{18}$$

Substituting in (9) *yields*

$$\begin{cases} \mathbb{E}(X) = \widehat{\lambda}\mu, \\ \mathbb{E}\left(X^2\right) = \widehat{\lambda}\mu\mathbb{E}(X) + \widehat{\lambda}\mu^2(1+\lambda\mu). \end{cases}$$

The parameters are then estimated via

$$\widehat{\mu} = \frac{\bar{X}}{\widehat{\lambda}} \text{ and } \widehat{\phi} = \frac{\widehat{\lambda}m_2 - \bar{X}^2}{\bar{X}}. \tag{19}$$

These estimators do not involve the third order moment, and if λ *is in a reasonable range* ($\lambda > \frac{\bar{X}^2}{m_2}$) *, their values will be non-negative.*

3 Goodness-of-fit tests for aggregate claims

3.1 Tests based on the distribution function

As already mentioned, a **DF**-based **GOF** test compares the population **DF** $F_0^X(x;\vartheta) = \mathbb{P}_0(X \leq x \big| \vartheta)$, $x \in \mathbb{R}$, to its empirical counterpart, the empirical **DF**, defined by

$$F_n^X(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i \leq x). \tag{20}$$

Note that the empirical **DF** may be rewritten as

$$F_n^X(x) = \frac{n_0}{n} + \frac{n-n_0}{n} F_n^{X|N>0}(x),$$

where $F_n^{X|N>0}$ denotes the empirical **DF** of X given that $N > 0$, which can be estimated via

$$F_n^X(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i^+ \leq x). \tag{21}$$

Denote by $X^+_{1:n-n_0}, \ldots, X^+_{n-n_0:n-n_0}$ the order statistics associated to the sample $X_1^+, \ldots, X^+_{n-n_0}$. Finally, we estimate the population **DF** as $\widehat{F}^X(x) := F^X(x;\widehat{\vartheta})$ where $\widehat{\vartheta} = \widehat{\vartheta}(X_1, \ldots X_n)$ is some asymptotically efficient estimator. As we assumed that the distribution of N belongs to Katz's family, the population **DF** may be approximated via the so-called Panjer algorithm, see [28] for more details.

3.1.1 *Kolmogorov-Smirnov* test for compound distributions

The Kolmogorov-Smirnov **GOF** test employs the distance

$$\mathrm{KS}\left(F^X, F_n^X\right) = \sqrt{n} \sup_{x\in(0,\infty)} |F^X(x) - F_n^X(x)| := \sqrt{n} \sup_{x\in(0,\infty)} D_n^{\mathrm{KS}}(x). \tag{22}$$

Define the intervals $I_i = [X^+_{i-1:n-n_0}, X^+_{i:n-n_0})$ for $i = 1, \ldots, n-n_0$, with the convention $X_{0:n-n_0} = 0$. Then for $x \in I_i$,

$$D_n^{\mathrm{KS}}(x) = \left|\frac{n_0 + i}{n} - F^X(x)\right|, \; i = 0, \ldots, n-n_0,$$

so that

$$\sup_{x \in I_i} D_n^{\mathrm{KS}}(x) = \max \left[\frac{n_0 + i}{n} - F^X \left(X_{i:n-n_0}^+ \right), F^X \left(X_{i+1:n-n_0}^+ \right) - \frac{n_0 + i}{n} \right].$$

Considering successively the intervals I_i, we estimate the **KS** distance (22) by

$$\mathrm{KS}_n = \max \left(D_-, D_+ \right),$$

where

$$D_- = \max_{0 \le i \le n-n_0} \frac{n_0 + i}{n} - \widehat{F}^X \left(X_{i:n-n_0}^+ \right), \text{ and } D_+ = \max_{0 \le i \le n-n_0} \widehat{F}^X \left(X_{i+1:n-n_0}^+ \right) - \frac{n_0 + i}{n}.$$

3.1.2 *Cramér-von Mises* test for compound distributions

The *Cramér-von Mises* **GOF** test uses the criterion

$$\mathrm{CM} \left(F^X, F_n^X \right) = n \int_0^{+\infty} \left[F^X(x) - F_n^X(x) \right]^2 \mathrm{d}F^X(x). \tag{23}$$

Given the mixed nature of the distribution of X, the probability measure follows from differentiation in (3) with

$$\mathrm{d}F^X(x) = p^N(0)\delta_0(x) + \left[1 - p^N(0) \right] \mathrm{d}F^{X|N>0}(x), \; x \ge 0, \tag{24}$$

where $\delta_0(x)$ denotes the Dirac measure at 0. Reinserting (24) into the integral (23) yields

$$\begin{aligned} \mathrm{CM} \left(F^X, F_n^X \right) &= n \Bigg\{ p^N(0) \left[p^N(0) - \frac{n_0}{n} \right]^2 \\ &+ \left[1 - p^N(0) \right] \int_0^{+\infty} \left[F^{X|N>0}(x) - F_n^{X|N>0}(x) \right]^2 \mathrm{d}F^{X|N>0}(x) \Bigg\}. \end{aligned} \tag{25}$$

Because the non-negative data points $X_1^+, \ldots, X_{n-n_0}^+$ are assumed to be distributed as $F^{X|N>0}$ under the null hypothesis, we can estimate the integral in (25) by

$$\begin{aligned} \int_0^{+\infty} \left[F^X(x) - F_n(x) \right]^2 \mathrm{d}F^{X|N>0}(x) &= \frac{1}{n-n_0} \sum_{i=1}^{n-n_0} \left[F^X(X_{i:n-n_0}^+) - \frac{2i-1}{2(n-n_0)} \right]^2 \\ &+ \frac{1}{12(n-n_0)^2}. \end{aligned}$$

Finally putting these together, the **CvM** criterion is estimated by

$$\begin{aligned} \mathrm{CM}_n &= n \Bigg(\widehat{p}^N(0) \left[\widehat{p}^N(0) - \frac{n_0}{n} \right]^2 \\ &+ \frac{\left[1 - \widehat{p}^N(0) \right]}{n-n_0} \left\{ \sum_{i=1}^{n-n_0} \left[\widehat{F}^X(X_{i:n-n_0}^+) - \frac{2i-1}{2(n-n_0)} \right]^2 + \frac{1}{12(n-n_0)} \right\} \Bigg), \end{aligned} \tag{26}$$

where $\widehat{F}^X(x) := F^X(x; \widehat{\vartheta})$ and $\widehat{p}^N(0) := p^N(0; \widehat{\vartheta}_1)$ is the parametric estimator of the probability that $N = 0$ under the null hypothesis.

3.2 Tests based on the Laplace transform

LT–based **GOF** tests are based on a distance between the **LT** $L^X(t) := L^X(t;\vartheta) = \mathbb{E}(e^{-tX}|\vartheta),\ t > 0$, corresponding to the null hypothesis, and its empirical counterpart, the empirical **LT**, given by

$$L_n^X(t) = \frac{1}{n}\sum_{i=1}^{n} e^{-tX_i}.$$

Typically, such a test statistic is expressed as an integrated distance between $L_n^X(t)$ and $L^X(t)$ involving a weight function $w(t) > 0,\ t \geq 0$. The main motivation of the **LT** approach lies in tractability of the **LT** of X, given by (4) where, recall, $G^N(t) := G^N(t;\vartheta_N) = \mathbb{E}(t^N|\vartheta_N)$ is the probability generating function of N and $L^U(t) := L^U(t;\vartheta_U)$ stands for the **LT** of the individual claim U.

3.2.1 L^2 dissimilarity measure

An obvious choice to measure the distance between the theoretical and empirical Laplace transform is the squared difference defined as

$$\mathrm{SE}_n(t) = \left[L_n^X(t) - \widehat{L}^X(t)\right]^2, \tag{27}$$

integrated against the weight function $w(t)$ as

$$S_{w,n} = n\int_0^\infty \mathrm{SE}_n(t)w(t)\mathrm{d}t, \tag{28}$$

where $\widehat{L}^X(t) = L^X(t;\widehat{\vartheta})$. Choosing an exponential weight function allows us to write the test statistic in (28) as

$$S_{w,n} = \frac{1}{n}\sum_{i,j=1}^{n}\frac{1}{X_i + X_j + \beta} - 2\sum_{i=1}^{n}\int_0^\infty \widehat{L}^X(t)e^{-(X_i+\beta)t}\mathrm{d}t + \int_0^\infty \left[\widehat{L}^X(t)\right]^2 e^{-\beta t}\mathrm{d}t. \tag{29}$$

Numerical integration is required for the evaluation. A classical work-around in **LT**-based **GOF** testing to avoid numerical integration is to define a dissimilarity measure relying on a differential equation .

The weight function figuring in (28) is often choosen to be exponential as $w(t) = e^{-\beta t}$ to ensure integrability. Tuning the parameter may also increase the power. The optimality of the value of this parameter is an open problem. The common approach is to run extensive Monte Carlo studies evaluating the power performance across a grid of values of the tuning parameter and suggests a compromise choice by selecting a value that fares well for a majority of the alternative choices. The discussion about the choice of the parameter is postponed to Section 4.

3.2.2 Dissimilarity measure based on a differential equation

If, under the null hypothesis, $N \sim \mathrm{KF}(a, b)$ then the **LT** of X satisfies a differential equation. Start by noting that

$$\mathrm{d}G^N(t) = \frac{a+b}{1-at}G^N(t), \tag{30}$$

where $\mathrm{d}f(t)$ denotes the first derivative of the function f with respect to t. Differentiating with respect to t on both sides of (4) yields

$$\mathrm{d}L^X(t) = \mathrm{d}L^U(t)\mathrm{d}G^N\left[L^U(t)\right], \ t \geq 0, \tag{31}$$

and reinserting (30) into (31) leads to the following differential equation

$$\mathrm{d}L^X(t)\left[1 - aL^U(t)\right] - (a+b)L^X(t)\mathrm{d}L^U(t) = 0. \tag{32}$$

Equation (32) motivates us to define a dissimilarity measure as

$$\mathrm{DE}_n(t) = \left[\mathrm{d}L_n^X(t)\frac{1-\widehat{a}\widehat{L}^U(t)}{\mathrm{d}\widehat{L}^U(t)} - (\widehat{a}+\widehat{b})L_n^X(t)\right]^2, \ t \geq 0, \tag{33}$$

where

$$\mathrm{d}L_n(t) = -\frac{1}{n}\sum_{i=1}^n X_i e^{-tX_i}, \ \widehat{L}^U(t) = L^U(t;\widehat{\vartheta_U}) \text{ and } \mathrm{d}\widehat{L}^U(t) = \mathrm{d}L^U(t;\widehat{\vartheta_U}).$$

The corresponding test statistic (analogous to (28)) is defined by

$$T_{n,w} = n\int_0^\infty \mathrm{DE}_n^2(t)w(t)\mathrm{d}t, \tag{34}$$

with rejection for large values of $T_{n,w}$. Straightforward integration then yields from (34),

$$T_{n,w} = \frac{1}{n}\sum_{i,j=1}^n X_iX_jK_w^{(2)}(X_i+X_j) + \frac{1}{n}\sum_{i,j=1}^n X_iK_w^{(1)}(X_i+X_j) + \frac{(a+b)^2}{n}\sum_{i,j=1}^n K_w^{(0)}(X_i+X_j), \tag{35}$$

where

$$K_w^{(k)}(x) = \int_0^\infty \left[\frac{1-a\widehat{L}^U(t)}{\mathrm{d}\widehat{L}^U(t)}\right]^k e^{-xt}w(t)\mathrm{d}t, \text{ for } k = 0,1,2. \tag{36}$$

The exponential weight function $w(t) = e^{-\beta t}$, $\beta > 0$, allows us to derive tractable formulas when the claim sizes distribution is gamma or inverse Gaussian as shown in the following example.

Example 3. *First note that with an exponential weight function we have that* $K_w^{(0)}(x) = (x+\beta)^{-1}$.

1. *Let U be gamma distributed $Gamma(r,\theta)$ with* ***LT*** *given by $L^U(t) = (1+\theta t)^{-r}$. We have that*

$$K_w^{(1)}(x) = a\frac{(x+\beta+\theta)}{r\theta(x+\beta)^2} - \frac{e^{(x+\beta)/\theta}\theta^r}{r(x+\beta)^{r+2}}\Gamma_u\left(r+2;\frac{x+\beta}{\theta}\right)$$

and

$$\begin{aligned} K_w^{(2)}(x) &= \frac{e^{(x+\beta)/\theta}\theta^{2r}}{r^2(x+\beta)^{2r+3}}\Gamma_u\left(2r+3;\frac{x+\beta}{\theta}\right) - 2a\frac{e^{(x+\beta)/\theta}\theta^r}{r^2(x+\beta)^{r+3}}\Gamma_u\left(r+3;\frac{x+\beta}{\theta}\right) \\ &+ a^2\frac{(x^2+2\theta x+2\theta^2)}{(r\theta)^2(x+\beta)^3}, \end{aligned}$$

where $\Gamma_u(r;x) = \int_x^{+\infty} y^{r-1}e^{-y}dy$ denotes the upper incomplete gamma function.

2. *Let U be inverse Gaussian distributed $IG(\mu,\phi)$ with* **LT** *given by* $L^U(t)=\exp\left(\frac{1-\sqrt{1+\phi\mu^2 t}}{\mu\phi}\right)$. *Denote by* $c=\sqrt{2}\frac{x+\beta-\mu}{\mu\sqrt{\phi(x+\beta)}}$ *and* $d=\sqrt{2}\frac{x+\beta-2\mu}{\mu\sqrt{\phi(x+\beta)}}$. *We have that*

$$
\begin{aligned}
K_w^{(1)}(x) &= a\frac{\sqrt{\phi}e^{\frac{(x+\beta)}{2\mu^2\phi}}}{(x+\beta)^{3/2}}\left\{\frac{e^{-\frac{(x+\beta)}{2\mu^2\phi}}\sqrt{x+\beta}}{\mu\sqrt{\phi}}+\frac{1}{\sqrt{2}}\mathrm{erfc}\left(\sqrt{2}\frac{x+\beta}{\mu^2\phi}\right)\right\}\\
&+ \frac{e^{c^2}}{\mu^2\sqrt{(x+\beta)\phi}}\left\{\frac{\mu^2\phi}{x+\beta}\left[\frac{c}{\sqrt{2}}e^{-c^2}+\frac{1}{\sqrt{2}}\mathrm{erfc}\,(c)\right]\right.\\
&+ \left.\frac{2\mu^2\sqrt{\phi}}{(x+\beta)^{3/2}}e^{-c^2}+\frac{\mu^2}{(x+\beta)^2\sqrt{2}}\mathrm{erfc}(c)\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
K_w^{(2)}(x) &= \frac{2^3 e^{d^2}}{(x+\beta)^3\sqrt{\phi(x+\beta)}}\left(\frac{1}{\sqrt{2}}\mathrm{erfc}(d)+\frac{3\sqrt{\phi(x+\beta)}}{2}e^{-d^2}\right.\\
&+ \left.\frac{3\phi(x+\beta)}{4}\left\{\frac{d}{\sqrt{2}}e^{-d^2}+\frac{1}{\sqrt{2}}\mathrm{erfc}(d)+\frac{\sqrt{\phi(x+\beta)}}{2}\frac{d^2+4}{2}e^{-d^2}\right\}\right)\\
&- 2ae^{c^2}\left(\frac{1}{\sqrt{2}}\mathrm{erfc}(c)+3\sqrt{\phi x}e^{-c^2}\right.\\
&+ \left.3\phi x\left\{\frac{c}{\sqrt{2}}e^{-c^2}+\frac{1}{\sqrt{2}}\mathrm{erfc}(c)\right\}+[\phi(x+\beta)]^{3/2}\frac{c^2+4}{2}e^{-c^2}\right)\\
&+ a^2\left(\frac{1}{\mu^2(x+\beta)}+\frac{2\phi}{(x+\beta)^2}\right),
\end{aligned}
$$

where $\mathrm{erfc}(x)=\frac{2}{\sqrt{\pi}}\int_x^{+\infty}e^{-t^2}\,dt$ *denotes the complementary error function.*

Although tractable, the formulae given above remain too involved to derive the asymptotic distribution of the test statistic. We resort to parametric bootstrap to compute the critical values. In this work, we restrict ourself to the case where the claim sizes are gamma and Inverse Gaussian distributed, because there are computable formulae available, but also since these distributions are amongst the most popular for claim–size modelling.

4 Simulation study

This section presents the result of a Monte-Carlo experiment designed to assess the power of the **GOF** procedures. In the first subsection, we investigate the impact of the choice of the parameter β in the weight function on the performance of the **LT** based **GOF** procedures. In the second subsection, the **DF** and **LT** based **GOF** tests are compared in terms of power. Parametric bootstrap resampling is used to approximate the distribution of the test statistic under the null hypothesis. This type of resampling has been set on a firm theoretical basis, see e.g., Stute et al. [35], Henze [18], and Genest and Rémillard [10], and is typically called upon when the asymptotic null distribution of any given test is too complicated to apply in practice. For the sake of completeness, the parametric bootstrap principle is recalled hereafter. Say we wish to assess the fit of an **iid** sample

$X_1, \ldots, X_n$ of aggregated claim data to a parametric model characterized by its **DF** $F^X(x, \vartheta)$. The parameter of the model is infered as $\widehat{\vartheta} = \widehat{\vartheta}(X_1, \ldots, X_n)$ and the test statistic $\mathrm{TS} \in \{\mathrm{CM}_n, \mathrm{KS}_n, S_{n,w}, T_{n,w}\}$ is computed. Bootstrap samples are then drawn from $F^X(x, \widehat{\vartheta})$. We compute the test statistic for each one of the samples and the critical values follows from quantile estimation. The steps of the parametric bootstrap routine are given in Algorithm 1 where $B \in \mathbb{N}$ denotes the number of bootstrap samples and $\alpha \in (0, 1)$ is the confidence level of the **GOF** test. In the sequel, we study the probability

Algorithm 1 Parametric bootstrap for goodness-of-fit test

compute $\widehat{\vartheta} := \widehat{\vartheta}(X_1, \ldots, X_n)$
compute TS.
for $k = 1 \to B$ **do**
  simulate $X^*_{k,1}, \ldots, X^*_{k,n}$ from $F^X(x, \widehat{\vartheta})$
  compute $\widehat{\vartheta}^*_k := \widehat{\vartheta}(X^*_{k,1}, \ldots, X^*_{k,n})$
  compute TS^*_k
end for
compute $TS^*_\alpha := \mathrm{Quantile}(\mathrm{TS}^*_1, \ldots, \mathrm{TS}^*_n; \alpha)$
if $\mathrm{TS} > TS^*_\alpha$ **then reject** H_0
else accept H_0
end if

of rejection of a sample generated by a known model F when the model tested is $F^X(x, \vartheta)$. It requires to generate $M \in \mathbb{N}$ samples $X_{k,1}, \ldots, X_{k,n},\ k = 1, \ldots, M$ drawn from F and apply Algorithm 1. The *warp-speed* strategy suggested by Giacomini et al. [12] allows us to reduce the running time required for our experiment. The idea is to generate only one bootstrap sample from F^X for each Monte Carlo sample simulated from F. The parametric bootstrap routine augmented by the warp-speed strategy is provided in Algorithm 2.

Algorithm 2 Rejection probability via parametric bootstrap and warp-speed method

for $k = 1 \to M$ **do**
  simulate $X_{k,1}, \ldots, X_{k,n}$ from F
  compute $\widehat{\vartheta}^*_k := \widehat{\vartheta}(X_{k,1}, \ldots, X_{k,n})$
  compute TS_k
  simulate $X^*_{k,1}, \ldots, X^*_{k,n}$ from $F^X(x, \widehat{\vartheta})$
  compute $\widehat{\vartheta}^*_k := \widehat{\vartheta}(X^*_{k,1}, \ldots, X^*_{k,n})$
  compute TS^*_k
end for
compute $\mathrm{TS}^*_\alpha := \mathrm{Quantile}(\mathrm{TS}^*_1, \ldots, \mathrm{TS}^*_M; \alpha)$
return The probability of rejection $\frac{1}{M}\sum_{k=1}^{M} \mathbb{I}_{\mathrm{TS}_k > \mathrm{TS}^*_\alpha}$

4.1 The value of the parameter in the weight function

The goal of this subsection is to determine the values of the parameter β that lead to acceptable results when using the **LT** based GOF procedures. It is well known from

Tauberian theorems, see for instance Feller [9, Chapter XIII.5] that the tail behavior of a probability distribution concentrated on the positive half-line is reflected by the behavior of the Laplace transform at 0 and vice-versa. It is especially true in our context, for instance, the $\mathrm{SE}_n(t)$ distance reduces to

$$\mathrm{SE}_n(t) \to \left[\frac{n_0}{n} - \widehat{p}^N(0)\right]^2, \text{ as } t \to +\infty.$$

Choosing small values of β leads to capture a difference in the atom of probability at 0, while choosing a large β allows to detect variations in the right tail. Note also that opting for the partial-MME method will make the $\mathrm{SE}_n(t)$ distance tends toward 0 for large values of t. The inference method will therefore affect the outcome of the tests.

The discussion that follows sheds some extra light on the role of the weight function, and more generally on these type of statistics. For small values of t, a Taylor expansion $e^{-tX_i} = 1 - (t/1!)X_i + (-t^2/2!)X_i^2 + \ldots$, in (27) leads to

$$\mathrm{SE}_n(t) \sim \sum_{k,l=1}^{+\infty} \frac{(-t)^{k+l}}{k!l!} \left[\mu_k - \mathbb{E}_0(X^k)\right] \left[\mu_l - \mathbb{E}_0(X^l)\right], \text{ as } t \to 0, \tag{37}$$

where $\mu_k = n^{-1}\sum_{k=1}^{n} X_i$, $k \geq 1$ denote the empirical moments of the sample $X_1, \ldots, X_n$ and $\mathbb{E}_0(X)$ is the expectation of the aggregated claim sizes distribution under H_0. Integrating (37) term-by-term against the exponential weight function $w(t) = e^{-\beta t}$ yields

$$S_{n,w} \approx n \sum_{k,l=1}^{+\infty} \binom{k+l+1}{k} \frac{(-1)^{k+l}}{\beta^{k+l+1}} \left[\mu_k - \mathbb{E}_0(X^k)\right] \left[\mu_l - \mathbb{E}_0(X^l)\right]. \tag{38}$$

Thus the weight function tunes how the difference between the empirical and theoretical moments enter the test statistic $S_{n,w}$. Namely, lowering the value of β allows one to take into account higher order moments. This analysis holds too for the $\mathrm{DE}_n(t)$ distance with

$$\mathrm{DE}_n(t) \sim \sum_{k,l=0}^{+\infty} \frac{(-t)^{k+l}}{k!l!} \left[\frac{1}{n}\sum_{i=1}^{n} Q_{k+1}(X_i)\right] \left[\frac{1}{n}\sum_{i=1}^{n} Q_{l+1}(X_i)\right], \text{ as } t \to 0, \tag{39}$$

where $(Q_k)_{k\geq 1}$ is a sequence of polynomials satisfying $\mathbb{E}_0\left[Q_k(X)\right] = 0$, for $k \geq 1$. The polynomial $Q_k(x)$ is of order k in x and its coefficients may be expressed in terms of the parameters of the model specified under H_0. For instance, if a compound Poisson-exponential $\mathrm{Pois}(\widehat{\lambda}) - \exp(\widehat{\theta})$ is assumed under H_0 then we have that

$$Q_1(x) = x - \widehat{\lambda}\widehat{\theta},\ Q_2(x) = x^2 - (\widehat{\lambda}+2)\widehat{\theta}x,\ Q_3(x) = x^3 - \widehat{\theta}(\widehat{\lambda}+4)x^2 + 2\widehat{\theta}^2 x,\ \ldots \tag{40}$$

Integrating (39) term-by-term against the exponential weight function $w(t) = e^{-\beta t}$ yields

$$T_{n,w} \approx n \sum_{k,l=1}^{+\infty} \binom{k+l+1}{k} \frac{(-1)^{k+l}}{\beta^{k+l+1}} \left[\frac{1}{n}\sum_{i=1}^{n} Q_{k+1}(X_i)\right] \left[\frac{1}{n}\sum_{i=1}^{n} Q_{l+1}(X_i)\right].$$

The value of β is calibrated to select the moments that will influence the test decision. We note also the first four order terms in both (38) and (40) vanish when choosing the method of moments estimator.

A Monte Carlo experiments is further conducted to gain insight on how to choose β to optimize the performance of the Laplace transform **GOF** procedures. We test the adequacy of a compound Poisson-exponential model $\text{Pois}(\widehat{\lambda})-\exp(\widehat{\theta})$ to data coming from a Poisson-gamma model $\text{Pois}(\lambda=1)-\text{gamma}(r,\theta=1)$. The probability of rejection is computed when varying the value the shape parameter $r\in\{0.5,0.75,1,2,4\}$ for both of the Laplace transform based procedures as well as the two available inference method (MME and partial-MME). We set the sample size to $n=100$ and use Algorithm 2 with $M=10,000$ Monte Carlo runs. Figure 1 displays the level (when $r=1$) of the test for β ranging from 10^{-13} to 10^{3}. The probability of rejection (expected to be around 5%) is too

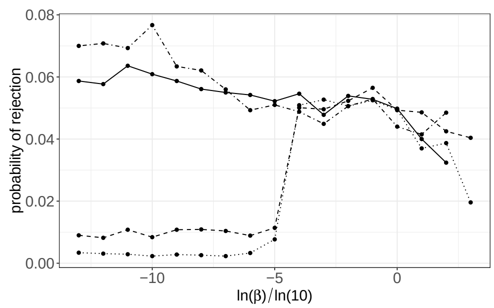


Figure 1: Level of the **LT** based **GOF** test depending on the distance between Laplace transform and inference techniques used: (dotted) $\text{SE}_n(t)$ and MME ; (dashed) $\text{SE}_n(t)$ and partial-MME; (solid) $\text{DE}_n(t)$ and MME ; (dotdash) $\text{DE}_n(t)$ and partial-MME.

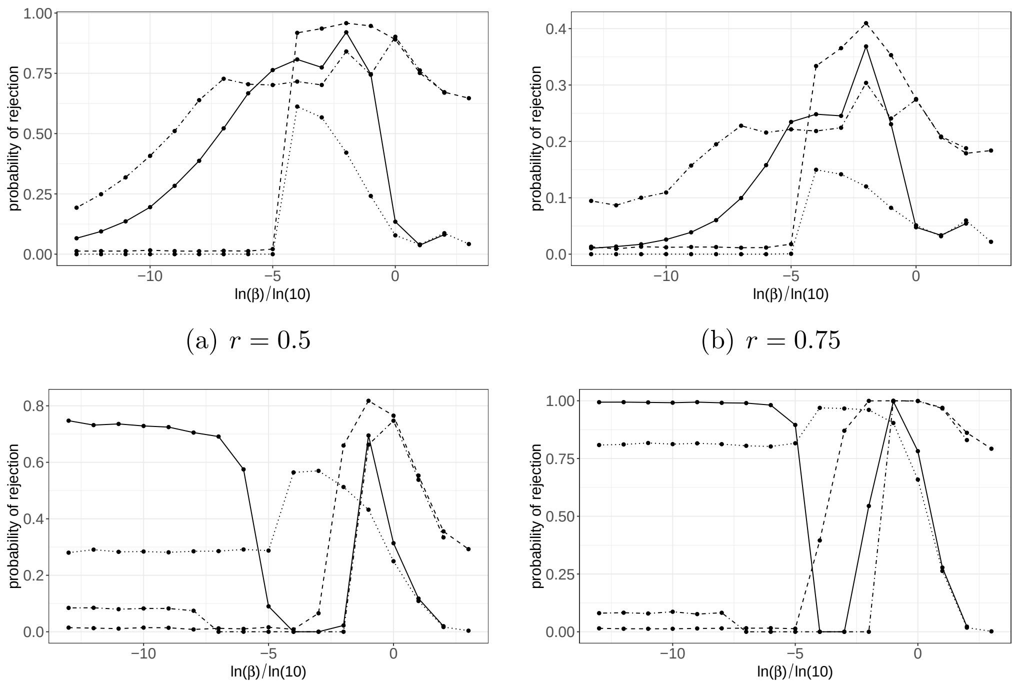
high when using the $\text{DE}_n(t)$ distance and too low when using the $\text{SE}_n(t)$ when $\beta<10^{-7}$. The sampling error on the parameter estimates might explain this fact as the variance is higher for large integrated distance. Figure 2 displays the power (when $r\neq 1$) of the tests. The power of the test always decreases with β which reflects that the distance between the Laplace transforms vanishes as t approaches 0. Two behaviors may be observed on Figure 2 depending on the value of r. When $r<1$, the rejection probability increases before reaching a maximum and decreasing. When $r>1$, the power admits a plateau before decreasing. It is common in the goodness-of-fit testing literature to opt for the values of β that fare well in the majority of the cases. We therefore retain the values $\beta=10^{-3},10^{-2}$ when using the $\text{SE}_n(t)$ distance and $\beta=0.1,1$ when using the $\text{DE}_n(t)$ to move on to the comparative study with the **DF** based **GOF** procedures.

4.2 Comparison of the GOF procedures

In this subsection, we compare the **GOF** procedures in terms of probability of rejection when the input samples differ from the model stated under the null hypothesis.

Test 1. *In this first test, we generate samples from a Poisson-Weibull model Pois*$(\lambda=1)-$*Weibull*$(r,\theta=1)$ *and test with our GOF method the adequacy of a Poisson-exponential, Poisson-gamma and Poisson-inverse Gaussian model. The Weibull distribution Weibull*(r,θ) *admits a probability density function given by*

$$f_U(x)=\frac{r}{\theta}\left(\frac{x}{\theta}\right)^{r-1}\exp\left[-\left(\frac{x}{\theta}\right)^{r}\right],\ \textit{for } x>0. \tag{41}$$



(a) $r = 0.5$ (b) $r = 0.75$

(c) $r = 2$ (d) $r = 4$

Figure 2: Power of the **LT** based **GOF** test for the different combinations of distances and inference techniques: (dotted) $\mathrm{SE}_n(t)$ and MME ; (dashed) $\mathrm{SE}_n(t)$ and partial-MME; (solid) $\mathrm{DE}_n(t)$ and MME ; (dotdash) $\mathrm{DE}_n(t)$ and partial-MME.

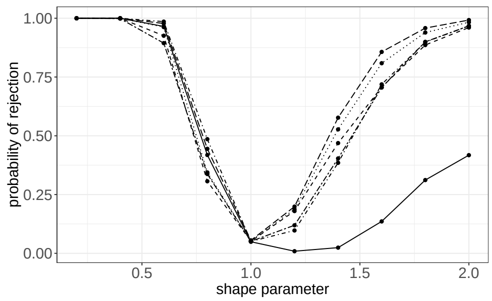
We set the sample size to 100 *and use the partial-MME to infer the parameters in the model specified in* H_0. *Figure 3 displays the powers computed via our parametric bootstrap routine with* 10,000 *Monte Carlo runs changing the shape parameter* r *in the claim sizes distribution. The Poisson-Weibull coincides with the Poisson-exponential and Poisson-gamma models when the shape parameter is set to* 1 *so the power tends toward* 5% *on Figures 3(a) and 3(b) as* r *gets closer to* 1. *The GOF procedures do well when testing for a Poisson-exponential with very high power as* r *gets farther from* 1*, see Figure 3(a). The results are a bit disapointing when testing for a Poisson-gamma distribution, the* ***DF*** *based procedures achieve greater power in this case, see Figure 3(b). The procedures associated to the* $SE_n(t)$ *distance outperform greatly the other methods when testing for a Poisson-inverse Gaussian model.*

Test 2. *In this second test, samples are generated from zero-modified Poisson-exponential* $zmpois(\lambda = 5, p_0) - \exp(\theta = 1)$ *and mixed Poisson-exponential* $mpois(p, \lambda_1 = 1, \lambda_2 = 5) - \exp(\theta = 1)$ *and we assess the adequacy of a Poisson-exponential model. The probability mass function of the zero-modified Poisson distribution* $zmpois(\lambda, p_0)$ *is given by*

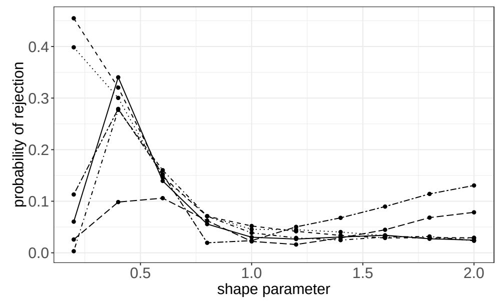
$$p^N(k) = \begin{cases} p_0, & \text{for } k = 0, \\ \frac{1-p_0}{1-e^{-\lambda}} \frac{\lambda^k e^{-\lambda}}{k!}, & \text{for } k \geq 1, \end{cases} \tag{42}$$

and the probability mass function of the mixed Poisson distribution $mpois(p, \lambda_1, \lambda_2)$ *is given by*

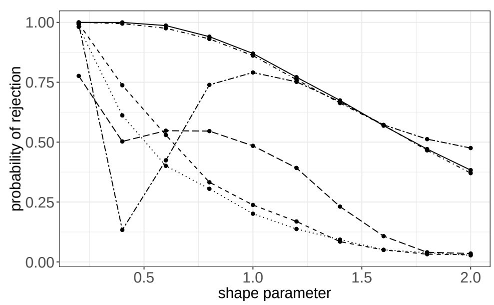
$$p^N(k) = p\frac{\lambda_1^k e^{-\lambda_1}}{k!} + (1-p)\frac{\lambda_2^k e^{-\lambda_2}}{k!}, \text{ for } k \geq 0. \tag{43}$$



(a) H_0: Poisson-exponential



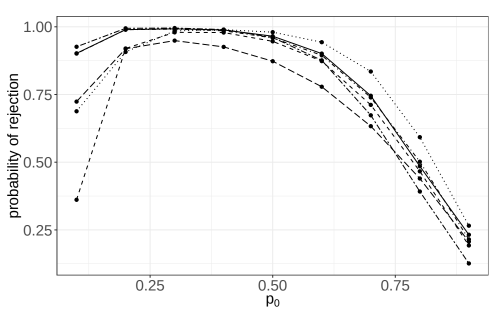
(b) H_0: Poisson-gamma



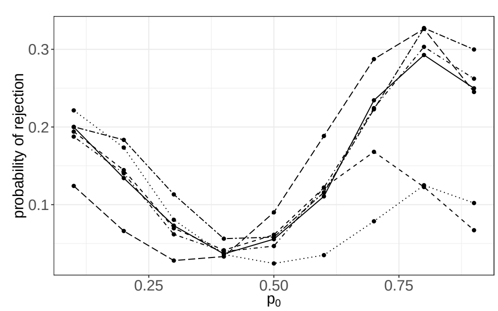
(c) H_0: Poisson-inverse Gaussian

Figure 3: Power of the various **GOF** tests: (dotted) Cramer-von Mises ; (dash) Kolmogorov-Smirnov; (solid) $\mathrm{SE}_n(t)$ distance and $\beta = 10^{-3}$; (dotdash) $\mathrm{SE}_n(t)$ distance and $\beta = 10^{-2}$; $\mathrm{DE}_n(t)$ distance and $\beta = 0.1$; $\mathrm{DE}_n(t)$ distance and $\beta = 1$.

The Poisson-exponential under H_0 is fitted using the MME based on samples of size 100*. Figure 4 displays the probability of rejection are computed via our parametric bootstrap routine with* 10,000 *Monte Carlo runs letting the parameter p vary in the alternative claim frequency distributions. The GOF procedures all detect reasonably well the modification*



(a) Samples drawn from a zmpois$(p, \lambda = 5)-$ exp$(\theta = 1)$ model



(b) Samples drawn from a mpois$(p, \lambda_1 = 1, \lambda_2 = 5) -$ exp$(\theta = 1)$ model

Figure 4: Power of the various **GOF** tests: (dotted) Cramer-von Mises ; (dash) Kolmogorov-Smirnov; (solid) $\mathrm{SE}_n(t)$ distance and $\beta = 10^{-3}$; (dotdash) $\mathrm{SE}_n(t)$ distance and $\beta = 10^{-2}$; (two dash) $\mathrm{DE}_n(t)$ distance and $\beta = 0.1$; (long dash) $\mathrm{DE}_n(t)$ distance and $\beta = 1$.

at 0*, see Figure 4(a). The Laplace transform based techniques outperform the* ***DF*** *based* *one in the mixed Poisson example, see Figure 4(b). The probability of rejection computed remain relatively small. The shape of the power is on the low side when* p *is close to* 0 *or* 1 *which makes sense since the Mixed Poisson distribution is then very close to a Poisson distribution. The downfall at* 0.5 *indicates that a balanced mixture of two Poisson random variables may be approximated well by one Poisson random variable.*

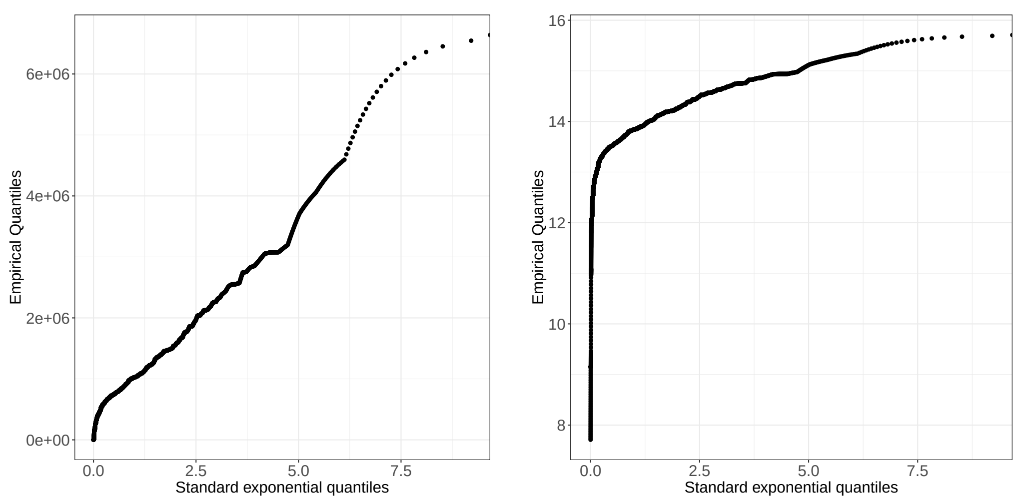
Other cases have been studied, the simulation data may be found in the online supplements [14]. The main conclusion is that none of the procedures stands out in all and every case. We suggest, in a practical situation, to apply all the procedures to see if they agree.

5 Application on a real insurance data set

We illustrate our inference and goodness-of-fit procedures on an actuarial data set called `itamtplcost` accessible from the `R` package `CASdatasets` and associated to the book of Charpentier [3]. This dataset contains losses (in excess of $500,000$ euros) of an Italian Motor-TPL company since 1997. It comprises two variables `Date` and `UltimateCost`, and 457 observations. Table 1 shows the first 5 observations of the dataset `itamtplcost`. Figure 5 displays the exponential and Pareto Quantile-Quantile plots. A linear relation-

Date	Claim size
08/01/1997	726,986.95
02/03/1997	1,222,682.37
18/03/1997	428,543.10
07/04/1997	258,786.06
11/04/1997	637,117.61

Table 1: The first five observations from the dataset `itamtplcost`.

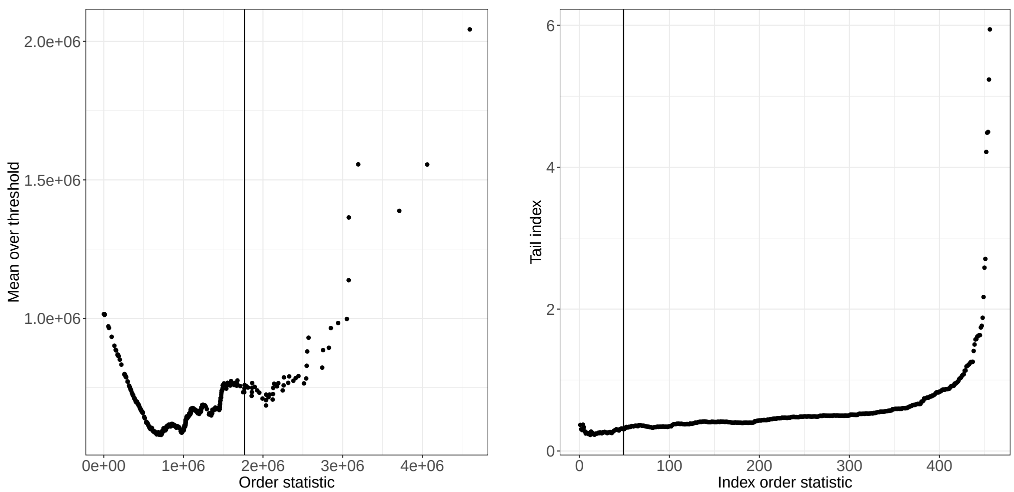


(a) Q-Q plot to test the adequacy to the exponential distribution.

(b) Q-Q plot to assess the adequacy to the Pareto distribution.

Figure 5: Quantile-Quantile Plots.

ship is observed between the lower order quantiles on Figure 5(a) and between the higher order quantiles on Figure 5(b). This, in turns, suggests the use of a splicing model with an exponential-type distribution to model the small claims and a Pareto-type distribution to fit the larger losses. The claim sizes distribution tested in the sequel are the exponential, gamma and inverse Gaussian distributions. Due to the heavy tail of the data, these distributions are not likely to result in a good fit. Hence, we decided to conduct the analysis over the whole data set and then on the smaller claims only. The small claims are defined on the basis of a threshold. The cut-off point between small and large claims is the upper order statistic that minimizes the asymptotic mean squared error of the Hill estimator, as it is a standard procedure in extreme value theory; see for instance the case study in the book by Beirlant et al. [1, Chapter 6]. Figure 6 displays the mean-excess plot and the Hill plot of the loss data. The threshold is set $1,766,751$ euros (corresponding to the vertical line on Figures 6(a) and 6(b)). A statistical summary over the whole dataset and the small claims is provided in Table 2. We note the swift decrease of variance when considering the small claims solely. The quartiles are stable while the mean in the small claims subset decreases to get closer to the median. Table 3 reports the method of moments estimators of the parameters of the exponential, gamma and inverse Gaussian distribution. Figure 7 displays the histograms of the claim sizes, on which are



(a) Mean excess plot

(b) Hill plot

Figure 6: Mean-Excess plot and Hill plot.

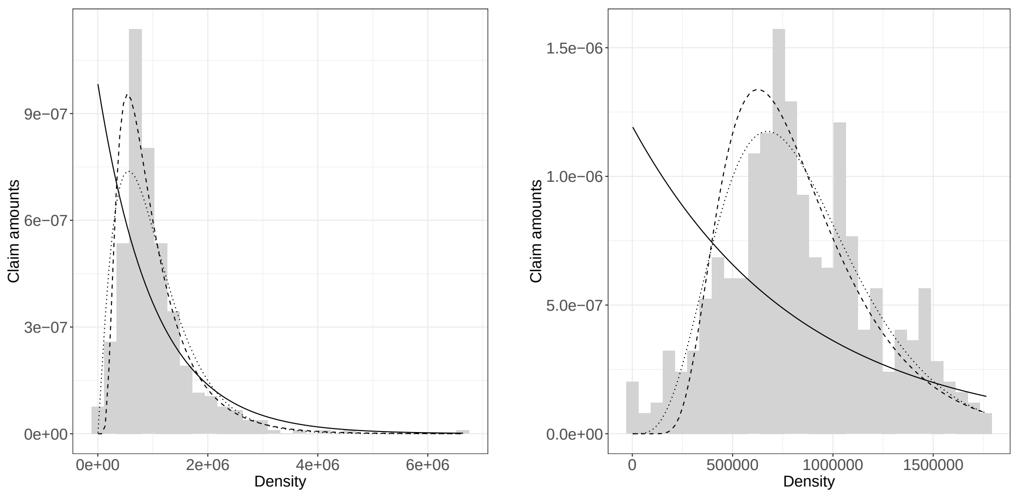
	Overall	*Small claims*
No of observations	457	408
E(U)	$1,015,352.15$	$837,255.78$
St.Dev(U)	$680,742.00$	$371,646.54$
Skewness(U)	2.55	0.22
Kurtosis(U)	12.65	-0.35
Minimum	$2,160.73$	$2,160.73$
Maximum	$6,639,499.57$	$1,764,900.46$
Q1	$627,718.53$	$595,464.36$
median	$844,010.92$	$790,992.29$
Q3	$1,224,316.09$	$1,066,904.00$

Table 2: Descriptive statistics of the claim data.

Model	parameters	*Overall*	*Small claims*
exponential	scale	$1,015,352.15$	$837,255.78$
gamma	shape	2.23	5.09
	scale	$455,404.21$	$164,564.55$
inverse Gaussian	mean	$1,015,352.15$	$837,255.78$
	dispersion	4.42E-07	2.35E-07

Table 3: Estimated parameters for the exponential and gamma distributions.

superposed the densities of the exponential, gamma and inverse Gaussian distributions with the parameters provided in Table 3. Table 4 reports the value of the AIC as well



(a) Histogram of the claim sizes. (b) Histogram of the small claim sizes.

Figure 7: Histograms of the data along with the density of the exponential (solid), gamma (dotted) and inverse Gaussian (dashed) distributions.

as the outcome of the **KS** and **CvM GOF** test for the exponential, gamma and inverse Gaussian distributions. On the basis of these results it seems that the gamma is the single distribution that provides a better fit for both of the datasets. We note how the rejection when using the **KS GOF** test is a close call when testing the gamma and inverse Gaussian model for the small claims. In order to apply our method, the original data

			KS		**CvM**	
Data	Model	AIC	Value	Critical value	Value	Critical value
Overall	exponential	$13,557.30$	0.24	0.05	7.79	0.22
	gamma	$13,403.02$	0.10	0.04	1.06	0.15
	inverse Gaussian	$15,498.63$	**0.08**	0.05	**0.58**	0.21
Small claims	exponential	$11,946.51$	**0.28**	0.05	**10.97**	0.22
	gamma	$11,765.60$	**0.05**	0.04	**0.24**	0.13
	inverse Gaussian	$13,859.68$	**0.07**	0.05	**0.62**	0.15

Table 4: Measure of the adequacy of the exponential and the gamma distributions.

needs to be processed so as to consider claim amounts aggregated monthly. For each time period, we collect the claim frequency as well as the sums of the incurred claims. Table 5 provides an overview of the processed data. Table 6 reports the estimated parameters of

Time period	claim frequency	total claim size
01/1997	1	726,986.95
02/1997	0	0.00
03/1997	2	1,651,225.47
04/1997	2	895,903.67
05/1997	0	0.00

Table 5: Monthly aggregated data.

the Poisson and geometric distributions accompanied by the Akaike information criterion and the χ^2 distance. The Poisson is better suited than the geometric distribution. Table

Data	model	parameters	χ^2 distance	AIC
Overall	Poisson	$\widehat{\lambda} = 2.38$	15.17	752.79
	geometric	$\widehat{p} = 0.70$	45.42	790.27
Small claims	Poisson	$\widehat{\lambda} = 1.93$	17.14	705.48
	geometric	$\widehat{p} = 0.66$	32.67	724.57

Table 6: Inference and measure of adequacy of the Poisson and geometric distribution over the claim frequency data.

7 gives the partial-MMEs for the Poisson-exponential, Poisson-gamma, Poisson-inverse Gaussian and geometric-exponential compound models. We note that these values are

model		*Overall*	*Small claims*
Poisson-exponential	$\widehat{\lambda} =$	2.42	2.52
	$\widehat{\theta} =$	999,750.14	630,059.684
Poisson-gamma	$\widehat{\lambda} =$	1.86	1.52
	$\widehat{r} =$	1.89	5.03
	$\widehat{\theta} =$	687,167.20	207,806.41
Poisson-inverse Gaussian	$\widehat{\lambda} =$	1.86	1.37
	$\widehat{\mu} =$	1,301,919.02	1,058,724.32
	$\widehat{\phi} =$	4.05E-07	1.90E-07
geometric-exponential	$\widehat{p} =$	0.84	0.78
	$\widehat{\theta} =$	447,546.23	445,020.10

Table 7: Inference and measure of adequacy of the Poisson and geometric distribution over the claim frequency data.

very different from the values estimated via the individual claim sizes and frequency data given in Tables 3 and 6. Tables 8, 9 and 10 provides a summary of the **GOF** procedures applied on the data. The critical values are computed using Algorithm 1 with $10,000$ bootstrap loops. The Poisson-gamma and Poisson-Inverse Gaussian models cannot be discarded according to all the methods. The exponential claim sizes is discarded for all the methods except when using the $\mathrm{SE}_n(t)$ distance in the Laplace transform based test, see Table 8.

Data	Model	$\beta = 10^{-3}$ Value	$\beta = 10^{-3}$ Critical value	$\beta = 10^{-2}$ Value	$\beta = 10^{-2}$ Critical value
overall	geometric-exp	2.13E+01	**2.44E+01**	2.12E+01	**2.45E+01**
	Poisson-exp	**3.62E+00**	2.61E+00	**3.21E+00**	1.41E+00
	Poisson-gamma	4.63E-02	**4.47E-01**	4.50E-02	**3.75E-01**
	Poisson-inverse Gaussian	2.53E-02	**1.64E-01**	2.27E-02	**1.58E-01**
small claims	geometric-exp	2.56E+01	**2.74E+01**	2.53E+01	**2.72E+01**
	Poisson-exp	**6.39E+00**	2.81E+00	**5.80E+00**	1.52E+00
	Poisson-gamma	5.30E-03	**4.66E-02**	5.22E-03	**4.90E-02**
	Poisson-inverse Gaussian	5.13E-02	**8.06E-02**	4.81E-02	**8.15E-02**

Table 8: Summary of the Laplace transform based **GOF** procedures using the $\mathrm{SE}_n(t)$ distance.

Data	Model	$\beta = 10^{-3}$ Value	$\beta = 10^{-3}$ Critical value	$\beta = 10^{-2}$ Value	$\beta = 10^{-2}$ Critical value
overall	geometric-exp	**1.97E+01**	9.49E+00	**1.01E+00**	1.95E-01
	Poisson-exp	**4.61E+01**	2.04E+01	**2.90E+00**	9.97E-01
	Poisson-gamma	1.43E+01	**7.89E+03**	1.45E-01	**2.75E+00**
	Poisson-inverse Gaussian	**1.24E+04**	1.90E+03	8.94E-03	**2.62E-01**
small claims	geometric-exp	**3.62E+01**	1.20E+01	**1.87E+00**	2.08E-01
	Poisson-exp	**8.94E+01**	2.19E+01	**5.49E+00**	7.72E-01
	Poisson-gamma	4.05E+04	**1.51E+07**	5.26E+01	**1.33E+06**
	Poisson-inverse Gaussian	1.72E+10	**7.53E+05**	3.34E-02	**1.41E+00**

Table 9: Summary of the Laplace transform based **GOF** procedures using the $\mathrm{DE}_n(t)$ distance.

Data	Model	**KS** Value	**KS** Critical value	**CvM** Value	**CvM** Critical value
overall	geometric-exp	**1.30E-01**	7.20E-02	**5.27E-03**	1.14E-03
	Poisson-exp	**9.88E-02**	7.08E-02	**2.13E-03**	1.12E-03
	Poisson-gamma	4.32E-02	**6.08E-02**	2.66E-04	**7.96E-04**
	Poisson-inverse Gaussian	3.16E-02	**6.05E-02**	2.23E-04	**7.74E-04**
small claims	geometric-exp	**1.69E-01**	6.91E-02	**9.29E-03**	1.16E-03
	Poisson-exp	**1.45E-01**	6.77E-02	**5.99E-03**	1.09E-03
	Poisson-gamma	3.17E-02	**7.20E-02**	2.01E-04	**1.19E-03**
	Poisson-inverse Gaussian	3.09E-02	**7.21E-02**	2.09E-04	**1.17E-03**

Table 10: Summary of the **DF** based **GOF** procedures.

6 Conclusion and perspectives

We develop **GOF** test procedures that simulaneously identify the components of an aggregate claim probability distribution, namely the claim frequency and the claim size distribution, when the only available data are the aggregated losses. Several tests were investigated, both classical as well as tests based on the **LT**. In either case the test criteria were tailored to the null hypothesis of compound distributions. The message drawn from a detailed Monte Carlo study was that all criteria respect the nominal level of the test and at the same time have reasonable power against some interesting alternatives, with the **LT**–based test having a certain edge in terms of power. The longstanding problem remains the choice of the parameter in the weight function. The real–data application also illustrates the potential of the suggested methods for practitioners.

There is scope for further improvement and research. First, a reliable inference procedure should be put together to efficiently estimate the parameters of any compound distribution based on the aggregated data. We proposed here a moment-based method, but we would like to review and compare likelihood-based and Bayesian methods to address this inferential problem. Second, we restricted ourselves to the Katz's family of counting distribution but we probably could consider generalization of these distributions such as presented in the works of Sundt [36, 37, 38].

More exotic count distributions may also be envisaged, but one will have to turn to numerical methods that inverse the Laplace transform. Many methods have been proposed relying on orthogonal polynomials as in Goffard and Laub [15] and Provost et al. [21], on Fourier series as in Rolski et al. [30], on exponential (Mnatsakanov and Sarkisian [26]) or fractional (Gzyl and Tagliani [16]) moments. These methods often lead to better accuracy that may, in turn, help the cause of **DF**-based **GOF** tests.

We close this article with a note on multivariate extension: A more general model for total claims is as follows. Let $Y = (X_1, X_2)$ be a pair of total claims $X_m = \sum_{k=1}^{N_m} U_{m,k},\ m = 1,2,$ where the two sequences of individual claims are mutually independent and for $m = 1,2$, each sequence $U_{m,k},\ k \geq 1$, is as in eqn. (1), but the claim frequencies are allowed to be dependent, say by following a bivariate Poisson distribution. Under this model the joint **LT** of Y is formulated as $L^Y(t_1,t_2) = g\left(L^{U_1}(t_1), L^{U_2}(t_2)\right)$ where $L^{U_m}(\cdot)$ denotes the **LT** of $U_{m,k}$, $m = 1,2$, and where

$$g(t_1,t_2) = e^{\lambda_1(t_1-1)+\lambda_2(t_2-1)+\lambda_{12}(t_1t_2-1)},$$

is the probability generating function of the bivariate Poisson distribution. Clearly **LT**–based methods are particularly suited for this model as well as for its obvious generalizations.

Acknowledgements

The work of Pierre-Olivier Goffard was partially funded by a CAE educational grant issued by the Society of Actuaries. This work was initiated while Pierre-Olivier Goffard and Simos Meintanis were visiting the department of Statistics and Applied Probability at UC Santa Barbara. The authors are grateful for the warm welcome they received there.

References

[1] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications.* John Wiley & Sons, 2006.

[2] P. Besbeas and B. J. T. Morgan. Efficient and robust estimation for the one-sided stable distribution of index 12. *Statistics & Probability Letters*, 66(3):251–257, 2004.

[3] A Charpentier. *Computational actuarial science with R.* CRC press, Boca Raton, 2014.

[4] M. Chaudhury. A review of the key issues in operational risk capital modeling. *The Journal of Operational Risk*, 5(3):37, 2010.

[5] V. Choulakian, R. Lockhart, and M.A. Stephens. Cramér-von Mises statistics for discrete distributions. *Canadian Journal of Statistics*, 22(1):125–137, 1994.

[6] W. Conover. A Kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association*, 67(339):591–596, 1972.

[7] R. B. D'Agostino and R. B. Stephens. *Goodness-of-Fit Techniques*, volume 68. Statistics, textbooks and monograph, 1986.

[8] D. S. Dimitrova, V.K. Kaishev, and S. Tan. Computing the Kolmogorov-Smirnov distribution when the underlying cdf is purely discrete, mixed or continuous. Available at http://openaccess.city.ac.uk/18541/, 2017.

[9] W. Feller. *An Introduction to Probability Theory and its Applications*, volume 2. John Wiley & Sons, New York, 2008.

[10] C. Genest and B. Rémillard. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 44(6):1096–1127, 2008.

[11] S. Ghosh and J. Beran. On estimating the cumulant generating function of linear processes. *Annals of the Institute of Statistical Mathematics*, 58(1):53–71, 2006.

[12] R. Giacomini, D. Politis, and H. White. A warp-speed method for conducting Monte Carlo experiments involving bootstrap. *Econometric Theory*, 29(3):567–589, 2013.

[13] L. Gleser. Exact power of goodness-of-fit tests of Kolmogorov type for discontinuous distributions. *Journal of the American Statistical Association*, 80(392):954–958, 1985.

[14] P.-O. Goffard. *Online accompaniment for "Goodness-of-fit tests for compound distributions with applications in insurance"*, 2019. Available at `https://github.com/LaGauffre/Online_accoompaniement_GOF_Test_Compound_Distribution`.

[15] P.-O. Goffard and P.J. Laub. Two numerical methods to evaluate stop-loss premiums. *arXiv preprint arXiv:1712.03468*, 2017. available at http://pierre-olivier.goffard.me.

[16] H. Gzyl and A. Tagliani. Determination of the distribution of total loss from the fractional moments of its exponential. *Applied Mathematics and Computation*, 219(4):2124 – 2133, 2012.

[17] N. Henze. A new flexible class of omnibus tests for exponentiality. *Communications in Statistics - Theory and Methods*, 22(1):115–133, 1992.

[18] N. Henze. Empirical-distribution-function goodness-of-fit tests for discrete models. *Canadian Journal of Statistics*, 24(1):81–93, 1996.

[19] N. Henze and B. Klar. Goodness-of-fit tests for the inverse Gaussian distribution based on the empirical laplace transform. *Annals of the Institute of Statistical Mathematics*, 54(2):425–444, 2002.

[20] N. Henze and S. G. Meintanis. Tests of fit for exponentiality based on the empirical Laplace transform. *Statistics*, 36(2):147–161, 2002.

[21] T. Jin, S. B. Provost, and J. Ren. Moment-based density approximations for aggregate losses. *Scandinavian Actuarial Journal*, 2016(3):216–245, 2016.

[22] L. Katz. *Unified treatment of a broad class of discrete probability distributions.* in: Classical and Contagious Discrete Distributions, Pergamon Press, Oxford, 1965.

[23] A. F. Laurence and B. J. T. Morgan. Selection of the transformation variable in the Laplace transform method of estimation. *Australian Journal of Statistics*, 29(2):113–127, 1987.

[24] R. Lockhart, J. Spinelli, and M. Stephens. Cramér-von Mises statistics for discrete distributions with unknown parameters. *Canadian Journal of Statistics*, 35(1):125–133, 2007.

[25] S. Meintanis and G. Iliopoulos. Tests of fit for the Rayleigh distribution based on the empirical Laplace transform. *Annals of the Institute of Statistical Mathematics*, 55(1):137–151, 2003.

[26] R. M. Mnatsakanov and K. Sarkisian. A note on recovering the distributions from exponential moments. *Applied Mathematics and Computation*, 219(16):8730 – 8737, 2013.

[27] G. Noether. Note on the Kolmogorov statistic in the discrete case. *Metrika*, 7(1):115–116, 1963.

[28] H. H. Panjer. Recursive evaluation of a family of compound distributions. *ASTIN Bulletin*, 12(1):22–26, 1981.

[29] N. De Pril. Moments of a class of compound distributions. *Scandinavian Actuarial Journal*, (2):117–120, 1986.

[30] T. Rolski, H. Schmidli, V. Schmidt, and J. Teugels. *Stochastic Processes for Insurance & Finance*, pages 147–204. John Wiley & Sons, Inc., 2008.

[31] P. Schmid. On the Kolmogorov and Smirnov limit theorems for discontinuous distribution functions. *The Annals of Mathematical Statistics*, 29(4):1011–1027, 1958.

[32] M. J. Slakter. A comparison of the Pearson chi-square and Kolmogorov goodness-of-fit tests with respect to validity. *Journal of the American Statistical Association*, 60(311):854–858, 1965.

[33] J. Spinelli. Testing fit for the grouped exponential distribution. *Canadian Journal of Statistics*, 29(3):451–458, 2001.

[34] J. Spinelli and M. Stephens. Cramér-von Mises tests of fit for the Poisson distribution. *Canadian Journal of Statistics*, 25(2):257–268, 1997.

[35] W. Stute, W. G. Manteiga, and M. P. Quindimil. Bootstrap based goodness-of-fit-tests. *Metrika*, 40(1):243–256, 1993.

[36] B. Sundt. On some extensions of Panjer's class of counting distributions. *Astin Bulletin*, 22(1):61–80, 1992.

[37] B. Sundt. Recursive evaluation of aggregate claims distributions. *Insurance: Mathematics and Economics*, 30(3):297 – 322, 2002.

[38] B. Sundt. Some recursions for moments of compound distributions. *Insurance: Mathematics and Economics*, 33(3):487 – 496, 2003.

[39] B. Sundt and W. S. Jewell. Further results on recursive evaluation of compound distributions. *ASTIN Bulletin*, 12(1):27–39, 1981.

[40] O. Thas. *Comparing Distributions*. Springer, 2010.

[41] J. E. Walsh. Bounded probability properties of Kolmogorov-Smirnov and similar statistics for discrete data. *Annals of the Institute of Statistical Mathematics*, 15(1):153–158, 1963.

[42] Q. Yao and B. J. T. Morgan. Empirical transform estimation for indexed stochastic models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):127–141, 1999.