

Goodness-of-fit procedures for compound distributions with an application to Insurance

P.-O. Goffard¹, S. Rao Jammalamadaka², and S. G. Meintanis^{3,4}

¹Univ Lyon, Université Lyon 1, LSAF EA2429, Institut de Science Financière et d'Assurances, Lyon, France

²Department of Statistics and Applied Probability, University of California, Santa Barbara, USA

³Department of Economics, National and Kapodistrian University of Athens, Athens, Greece

⁴Unit for Business Mathematics and Informatics, North-West University Potchefstroom, South Africa*

May 2, 2020

Abstract

Goodness-of-fit procedures are introduced for testing the validity of compound models. New tests that utilize the Laplace transform as well as classical tests based on the distribution function are investigated. A major area of application of compound laws is in insurance, to model total claims resulting from specific claim frequencies and individual claim sizes. Monte Carlo simulations are used to compare the different test procedures under a variety of specifications for these two components of total claims. Also a detailed application to an insurance data set is presented.

MSC 2010: 60G55, 60G40, 12E10.

Keywords: Compound distributions; goodness-of-fit tests; Katz family; Laplace transform.

1 Introduction

Consider the random variable,

$$X = \sum_{k=1}^N U_k, \quad (1)$$

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 continuous 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

*On sabbatical leave from the University of Athens

amount of claim associated with a non-life insurance portfolio over a fixed time period. The random variable N represents the claim frequency while the U_k 's represent the individual claim sizes. The final objective of practitioners is to identify the components of model (1), namely the distribution of the claim frequency N and of the individual claim sizes U . While there may be theoretical situations whereby a given compound **DF** can not be traced back uniquely to a single pair of component **DF**'s for N and U_k 's, we will nevertheless and for all practical purposes treat the correspondence between X and (N, U) within model (1) as being unique. In this connection we note that incomplete data situations such as these of having observations only on the total claim X 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 [5], the data available to assess operational risk are typically incomplete as banks often report aggregate losses, often discarding small losses. Such loss of information also occurs when data is merged after acquiring another banking operation.

Assume we observe aggregated claim sizes X_1, \dots, X_n , and we wish to assess the conformity of a given compound model in the composite situation whereby the **DF**'s involved depend on unknown parameters. Specifically we write $p^N := p^N(\cdot; \vartheta_N)$ and $F^U := F^U(\cdot; \vartheta_U)$, for the component **DF**'s 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 composite null hypothesis

$$H_0 : \text{The DF of } X \text{ in (1.1) is } F_0^X \equiv F_0^X(\cdot; \vartheta), \text{ for some } \vartheta \in \Theta, \quad (2)$$

where Θ denotes an appropriate parameter space.

Two types of nonparametric 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 [8], or Thas [36] 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 Cramér-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 [29], Walsh [38], Noether [26], Slakter [30], Conover [7], Gleser [15], and Dimitrova et al. [9] for the **KS** **GOF** test. Regarding the **CvM** criterion, the reader is referred to the works of Choulakian et al. [6], Henze [18], Spinelli and Stephens [32], Spinelli [31] 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 [39], Henze and Klar [19], Henze and Meintanis [20], Meintanis and Il-iopoulos [25], Besbeas and Morgan [3], and Ghosh and Beran [13].

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. Section 5 presents an application of our **GOF** procedures to a real data set from the insurance industry. We conclude with discussion in Section 6. Asymptotic results are contained in the Appendix.

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 continuous 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

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

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 (**LT**) of X , defined as $L^X(t) := \mathbb{E}(e^{-tX})$, may be expressed as

$$L^X(t) = G^N(L^U(t)), \quad t \geq 0, \quad (4)$$

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 (apart from its parametric form), but model the claim frequency via a counting distribution from the broad class viz. Katz family. Recall that the distribution of N belongs to the Katz family [22], written $N \sim \text{KF}(a, b)$, if the probability mass function satisfies the recursive equation

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

A characterization is given in Sundt and Jewell [34]. Prominent members of this family (5) include

1. The binomial $N \sim \text{Bin}(\alpha, p)$ ($\alpha \in \mathbb{N}$, $0 < p < 1$) which satisfies (5) with $a = -p/(1-p)$ and $b = (\alpha+1)p/(1-p)$.
2. The Poisson $N \sim \text{Pois}(\lambda)$ ($\lambda > 0$) 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

$$p^N(k) = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(k+1)} p^\alpha (1-p)^k, \text{ for } k \geq 0, \quad (6)$$

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

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

Throughout this paper, we will denote by $n_0 \leq n$ the number of zeros and by $X_1^+, \dots, X_{n-n_0}^+$ the non-zero values within the sample X_1, \dots, X_n .

2.2 Moments based estimation for aggregate claims

The Method of Moments 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 X may be expressed in terms of the moments of U via the recursive relations

$$(1-a)\mathbb{E}(X^{k+1}) = \sum_{i=0}^k \binom{k}{i} \left(\frac{k+1}{i+1} a + b \right) \mathbb{E}(U^{i+1}) \mathbb{E}(X^{k-i}), \text{ for } k \geq 0, \quad (7)$$

provided in De Pril [28, Equation 3]. Solving the system (7) for $\vartheta_N = (a, b)$ and ϑ_U yields the Method of Moments Estimators (MMEs).

Denote by $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ and $m_k = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^k$ 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 $N \sim \text{Pois}(\lambda)$, with claim size following an exponential distribution $U \sim \text{Exp}(\theta)$, with density

$$f^U(x) = \frac{1}{\theta} e^{-x/\theta}, \text{ for } x \geq 0. \quad (8)$$

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 $\mathbb{E}(U^2) = 2\theta^2$. Substituting in (7) and solving yields the parameter estimates

$$\hat{\theta} = \frac{m_2}{2\bar{X}} \text{ and } \hat{\lambda} = \frac{2\bar{X}^2}{m_2}. \quad (9)$$

2. (*geometric-exponential*): Assume that the claim sizes follow an exponential distribution, $U \sim \text{Exp}(\theta)$, and that the claim frequency is geometric $N \sim \text{Neg-Bin}(1, p)$.

Thus we have that $a = 1 - p$, $b = 0$, in the Katz family parametrization in (5). Substituting in (7) and solving yields the parameters estimates

$$\hat{\theta} = \frac{m_2 - \bar{X}^2}{2\bar{X}} \text{ and } \hat{p} = \frac{\hat{\theta}}{\hat{\theta} + \bar{X}}. \quad (10)$$

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

$$f^U(x) = \frac{e^{-x/\theta} x^{r-1}}{\theta^r \Gamma(r)}, \text{ for } x \geq 0. \quad (11)$$

Let the claim frequency be Poisson distributed $N \sim \text{Pois}(\lambda)$. We have that $a = 0$, $b = \lambda$, $\mathbb{E}(U) = r\theta$, $\mathbb{E}(U^2) = r(r+1)\theta^2$, and $\mathbb{E}(U^3) = r(r+1)(r+2)\theta^3$. Substituting in (7) and solving yields the parameters estimates

$$\hat{r} = \frac{2m_2^2 - m_3\bar{X}}{m_3\bar{X} - m_2^2}, \quad \hat{\theta} = \frac{m_2}{\bar{X}(\hat{r} + 1)}, \quad \text{and } \hat{\lambda} = \frac{\bar{X}}{\hat{r}\hat{\theta}}. \quad (12)$$

Remark 2.1. The estimates of the parameters in Examples (10) and (12) 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 as follows: whenever the data consists of one or more X_i that take the value zero i.e. $n_0 > 0$, consider adding to the system of equations (7), 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

$$p^N(0) = p^\alpha, \quad (13)$$

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

$$p^N(0) = e^{-\lambda}. \quad (14)$$

This probability that $N = 0$ is estimated by $\hat{p}^N(0) = n_0/n$. 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 $N \sim \text{Neg-Bin}(1, p)$, then $b = 0$ and $a = 1 - p$, and p is estimated via

$$\hat{p} = \frac{n_0}{n},$$

and if the claim sizes are exponentially distributed $U \sim \text{Exp}(\theta)$, then θ is estimated via

$$\hat{\theta} = \frac{\hat{p}\bar{X}}{1 - \hat{p}}.$$

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

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

$$\hat{\lambda} = -\log\left(\frac{n_0}{n}\right),$$

and if the claim sizes are gamma distributed $U \sim \text{Gamma}(r, \theta)$, then using (7), the Gamma parameters are estimated via

$$\hat{r} = \frac{\bar{X}^2}{\hat{\lambda}m_2 - \bar{X}^2} \text{ and } \hat{\theta} = \frac{\bar{X}}{\hat{\lambda}\hat{r}}. \quad (15)$$

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 $N \sim \text{Pois}(\lambda)$, then $a = 0$ and $b = \lambda$, and λ is estimated via

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

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

$$f^U(x) = \left(\frac{1}{2\pi x^3 \phi}\right)^{1/2} \exp\left[-\frac{(x - \mu)^2}{2\mu^2 \phi x}\right], \quad x > 0. \quad (16)$$

Substituting in (7) and solving yields the parameter estimates

$$\hat{\mu} = \frac{\bar{X}}{\hat{\lambda}} \text{ and } \hat{\phi} = \frac{\hat{\lambda}m_2 - \bar{X}^2}{\bar{X}}. \quad (17)$$

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** $F_0^X(x; \vartheta) = \mathbb{P}_0(X \leq x | \vartheta)$, $x \in \mathbb{R}$, to its empirical counterpart, the empirical **DF**, defined by

$$F_n^X(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x). \quad (18)$$

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

$$F_n^{X|N>0}(x) = \frac{1}{n - n_0} \sum_{i=1}^{n - n_0} \mathbb{I}(X_i^+ \leq x). \quad (19)$$

We estimate the population **DF** as $\hat{F}_0^X(x) := F_0^X(x; \hat{\vartheta})$ where $\hat{\vartheta} = \hat{\vartheta}(X_1, \dots, X_n)$ is some asymptotically efficient estimator. As we assumed that the distribution of N belongs to the Katz family, the population **DF** may be approximated via the so-called Panjer algorithm, see [27] for more details.

3.1.1 *Kolmogorov-Smirnov* test for compound distributions

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

$$\text{KS} (F_0^X, F_n^X) = \sqrt{n} \sup_{x \in (0, \infty)} |F_0^X(x) - F_n^X(x)| := \sqrt{n} \sup_{x \in (0, \infty)} D_n^{\text{KS}}(x). \quad (20)$$

Denote by $X_{1:n-n_0}^+, \dots, X_{n-n_0:n-n_0}^+$, the order statistics associated to the sample $X_1^+, \dots, X_{n-n_0}^+$, and define the intervals $I_i = [X_{i:n-n_0}^+, X_{i+1:n-n_0}^+)$ for $i = 0, \dots, n - n_0$, with the convention $X_{0:n-n_0} = 0$ and $X_{n-n_0+1:n-n_0} = \infty$. Then for $x \in I_i$,

$$D_n^{\text{KS}}(x) = \left| \frac{n_0 + i}{n} - F_0^X(x) \right|, \quad i = 0, \dots, n - n_0,$$

so that

$$\sup_{x \in I_i} D_n^{\text{KS}}(x) = \max \left[\frac{n_0 + i}{n} - F_0^X(X_{i:n-n_0}^+), F_0^X(X_{i+1:n-n_0}^+) - \frac{n_0 + i}{n} \right].$$

Considering successively the intervals I_i , we estimate the **KS** distance (20) by

$$\text{KS}_n = \max(D_-, D_+),$$

where

$$D_- = \max_{0 \leq i \leq n-n_0} \frac{n_0 + i}{n} - \hat{F}_0^X(X_{i:n-n_0}^+), \text{ and } D_+ = \max_{0 \leq i \leq n-n_0} \hat{F}_0^X(X_{i+1:n-n_0}^+) - \frac{n_0 + i}{n}.$$

3.1.2 *Cramér-von Mises* test for compound distributions

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

$$\text{CM} (F_0^X, F_n^X) = n \int_0^{+\infty} [F_0^X(x) - F_n^X(x)]^2 dF_0^X(x). \quad (21)$$

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

$$dF_0^X(x) = p_0^N(0)\delta_0(x) + [1 - p^N(0)] dF_0^{X|N>0}(x), \quad x \geq 0, \quad (22)$$

where $\delta_0(x)$ denotes the Dirac measure at 0. Reinserting (22) into the integral (21) yields

$$\begin{aligned} \text{CM} (F_0^X, F_n^X) &= n \left\{ p^N(0) \left[p^N(0) - \frac{n_0}{n} \right]^2 \right. \\ &\quad \left. + [1 - p^N(0)] \int_0^{+\infty} [F_0^X(x) - F_n^X(x)]^2 dF_0^{X|N>0}(x) \right\}. \end{aligned} \quad (23)$$

Because the non-negative data points $X_1^+, \dots, X_{n-n_0}^+$ are assumed to be distributed as $F_0^{X|N>0}$ under the null hypothesis, we can write the integral in (23) as

$$\begin{aligned} &\left[p^N(0) - \frac{n_0}{n} \right] \left[\frac{2}{n} \sum_{i=1}^{n-n_0} F^{X|N>0}(X_i^+) - \frac{n - n_0}{n} \right] \\ &+ \int_0^{+\infty} \left\{ [1 - p^N(0)] F_0^{X|N>0}(x) - \frac{n - n_0}{n} F_n^{X|N>0}(x) \right\}^2 dF^{X|N>0}(x), \end{aligned} \quad (24)$$

By expanding the square and applying the change of variable $u = F^{X|N>0}(x)$, the integral (24) may be rewritten as

$$\begin{aligned} & \frac{[1 - p^N(0)]^2}{3} - [1 - p^N(0)] \frac{n - n_0}{n} \frac{1}{n - n_0} \sum_{i=0}^{n-n_0} [1 - F_0^{X|N>0}(X_i^+)^2] \\ & + \left(\frac{n - n_0}{n} \right) \left[n - n_0 + \sum_{i=0}^{n-n_0} (1 - 2i) F_0^{X|N>0}(X_{i:n}^+) \right], \end{aligned} \quad (25)$$

Finally combining (23), (24) and (25) allows to estimate the **CvM** statistics as

$$\begin{aligned} \text{CM}_n &= n \left(\hat{p}_0^N(0) \left[\hat{p}_0^N(0) - \frac{n_0}{n} \right]^2 \right. \\ &+ [1 - \hat{p}_0^N(0)] \left\{ \left[\hat{p}_0^N(0) - \frac{n_0}{n} \right] \left[\frac{2}{n} \sum_{i=1} \hat{F}_0^{X|N>0}(X_{i:n}^+) - \frac{n - n_0}{n} \right] \right. \\ &+ \frac{[1 - \hat{p}_0^N(0)]^2}{3} - [1 - \hat{p}_0^N(0)] \frac{n - n_0}{n} \frac{1}{n - n_0} \sum_{i=0}^{n-n_0} [1 - \hat{F}_0^{X|N>0}(X_i^+)^2] \\ &\left. \left. + \left(\frac{n - n_0}{n} \right) \left[n - n_0 + \sum_{i=0}^{n-n_0} (1 - 2i) \hat{F}_0^{X|N>0}(X_{i:n}^+) \right] \right\} \right), \end{aligned} \quad (26)$$

where $\hat{F}_0^{X|N>0}(x) := F_0^{X|N>0}(x; \hat{\vartheta})$ and $\hat{p}_0^N(0) := p_0^N(0; \hat{\vartheta}_N)$ 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** $L_0^X(t; \vartheta) := \mathbb{E}_0(e^{-tX}|\vartheta)$, $t > 0$, (we often write $L_0^X(t)$ for simplicity) 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 $L_0^X(t)$ involving a weight function $w(t) > 0$, $t \geq 0$. The main motivation of the **LT** approach lies in tractability of the **LT** $L_0^X(\cdot)$, given by (4) where, recall, $G^N(t) := \mathbb{E}(t^N)$ is the probability generating function of N and $L^U(t) := \mathbb{E}(e^{-tU})$ stands for the **LT** of the individual claim U .

3.2.1 L^2 dissimilarity measure

An obvious choice to consider the discrepancy between the theoretical and empirical Laplace transform

$$\text{SE}_n(t) = L_n^X(t) - \hat{L}_0^X(t), \quad (27)$$

and integrate $\text{SE}^2(\cdot)$ against the weight function $w(t)$ as

$$S_{n,w} = n \int_0^\infty \text{SE}_n^2(t) w(t) dt, \quad (28)$$

where $\widehat{L}_0^X(t) = L_0^X(t; \widehat{\vartheta})$. Choosing the exponential weight function $w(t) = e^{-\beta t}$, $\beta > 0$, allows us to write the test statistic in (28) as

$$S_{n,\beta} = \frac{1}{n} \sum_{i,j=1}^n \frac{1}{X_i + X_j + \beta} - 2 \sum_{i=1}^n \int_0^\infty \widehat{L}_0^X(t) e^{-(X_i + \beta)t} dt + n \int_0^\infty \left[\widehat{L}_0^X(t) \right]^2 e^{-\beta t} dt. \quad (29)$$

Depending on the hypothesized **LT**, numerical integration may be 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 which we discuss next. The issue of the choice of the weight parameter β is postponed to Section 4.

3.2.2 Dissimilarity measure based on a differential equation

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

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

where $df(t)$ denotes the first derivative of the function f with respect to t . Differentiating with respect to t on both sides of (4) yields

$$dL^X(t) = dL^U(t) dG^N[L^U(t)], \quad t \geq 0, \quad (31)$$

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

$$dL^X(t) [1 - aL^U(t)] - (a+b)L^X(t) dL^U(t) = 0. \quad (32)$$

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

$$\text{DE}_n(t) = dL_n^X(t) \frac{1 - \widehat{a}\widehat{L}_0^U(t)}{d\widehat{L}_0^U(t)} - (\widehat{a} + \widehat{b})L_n^X(t), \quad t \geq 0, \quad (33)$$

where

$$dL_n^X(t) = -\frac{1}{n} \sum_{i=1}^n X_i e^{-tX_i}, \quad \widehat{L}_0^U(t) = L_0^U(t; \widehat{\vartheta}_U) \text{ and } d\widehat{L}_0^U(t) = dL_0^U(t; \widehat{\vartheta}_U).$$

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

$$T_{n,w} = n \int_0^\infty \text{DE}_n^2(t) w(t) dt, \quad (34)$$

with rejection for large values of $T_{n,w}$.

Letting $w(t) = e^{-\beta t}$, $\beta > 0$ and by straightforward computations we have from (34),

$$T_{n,\beta} = \frac{1}{n} \sum_{i,j=1}^n X_i X_j K_\beta^{(2)}(X_i + X_j) + \frac{1}{n} \sum_{i,j=1}^n X_i K_\beta^{(1)}(X_i + X_j) + \frac{(\widehat{a} + \widehat{b})^2}{n} \sum_{i,j=1}^n K_\beta^{(0)}(X_i + X_j), \quad (35)$$

where

$$K_{\beta}^{(k)}(x) = \int_0^{\infty} \left[\frac{1 - a\widehat{L}_0^U(t)}{d\widehat{L}_0^U(t)} \right]^k e^{-(x+\beta)t} w(t) dt, \text{ for } k = 0, 1, 2. \quad (36)$$

The exponential weight function $e^{-\beta t}$, $\beta > 0$, 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 $K_{\beta}^{(0)}(x) = (x + \beta)^{-1}$.

1. Let U be gamma distributed $\text{Gamma}(r, \theta)$ with \mathbf{LT} given by $L_0^U(t) = (1 + \theta t)^{-r}$. We have that

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

and

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

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 $\text{IG}(\mu, \phi)$ with \mathbf{LT} given by $L_0^U(t) = \exp \left(\frac{1 - \sqrt{1 + \phi\mu^2 t}}{\mu\phi} \right)$. 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

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

and

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

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

Asymptotic results including the limit distribution of the **LT**-based test statistic $S_{n,w}$ under the null hypothesis are given in the Appendix. This distribution, as well as the limit distributions corresponding to the other **LT**- or **DF**- based tests considered here, is extremely complicated. Therefore, in the next section we resort to resampling techniques in order to obtain critical values and actually carry out the tests.

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. [33], Henze [18], and Genest and Rémillard [12], 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, \dots, X_n of aggregated claim data to a parametric model characterized by its **DF** $F_0^X(x, \vartheta)$. The parameter of the model is inferred as $\hat{\vartheta} = \hat{\vartheta}(X_1, \dots, X_n)$ and the test statistic $TS \in \{CM_n, KS_n, S_{n,w}, T_{n,w}\}$ is computed. Bootstrap samples are then drawn from $F_0^X(x, \hat{\vartheta})$. We compute the test statistic for each one of the samples and the critical values follow 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.

Algorithm 1 Parametric bootstrap for goodness-of-fit test

```

1: compute  $\hat{\vartheta} := \hat{\vartheta}(X_1, \dots, X_n)$ 
2: compute TS.
3: for  $k = 1 \rightarrow B$  do
4:   simulate  $X_{k,1}^*, \dots, X_{k,n}^*$  from  $F_0^X(x, \hat{\vartheta})$ 
5:   compute  $\hat{\vartheta}_k^* := \hat{\vartheta}(X_{k,1}^*, \dots, X_{k,n}^*)$ 
6:   compute  $TS_k^*$ 
7: end for
8: compute  $TS_\alpha^* := \text{Quantile}(TS_1^*, \dots, TS_B^*; \alpha)$ 
9: if  $TS > TS_\alpha^*$  then reject  $H_0$ 
10: else accept  $H_0$ 
11: end if

```

In the sequel, we study the probability of rejection of a sample generated by a fixed model F^X when the model tested is $F_0^X(x, \vartheta)$. It requires to generate $M \in \mathbb{N}$ samples $X_{k,1}, \dots, X_{k,n}$, $k = 1, \dots, M$ drawn from F^X and apply Algorithm 1. The *warp-speed* strategy suggested by Giacomini et al. [14] allows us to reduce the running time required for our experiment. The idea is to generate only one bootstrap sample from F_0^X for each Monte Carlo sample simulated from F^X . 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

```

1: for  $k = 1 \rightarrow M$  do
2:   simulate  $X_{k,1}, \dots, X_{k,n}$  from  $F^X$ 
3:   compute  $\hat{\vartheta}_k^* := \hat{\vartheta}(X_{k,1}, \dots, X_{k,n})$ 
4:   compute  $\text{TS}_k$ 
5:   simulate  $X_{k,1}^*, \dots, X_{k,n}^*$  from  $F_0^X(x, \hat{\vartheta})$ 
6:   compute  $\hat{\vartheta}_k^* := \hat{\vartheta}(X_{k,1}^*, \dots, X_{k,n}^*)$ 
7:   compute  $\text{TS}_k^*$ 
8: end for
9: compute  $\text{TS}_\alpha^* := \text{Quantile}(\text{TS}_1^*, \dots, \text{TS}_M^*; \alpha)$ 
10: return The probability of rejection  $M^{-1} \sum_{k=1}^M \mathbb{I}_{\text{TS}_k > \text{TS}_\alpha^*}$ 

```

4.1 Investigation of the impact of the weight function

The goal of this subsection is to investigate the role of the weight parameter β in **LT**-based procedures and prompt the discussion for the choice of good values of this parameter. It is well known from Tauberian theorems, see for instance Feller [11, 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, that for $\text{SE}_n(t)$ we have

$$\text{SE}_n^2(t) \rightarrow \left[\frac{n_0}{n} - \hat{p}_0^N(0) \right]^2, \text{ as } t \rightarrow +\infty.$$

Choosing small values of β leads to capturing 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 $\text{SE}_n(t)$ distance tend toward 0 for large values of t .

For further scrutiny on the role of the weight function and the weight parameter β , we consider the discrepancy $\text{SE}(t)$ in (27) and take the Taylor series expansion of the exponential function figuring in the empirical Laplace transform therein. This leads to

$$\text{SE}_n^2(t) = \sum_{k,\ell=1}^{\infty} \frac{(-t)^{k+\ell}}{k!\ell!} \left[\hat{\mu}_k - \hat{\mathbb{E}}_0(X^k) \right] \left[\hat{\mu}_\ell - \hat{\mathbb{E}}_0(X^\ell) \right], \quad (37)$$

where $\hat{\mu}_k = n^{-1} \sum_{i=1}^n X_i^k$, $k \geq 1$, denote the empirical moments of the sample X_1, \dots, X_n and $\hat{\mathbb{E}}_0(X^k) := \mathbb{E}_0(X^k; \hat{\vartheta})$ is the estimate of the corresponding expectation of the aggregated claim size under H_0 obtained by replacing ϑ by $\hat{\vartheta}$. Integrating (37) term-by-term

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

$$S_{n,\beta} = n \sum_{k,\ell=1}^{\infty} \binom{k+\ell}{k} \frac{(-1)^{k+\ell}}{\beta^{k+\ell+1}} \left[\hat{\mu}_k - \hat{\mathbb{E}}_0(X^k) \right] \left[\hat{\mu}_\ell - \hat{\mathbb{E}}_0(X^\ell) \right]. \quad (38)$$

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 $\text{DE}_n(t)$ distance with

$$\text{DE}_n^2(t) = \sum_{k,\ell=1}^{+\infty} \frac{(-t)^{k+\ell}}{k!\ell!} \left[\frac{1}{n} \sum_{i=1}^n Q_{k+1}(X_i) \right] \left[\frac{1}{n} \sum_{i=1}^n Q_{\ell+1}(X_i) \right], \quad (39)$$

where $(Q_k)_{k \geq 1}$ is a sequence of polynomials satisfying $\mathbb{E}_0[Q_k(X)] = 0$, 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 $\text{Pois}(\lambda) - \exp(\theta)$ is assumed under H_0 then we have that

$$Q_1(x) = x - \hat{\lambda}\hat{\theta}, \quad Q_2(x) = x^2 - (\hat{\lambda} + 2)\hat{\theta}x, \quad Q_3(x) = x^3 - \hat{\theta}(\hat{\lambda} + 4)x^2 + 2\hat{\theta}^2x, \quad \dots \quad (40)$$

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

$$T_{n,\beta} = n \sum_{k,\ell=1}^{\infty} \binom{k+\ell}{k} \frac{(-1)^{k+\ell}}{\beta^{k+\ell+1}} \left[\frac{1}{n} \sum_{i=1}^n Q_{k+1}(X_i) \right] \left[\frac{1}{n} \sum_{i=1}^n Q_{\ell+1}(X_i) \right]. \quad (41)$$

The value of β is calibrated to select the moments that will influence the test decision. We note also that with moment estimation, $\hat{\mathbb{E}}_0(X^k) = \hat{\mu}_k$, $k = 1, 2$, so that the corresponding terms in (38) and (41) vanish when choosing the Method of Moments estimator.

A Monte Carlo experiment 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 $\text{Pois}(\lambda) - \text{Exp}(\theta)$ 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 methods (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 high when using the $T_{n,\beta}(t)$ statistic and too low when using the $S_{n,\beta}$ 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. In this connection we note that there

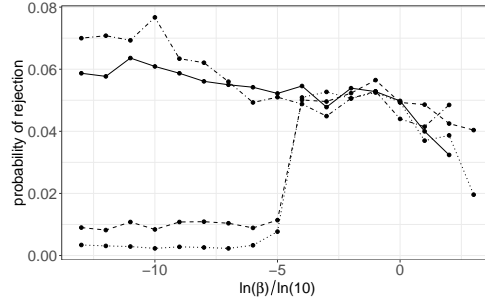


Figure 1: Level of the **LT** based **GOF** test depending on the distance between Laplace transform and inference techniques used: (dotted) $S_{n,\beta}$ and MME ; (dashed) $S_{n,\beta}$ and partial-MME; (solid) $T_{n,\beta}$ and MME ; (dotdash) $T_{n,\beta}$ and partial-MME.

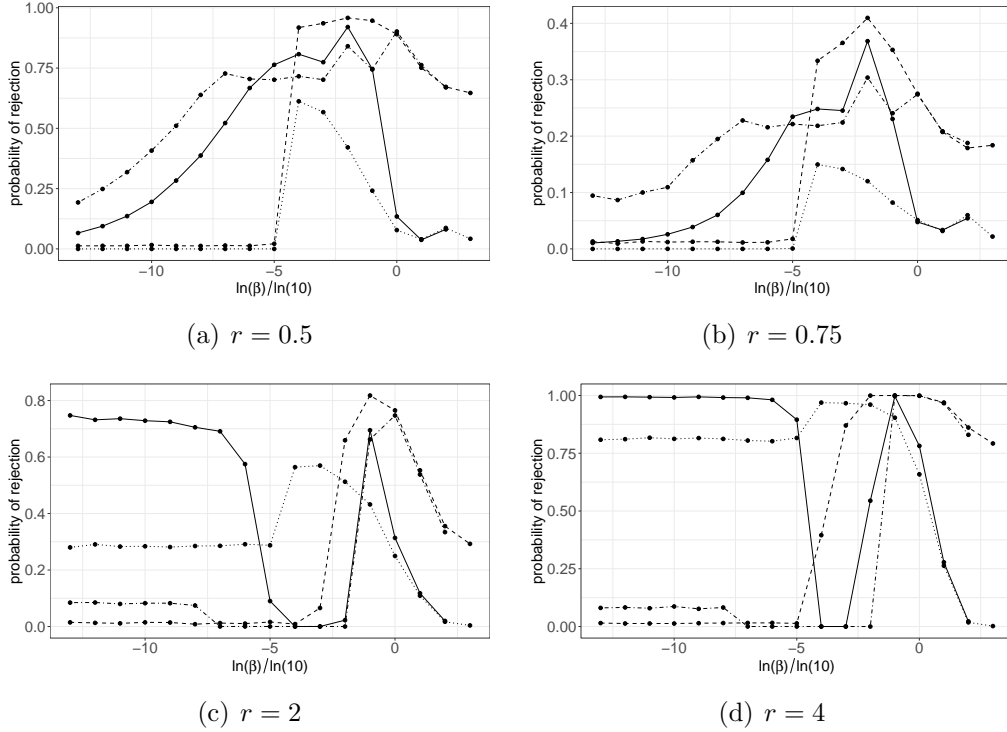


Figure 2: Power of the **LT** based **GOF** test for the different combinations of distances and inference techniques: (dotted) $S_{n,\beta}$ and MME ; (dashed) $S_{n,\beta}$ and partial-MME; (solid) $T_{n,\beta}$ and MME ; (dotdash) $T_{n,\beta}$ and partial-MME.

exist data-driven selection methods to determine proper values of the parameter β in **LT**-based goodness-of-fit tests, such as those proposed recently by Allison and Santana [1] and Tenreiro [35]. However, given the computing time associated with the test statistic we decided not to implement such search methods here, and employ in the comparative study below the values $\beta = 10^{-3}, 10^{-2}$ when using the $S_{n,\beta}$ and $\beta = 0.1, 1$ when using the $T_{n,\beta}$, which performed well in our preliminary study.

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 $\text{Pois}(\lambda = 1) - \text{Weibull}(r, \theta = 1)$ and test with our GOF method the adequacy of a Poisson-exponential, Poisson-gamma and Poisson-inverse Gaussian model. The Weibull distribution $\text{Weibull}(r, \theta)$ admits a probability density function given by

$$f^U(x) = \frac{r}{\theta} \left(\frac{x}{\theta}\right)^{r-1} \exp\left[-\left(\frac{x}{\theta}\right)^r\right], \text{ for } x > 0. \quad (42)$$

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-

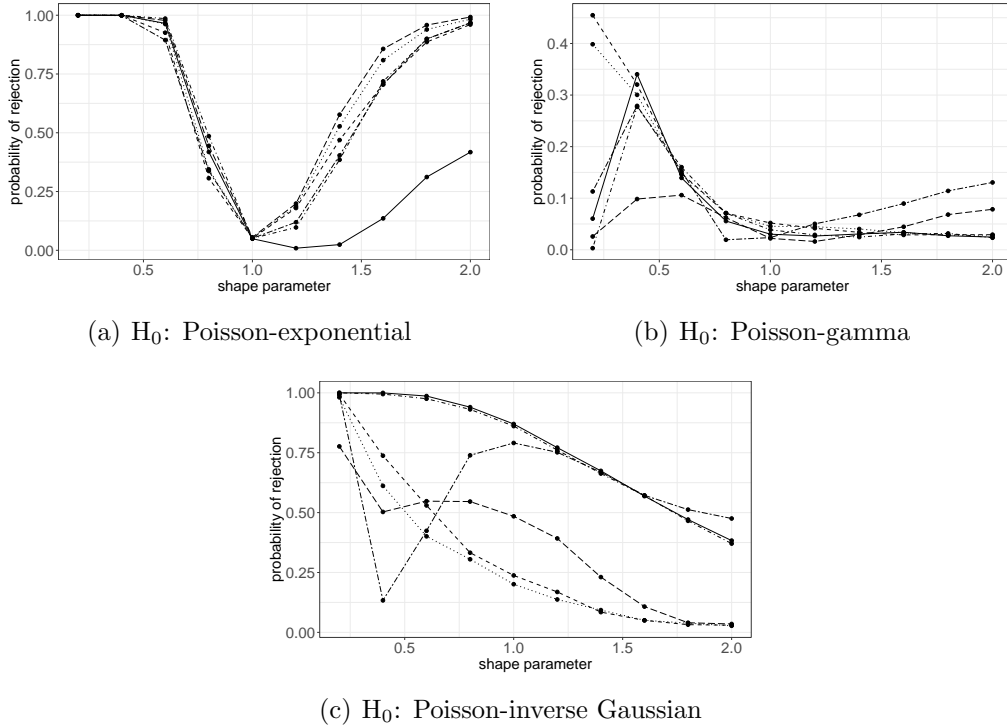


Figure 3: Power of the various **GOF** tests in **Test 1**: (dotted) Cramér-von Mises ; (dash) Kolmogorov-Smirnov; (solid) $S_{n,\beta}, \beta = 10^{-3}$; (dotdash) $S_{n,\beta}, \beta = 10^{-2}$; (two dash) $T_{n,\beta}, \beta = 0.1$; (long dash) $T_{n,\beta}, \beta = 1$.

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 disappointing 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 $S_{n,\beta}$ 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

$$p^N(k) = \begin{cases} p_0, & \text{for } k = 0, \\ \frac{1-p_0}{1-e^{-\lambda}} \frac{\lambda^k e^{-\lambda}}{k!}, & \text{for } k \geq 1, \end{cases} \quad (43)$$

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

$$p^N(k) = p \frac{\lambda_1^k e^{-\lambda_1}}{k!} + (1-p) \frac{\lambda_2^k e^{-\lambda_2}}{k!}, \text{ for } k \geq 0. \quad (44)$$

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

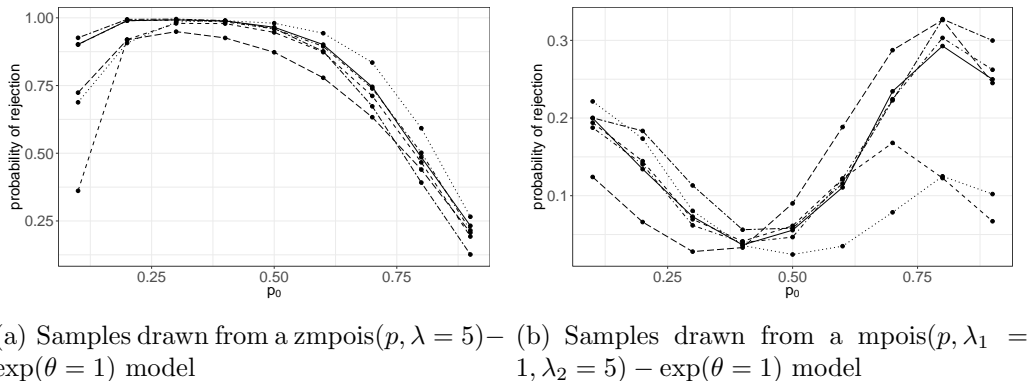


Figure 4: Power of the various **GOF** tests in **Test 2**: (dotted) Cramer-von Mises ; (dash) Kolmogorov-Smirnov; (solid) $S_{n,\beta}, \beta = 10^{-3}$; (dotdash) $S_{n,\beta}, \beta = 10^{-2}$; (two dash) $T_{n,\beta}, \beta = 0.1$; (long dash) $T_{n,\beta}, \beta = 1$.

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 [16]. The main conclusion is that none of the procedures stands out in all and every case. This conclusion is corroborated by analytic methods which lead to the conclusion that any given goodness-of-fit test has nontrivial power only towards a given direction away from the null hypothesis; see Janssen [21], and Escanciano [10]. Therefore we suggest, in a practical situation, to apply all the procedures to see if they lead to the same conclusion.

5 An application to insurance data

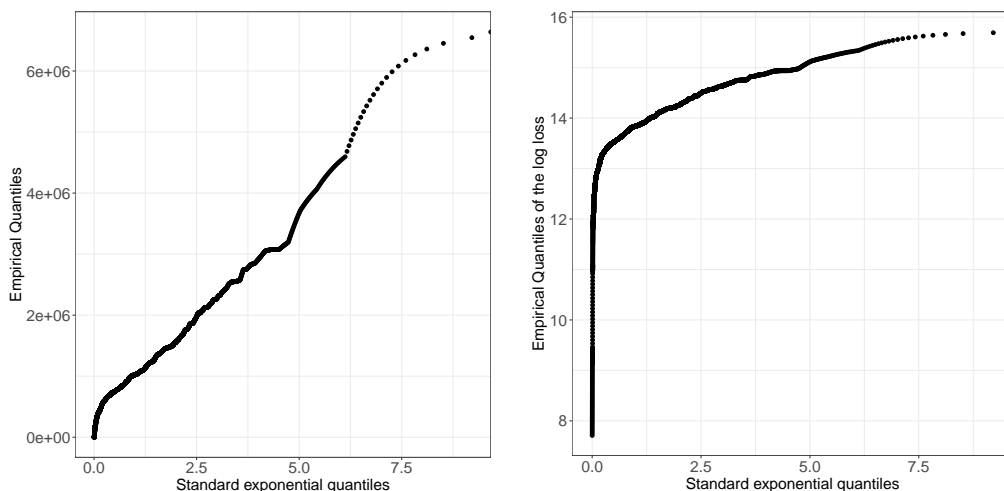
We illustrate our inference and goodness-of-fit procedures on an actuarial data set called `itamtplcost` accessible from the R package `CASdatasets` (see also the book of Charpentier [4]). This dataset contains losses (in excess of 500,000 euros) of an Italian Motor-TPL company since 1997. It comprises two variables `Date` and `UltimateCost`, and 457 observations. Table 1 shows the first 5 observations of the dataset `itamtplcost`.

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`.

We start by looking at the individual claim data before applying our methods to the monthly aggregated data, in the hope that they lead to similar inference and conclusions.

Figure 5 displays the exponential and Pareto Quantile-Quantile plots. A linear relation-



(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 turn, 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. [2, 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

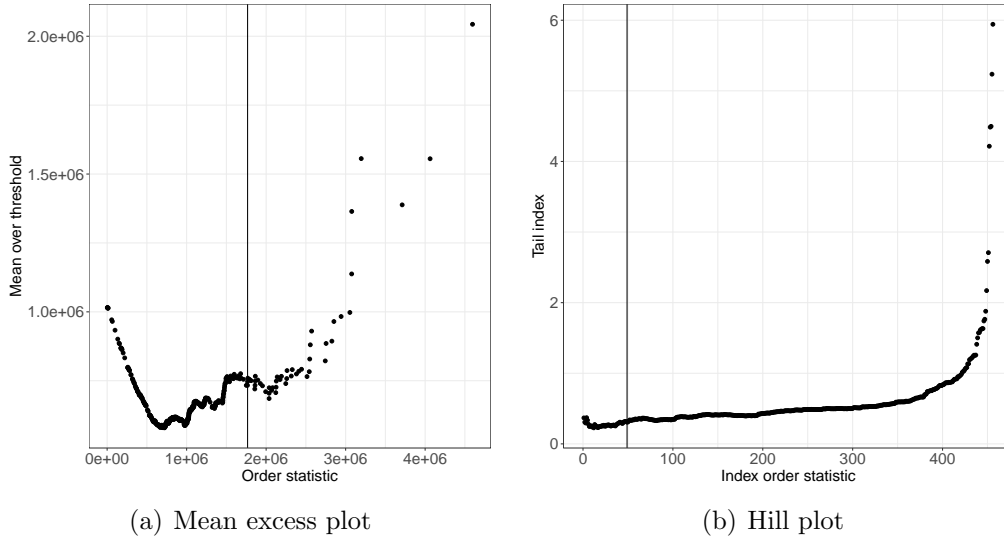


Figure 6: Mean-Excess plot and Hill plot.

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

	<i>Overall</i>	<i>Small claims</i>
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.

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

Model	parameters	<i>Overall</i>	<i>Small claims</i>
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.

7 displays the histograms of the claim sizes, on which are 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 as the outcome of the **KS** and

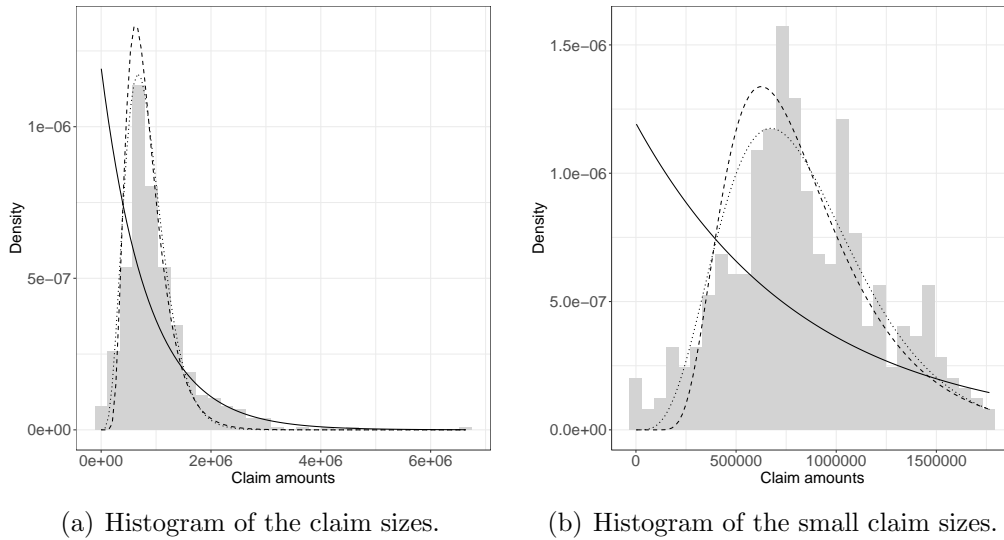


Figure 7: Histograms of the data along with the density of the exponential (solid), gamma (dotted) and inverse Gaussian (dashed) distributions.

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 needs to be processed so as to

Data	Model	AIC	KS		CvM	
			Value	Critical value	Value	Critical value
<i>Overall</i>	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
<i>Small claims</i>	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.

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 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 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 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

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.

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

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

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

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

Data	Model	$\beta = 10^{-3}$		$\beta = 10^{-2}$	
		Value	Critical value	Value	Critical value
<i>overall</i>	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
<i>small claims</i>	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 $S_{n,\beta}$ statistic.

10,000 bootstrap loops. The Poisson-gamma and Poisson-Inverse Gaussian models cannot be discarded according to all the methods. The exponential claim sizes are discarded for all the methods except when using the $S_{n,\beta}$ test statistic based on the Laplace transform, see Table 8.

Data	Model	$\beta = 10^{-3}$		$\beta = 10^{-2}$	
		Value	Critical value	Value	Critical value
<i>overall</i>	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
<i>small claims</i>	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 $T_{n,\beta}$ statistic.

Data	Model	KS		CvM	
		Value	Critical value	Value	Critical value
<i>overall</i>	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
<i>small claims</i>	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

Several goodness-of-fit tests for compound distributions were investigated, both classical as well as tests based on the Laplace transform. In either case the test criteria were tailored to specific versions of the null hypothesis that are popular in applications. The message drawn from a detailed Monte Carlo study is that all criteria respect the nominal level of the test and at the same time have reasonable power against some interesting alternatives, with the Laplace transform based test having a certain edge in terms of power. The real-data application shows the potential of the suggested methods for practitioners in order to also 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.

There are clearly several possible directions in which the current results can be extended. For instance, one can explore alternate estimation methods for the parameters, go outside the Katz family for counting models, or consider situations where multivariate data is available on the compound variable X .

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.

7 Appendix: Consistency and limit null distribution

In this section, we discuss the consistency and limiting distribution of the **LT**-based test statistics under the null hypothesis H_0 . We focus our attention on the test criterion $S_{n,w}$ defined in (28), but note that similar results may be obtained for the test statistic $T_{n,w}$. We begin with the consistency of the test based on $S_{n,w}$ under the following assumptions:

(A.1) Under alternatives $\widehat{\vartheta} \rightarrow \widetilde{\vartheta}$, a.s., as $n \rightarrow \infty$, for some $\widetilde{\vartheta} \in \Theta$.

(A.2) The **LT** $L_0^X(\cdot; \vartheta)$ is continuous in ϑ .

(A.3) The weight function satisfies,

(i) $w(t) > 0, \forall t > 0$, except for a set of measure zero,

(ii) $\int_0^\infty w(t)dt < \infty$.

Theorem 1. *Let $L^X(t)$ denote the **LT** of X . Then if assumptions (A.1)–(A.3) are satisfied, we have under alternatives*

$$\frac{S_{n,w}}{n} \rightarrow \int_0^\infty \left(L^X(t) - L_0^X(t; \widetilde{\vartheta}) \right)^2 w(t)dt, \quad (45)$$

a.s., as $n \rightarrow \infty$.

Proof. Clearly the strong consistency of the empirical Laplace transform and the continuity of $L_0(\cdot; \vartheta)$ imply that

$$SE_n \rightarrow L^X(t) - L_0^X(t; \widetilde{\vartheta}),$$

a.s., as $n \rightarrow \infty$. Then since $SE_n^2(t) \leq 4$, the result follows by Lebesgue's dominated convergence theorem. \square

The right-hand side of (45) is positive unless $L^X(t) = L_0^X(t; \widetilde{\vartheta})$, for all $t > 0$. However, by the uniqueness of the **LT**, the last identity holds true only under the null hypothesis H_0 , which implies the strong consistency of the test that rejects H_0 for large values of $S_{n,w}$.

We continue with the limit distribution of the test statistic $S_{n,w}$ under the null hypothesis H_0 . For simplicity we assume that ϑ is a scalar parameter. To this end assume that

(A.4) The estimator $\widehat{\vartheta} := \widehat{\vartheta}_n$ satisfies the Bahadur representation

$$\widehat{\vartheta}_n - \vartheta_0 = \frac{1}{n} \sum_{j=1}^n \ell(X_j; \vartheta_0) + o_P(1)$$

where ϑ_0 denotes the true value of ϑ and $\ell(\cdot; \cdot)$ are such that $\mathbb{E}(\ell(X; \vartheta_0)) = 0$ and $\mathbb{E}(\ell^2(X; \vartheta_0)) < \infty$.

(A.5) The **LT** $L_0^X(t; \vartheta)$ is twice differentiable with respect to ϑ with a continuous second derivative in the neighborhood of the true value ϑ_0 .

(A.6) The weight function is such that

$$\int_0^\infty \left(\frac{\partial L_0^X(t; \vartheta_0)}{\partial \vartheta} \right)^2 w(t) dt < \infty,$$

and

$$\int_0^\infty \left(\frac{\partial^2 L_0^X(t; \vartheta)}{\partial \vartheta^2} \right)^2_{\vartheta=\vartheta^*} w(t) dt < \infty,$$

for all ϑ^* in a neighborhood of ϑ_0 .

Theorem 2. *Under assumptions (A.4)–(A.6) we have*

$$Z_n(t) = \sqrt{n}(L_n^X(t) - L_0^X(t; \hat{\vartheta}_n)) \xrightarrow{\mathcal{L}} Z(t),$$

as $n \rightarrow \infty$, where $Z(t)$ is the zero-mean Gaussian process with covariance kernel $K(s, t; \vartheta_0) = \mathbb{E}(Y(t; \vartheta_0)Y(s; \vartheta_0))$ with

$$Y(t; \vartheta) = e^{-tX} - L_0^X(t; \vartheta) - \frac{\partial L_0^X(t; \vartheta)}{\partial \vartheta} \ell(X; \vartheta).$$

The covariance kernel is specified by

$$\begin{aligned} K(s, t; \vartheta) &= L_0^X(t+s, \vartheta) - L_0^X(t; \vartheta)L_0^X(s; \vartheta) \\ &- \frac{\partial L_0^X(s; \vartheta)}{\partial \vartheta} \mathbb{E}(e^{-tX} \ell(X; \vartheta)) - \frac{\partial L_0^X(t; \vartheta)}{\partial \vartheta} \mathbb{E}(e^{-sX} \ell(X; \vartheta)) \\ &+ \frac{\partial L_0^X(s; \vartheta)}{\partial \vartheta} \frac{\partial L_0^X(t; \vartheta)}{\partial \vartheta} \mathbb{E}(\ell^2(X; \vartheta)). \end{aligned}$$

Proof. In the proof we will write $Z_n^{(1)} \approx Z_n^{(2)}$ if the two random processes $(Z_n^{(k)}(t), k = 1, 2)$, satisfy $Z_n^{(1)}(t) - Z_n^{(2)}(t) = \varepsilon_n(t)$, and the remainder $\varepsilon_n(t)$ is such that it has no effect on the limit null distribution of the test statistic $S_{n,w}$.

With this understanding using assumption (A.5) and the second part of (A.6), a two-term Taylor expansion yields

$$Z_n \approx Z_n^*,$$

where

$$Z_n^*(t) = \sqrt{n}(L_n^X(t) - L_0^X(t; \vartheta_0)) - \sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \frac{\partial L_0^X(t; \vartheta_0)}{\partial \vartheta}. \quad (46)$$

In turn using assumption (A.4) and the first part of (A.6) in (46) leads to

$$Z_n^* \approx Z_n^{**},$$

where

$$Z_n^{**}(t) = \sqrt{n}(L_n^X(t) - L_0^X(t; \vartheta_0)) - \frac{\partial L_0^X(t; \vartheta_0)}{\partial \vartheta} \frac{1}{\sqrt{n}} \sum_{j=1}^n \ell(X_j; \vartheta_0). \quad (47)$$

The result now follows by applying the Central Limit Theorem in Hilbert spaces, (see e.g. Vaart and Wellner [37]) on the process $Z_n^{**}(t)$ given in (47). \square

Now the limit distribution of the test statistic follows from Theorem 2 and the Continuous Mapping theorem. Specifically we have

$$S_{n,w} = \int_0^\infty Z_n^2(t)w(t)dt \xrightarrow{\mathcal{L}} \int_0^\infty Z^2(t)w(t)dt := Z_w$$

where $Z(t)$ is the process defined in Theorem 2. The distribution of Z_w is the same as that of $\sum_{j=1}^\infty \lambda_j N_j^2$, where $\lambda_1, \lambda_2, \dots$, are the eigenvalues corresponding to the integral operator

$$Ag(s) = \int_0^\infty K(s, t)g(t)w(t)dt,$$

i.e. the solutions of the equation $Ag(s) = \lambda g(s)$, and where N_j , $j \geq 1$, are **iid** standard normal variates.

Remark 7.1. *The assumptions (A.1)–(A.3) made in order to prove consistency, as well as those pertaining to the limit null distribution, (A.4)–(A.6), are standard in the context of testing goodness-of-fit based on the empirical **LT**; see for instance Henze [17], Henze and Klar [19], and Henze and Meintanis [20],*

References

- [1] J.S. Allison and L. Santana. On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *Journal of Statistical Computation and Simulation*, 85(16):3276–3288, 2015.
- [2] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, 2006.
- [3] P. Besbeas and B. J. T. Morgan. Efficient and robust estimation for the one-sided stable distribution of index $1/2$. *Statistics & Probability Letters*, 66(3):251–257, 2004.
- [4] A Charpentier. *Computational actuarial science with R*. CRC press, Boca Raton, 2014.
- [5] M. Chaudhury. A review of the key issues in operational risk capital modeling. *The Journal of Operational Risk*, 5(3):37, 2010.
- [6] V. Choulakian, R. Lockhart, and M.A. Stephens. Cramér-von Mises statistics for discrete distributions. *Canadian Journal of Statistics*, 22(1):125–137, 1994.
- [7] W. Conover. A Kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association*, 67(339):591–596, 1972.
- [8] R. B. D’Agostino and R. B. Stephens. *Goodness-of-Fit Techniques*, volume 68. Statistics, textbooks and monograph, 1986.
- [9] 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.

- [10] J.C. Escanciano. On the lack of power of omnibus specification tests. *Econometric Theory*, 25:162–194, 2009.
- [11] W. Feller. *An Introduction to Probability Theory and its Applications*, volume 2. John Wiley & Sons, New York, 2008.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [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] A. Janssen. Global power function of goodness-of-fit tests. *Annals of Statistics*, 28:239–253, 2000.
- [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] G. Noether. Note on the Kolmogorov statistic in the discrete case. *Metrika*, 7(1):115–116, 1963.
- [27] H. H. Panjer. Recursive evaluation of a family of compound distributions. *ASTIN Bulletin*, 12(1):22–26, 1981.
- [28] N. De Pril. Moments of a class of compound distributions. *Scandinavian Actuarial Journal*, (2):117–120, 1986.
- [29] P. Schmid. On the Kolmogorov and Smirnov limit theorems for discontinuous distribution functions. *The Annals of Mathematical Statistics*, 29(4):1011–1027, 1958.
- [30] 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.
- [31] J. Spinelli. Testing fit for the grouped exponential distribution. *Canadian Journal of Statistics*, 29(3):451–458, 2001.
- [32] J. Spinelli and M. Stephens. Cramér-von Mises tests of fit for the Poisson distribution. *Canadian Journal of Statistics*, 25(2):257–268, 1997.
- [33] W. Stute, W. G. Manteiga, and M. P. Quindimil. Bootstrap based goodness-of-fit-tests. *Metrika*, 40(1):243–256, 1993.
- [34] B. Sundt and W. S. Jewell. Further results on recursive evaluation of compound distributions. *ASTIN Bulletin*, 12(1):27–39, 1981.
- [35] C. Tenreiro. On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests. *Journal of Statistical Computation and Simulation*, 89(10):1780–1797, 2019.
- [36] O. Thas. *Comparing Distributions*. Springer, 2010.
- [37] A. W. van der Vaart and J. A. Wellner. *Weak Convergence*. Springer New York, New York, NY, 1996.
- [38] 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.
- [39] 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.