Gomes-Gonçalves, E., Gzyl, H., & Mayoral, S. (2015). Two maxentropic approaches to determine the probability density of compound risk losses. *Insurance: Mathematics and Economics*, 62, pp. 42-53.

DOI: 10.1016/j.insmatheco.2015.03.001

Two maxentropic approaches to determine the probability density of compound risk losses

Erika Gomes[a], Henryk Gzyl[b], Silvia Mayoral[a],

[a] Business Administration, Universidad Carlos III de Madrid,

[b] Centro de Finanzas, IESA, Caracas

May 25, 2014

Abstract

Here we extend previous application of maxentropic methods to the reconstruction of probability densities of the total severity of losses, when the information available consists of the knowledge of a few values of the Laplace transform, obtained numerically. Therefore we previous work in which the analytic expression of the Laplace transform was available. In particular we shall obtain the probability density of the compound loss when the individual losses are lognormal random variables, and compute its relevant tail quantiles.

Since the data we consider is simulated, and the amount of simulated data is large, which makes statistical tests of quality easy most of the time, this work can be considered as a first step in the direction of analyzing the adequacy of the maxentropic procedures for the case in which the data is scarce and sample error must be taken into account.

1 Introduction

The first step in the implementation of the AMA methodology proposed by the Basel Committee, consists of selecting a line of activity and a type of work, and then to compute the probability distribution of total losses, and the next step is to aggregate these losses into one total loss and compute its probability distribution. The same problem is important as well for the insurance industry, in which the determination of the distribution of total severity is the first step towards premium computation and compliance with Solvency II requirements.

In a previous note (Gzyl et al, 2013), one of us was involved in a comparative study to solve the first step of the capital charge computation in the following situation: We supposed that we knew an analytic model for the combination of line/type operational risks, and moreover, that the Laplace transform of both the distributions of the frequency of losses and of the individual losses could be determined. In this case, we compared to possible ways of determining the distribution of the total loss.

As mentioned, here we shall consider a situation in which the model is known, but the Laplace transform of the individual losses doesn't seem to have a simple, closed form. We shall consider a lognormal model for individual losses. See Lipnik (1991) for a numerical procedure to determine the characteristic function of such variables. Thus, to carry on the program initiated in Gzyl et al, we have to go numerical. Instead of considering an approximate method to compute the Laplace transform, we shall consider simulated data. Here we shall consider a large sample of simulated data and compute the Laplace transform form it. In the sequel to this work we shall extend the maxentropic artillery to deal with error in the data.

Let us begin by reviewing the basics. For the implementation of the AMA methodology, one considers the total severity to be of the type $S = \sum_{n\geq 0}^{N} X_n$, where N denotes the (annual) frequency of losses, and X_n the individual losses. The Laplace transform of S, is readily expressed in terms of those of N and the X_n under the usual independence

and identical distributions.

Then, starting form the knowledge of

$$E[e^{-\alpha_i S}] = \psi(\alpha_k), \quad i = 1, ..., K. \tag{1}$$

we showed how the knowledge of $\psi(\alpha_i)$ for $K = 8$ values of the α's yielded reasonably good approximation to the density of S. As example we used a combination of Poisson model for the frequency of events and a Gamma model for the individual losses, and we compared our procedure with a good approximation of the true total density of losses. The results were quite satisfactory: Not only were the reconstructed densities close to the true ones, but also quantities like the VaR and $TVaR$ were quite close.

Regretfully, In the case when $E[e^{-\alpha X}]$ can not be computed analytically we have to deal with either an empirical sample, or of we trust a given model, with an approximate calculation. This is what happens when we know that the individual losses are, say, lognormal. As a first step into this program, we shall consider first the case in which the frequency is Poisson and the individual losses are lognormal. But we shall consider a situation in which he amount of data is very large, in order to be able to compare our procedure against empirical data analysis. In this case we can estimate the Laplace transform $\psi(\alpha_i)$ at a finite number of points, and apply the maxentropic procedure proposed before, and worry about errors in the data later.

To state the program that we carry on below, let us recall some notation from Gzyl (2013). As S is positive, we shall think of $Y = e^{-S}$ as a variable in $[0, 1]$ whose density $f_Y(y)$ we want to infer from fractional moments, that is, from

$$\int_0^1 y^\alpha f_Y(y)dy = \mu(\alpha_k), \quad i = 1, ..., K. \tag{2}$$

The relationship between the $\psi(\alpha)$ and $\mu(\alpha)$ is established noticing that the distribution of S has a point mass $e^{-\ell}$ at $S = 0$ corresponding to the event $N = 0$. Eliminating this, or equivalently, instead of (1, we consider the conditioned version

$$E[e^{-\alpha_i S}, \quad S > 0] = \int_0^\infty e^{-\alpha s} f_S(s)ds = \frac{\psi(\alpha_k) - e^{-\ell}}{1 - e^{-\ell}} \equiv \mu(\alpha_i), \quad i = 1, ..., K. \tag{3}$$

where, throughout we used ℓ for the parameter of the Poisson variable describing the frequency of events.

Thus, the generic program that we want to develop here goes as follows: We shall compute the moments $\mu(\alpha)$ from simulated data, but this time with a lognormal distribution for the individual losses, carry out our maxentropic reconstruction procedures, each of which will provide us with a possible solution to (2).

To make this note self-contained, in section 2 we recall the basics of the maximum entropy procedures, and in section 3 we describe the results of the implementation of the procedure to determine the distribution of total losses. At this point we mention the first of the procedures has been applied successfully in a large variety of problems, see Kapur (199X) or Gzyl and Velásquez (2011) for details and references.

Section 4 is devoted to the computation of several risk measures as well as the "price" of the operational risk. This is interesting for the risk manager considering insuring the operational risk losses to decrease the capital charges. Section 5 is devoted to a complementary problem. Suppose that the data is collected aggregated, that is, only S and N are measured, but the risk manager may want to know the distribution of the individual losses to decide on the result of the particular corrective loss prevention policy. We do this by applying a maxentropic decompounding procedure, which will play the role of check up test for our procedure.

2 The maxentropic approaches

Here we collect the basic results about the two possible ways of solving the problem using the maximum entropy method.

2.1 The standard maxent approach

That is the name of a variational procedure proposed by Jaynes (1957) to solve the (inverse) problem consisting of finding a probability density $f_Y(y)$ (on $[0,1]$ in this case), satisfying the following integral constraints:

$$\int_0^1 y^{\alpha_k} f_Y(y) dy = \mu_Y(\alpha_k) \qquad \text{for} \qquad k = 0, 1, ..., K. \tag{4}$$

We set $\alpha_0 = 0$ and $\mu_0 = 1$ to take care of the natural requirement on $f_Y(y)$. The intuition is rather simple: The class of probability densities satisfying (4) is convex. One can pick up a point in that class one by maximizing (or minimizing) a concave (convex) functional (an "entropy") that achieves a maximum (minimum) in that class. That extremal point is the "maxentropic" solution to the problem. It actually takes a standard computation to see that when the problem has a solution it is of the type

$$f_K^*(y) = \exp\left(-\sum_{k=0}^{K} \lambda_k^* y^{\alpha_k}\right) \tag{5}$$

in which the number of moments K appears explicitly. It is usually customary to write $e^{-\lambda_0^*} = Z(\boldsymbol{\lambda}^*)^{-1}$, where $\boldsymbol{\lambda^*} = (\lambda_1^*, ..., \lambda_K^*)$ is a $K-$dimensional vector. Clearly, the generic form of the normalization factor is given by

$$Z(\boldsymbol{\lambda}) = \int_0^1 e^{-\sum_{k=1}^{K} \lambda_k y^{\alpha_k}} dy. \tag{6}$$

With this notation the generic form of the solution looks like

$$f_M^*(y) = \frac{1}{Z(\boldsymbol{\lambda}^*)} e^{-\sum_{k=1}^{M} \lambda_k^* y^{\alpha_k}} = e^{-\sum_{k=0}^{M} \lambda_k^* y^{\alpha_k}}. \tag{7}$$

To complete, it remains to specify how the vector $\boldsymbol{\lambda}^*$ can be found. For that one has to minimize the dual entropy:

$$\Sigma(\boldsymbol{\lambda}, \boldsymbol{\mu}_Y) = \ln Z(\boldsymbol{\lambda}) + < \boldsymbol{\lambda}, \boldsymbol{\mu}_Y > \tag{8}$$

where $< \mathbf{a}, \mathbf{b} >$ denotes the standard Euclidean scalar product and $\boldsymbol{\mu}$ is the $K-$vector with components μ_k, and obviously, the dependence on $\boldsymbol{\alpha}$ is through $\boldsymbol{\mu}_Y$.

2.2 The method of maximum entropy in the mean

The MEM provides another interesting approach to solve the problem of determining $f_Y(y)$ such that (2) holds. It an be summed up by saying that it consists of a technique to obtain $f_Y(y)$ or its discretized version, as the expected value of an auxiliary probability distribution which is determined by an entropy maximization procedure. To implement MEM numerically, the first step consists of discretizing the problem. This leads to a system of equations like:

$$\sum_{1}^{M} A_{i,j}x_j = \mu_i, i = 0, 2, ..., K \text{ with } x_j \geq 0 \text{ for } j = 1, ..., M. \tag{9}$$

Here we have set $x_j = (1/N)f((j-1)/M)$ and $A_{i,j} = (\frac{2j-1}{2M})^{\alpha_i}$, for j=1,...,M. The first factor in front of the definition of x_j comes from the discretization $dy \approx 1/M$, and $A_{i,j}$ is obtained as the midpoint approximation of y^{α_i} for the chosen partition. We have added a normalization constraint by choosing $\alpha_0 = 0$. Note that with this choice we have $A_{0,j} = 1$ for $j = 1, ..., M$ and $\sum_{j=0}^{M} x_j = \mu_0 = 1$ as normalization condition for $\alpha_0 = 0$. Observe that if we consider a partition of size 200 say, as we have $K = 9$ moments, we have a system of 9 equations to determine 200 unknowns subject to a positivity constraint. Actually, the original problem consists of $K = 9$ equations to determine a continuous function, so the discretization may seem to be an improvement from the dimensionality point of view.

To continue, to take care of the positivity constraint, we consider a space $\Omega = [0, \infty)^M$ with its Borel sets $\mathcal{F}$. Denote by $\boldsymbol{\xi} = (\xi_1, ...\xi_M)$ a generic point in Ω and define the coordinate maps $X_j(\boldsymbol{\xi}) : \Omega \to [0, \infty)$ by $X_j(\boldsymbol{\xi}) = \xi_j$. On $(\Omega, \mathcal{F})$ we place a reference measure $dQ(\boldsymbol{\xi})$ and we search for a measure $P << Q$ such that

$$\sum_{j=1}^{M} A_{i,j}E_P[X_j] = \mu_i \text{ for } i = 1, ..., K. \tag{10}$$

The choice of the measure Q is up to the modeler, and it may be thought of as the first guess of P. The only restriction upon it is that the convex hull generated by its support is Ω. The purpose of that condition is to ensure that any strictly positive density $\rho(\boldsymbol{\xi})$ of

P respect to Q is such that

$$E_P[X_j] = \int_\Omega \xi_j \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} \in [0, \infty),$$

that is, the positivity constraint is automatically satisfied. The other constraints will be achieved by a special choice of P. With the notations introduced above, we note that the class of probabilities

$$\mathcal{P} = \{P << Q \text{ such that 10 holds true}\}$$

is a convex, closed set if not empty, which we **suppose** to be the case. On this set we define the entropy function

$$S_Q(P) = -\int_\Omega \rho(\boldsymbol{\xi}) \ln\left(\rho(\boldsymbol{\xi})\right) dQ(\boldsymbol{\xi})$$

whenever the integral is finite or ∞ otherwise.

$$\text{Find } P^* \in \mathcal{P} \text{ which maximizes } S_Q(P). \tag{11}$$

From now on the routine is pretty much as in the previous section. It is clear that the MEM uses the SME as a stepping stone. The problem is similar but in another setup. The generic solution is of the type

$$\rho^*(\boldsymbol{\xi}) = \frac{1}{Z(\boldsymbol{\lambda}^*)} e^{-(<(\boldsymbol{A}^t \boldsymbol{\lambda}^*), \boldsymbol{\xi}>} \tag{12}$$

where, again $\boldsymbol{\lambda}^*$ is obtained minimizing the dual entropy function

$$\Sigma(\boldsymbol{\lambda}, \boldsymbol{\lambda}) = Z(\boldsymbol{\lambda}) + < \boldsymbol{\lambda}, \boldsymbol{\mu} >, \tag{13}$$

where, recall, $\boldsymbol{\lambda}$ is a $K = 9-$dimensional vector, and $\boldsymbol{\mu}$ is the $K = 9-$dimensional vector of constraints defined (9). This time, the function $Z(\boldsymbol{\lambda})$ is defined by

$$Z(\boldsymbol{\lambda}) = \int_\Omega e^{-(<(\boldsymbol{A}^t \boldsymbol{\lambda}^*), \boldsymbol{\xi}>} dQ(\boldsymbol{\xi}).$$

The following result is a simplified version of the duality theory presented in chapter 4 of Borwein and Lewis (2000). It backs up the computations corresponding to the two example presented below.

Theorem 1. *Let us suppose that* $\inf \Sigma(\boldsymbol{\lambda}, \boldsymbol{\mu})$ *is achieved at an interior point* $\boldsymbol{\lambda}^*$ *of* $\{\boldsymbol{\lambda} \in \mathbb{R}^K | Z(\boldsymbol{\lambda}) < \infty\}$. *In this case, the probability* P^* *solving (11) has density* $\rho^*(\boldsymbol{\xi})$ *given by (12), and*

$$S_Q(P^*) = \Sigma(\boldsymbol{\lambda}^*, \boldsymbol{\mu}).$$

In the next section we shall illustrate the MEM in one specific setups. The basic idea is that the x_j are to be estimated as expected values with respect to a probability P to be determined as explained above.

2.3 Poisson reference measure

This time, instead of a product of exponentials, we shall consider a product of Poisson measures, i.e., we take

$$q(d\xi) = e^{-\eta} \sum_{k \geq 0} \frac{\eta^k}{k!} \epsilon_{\{k\}}(d\xi).$$

Here we use $\epsilon_{\{a\}}(d\xi)$ to denote the unit point mass (Dirac delta) at a. Certainly the convex hull of the non-negative integers is $[0, \infty)$. Notice that now

$$Z(\boldsymbol{\lambda}) = \prod_{j \geq 0}^{M} \exp\Big(- \eta(1 - e^{-(\boldsymbol{A^t \lambda})_j})\Big)$$

from which we obtain

$$\Sigma(\boldsymbol{\lambda}) = -\eta \sum_{j-1}^{M} \Big(1 - e^{-(\boldsymbol{A^t \lambda})_j}\Big) + < \boldsymbol{\lambda}, \boldsymbol{m} >$$

Notice now that if $\boldsymbol{\lambda}^*$ minimizes that expression, then the estimated solution to (9) is

$$x_j^* = e^{-(\boldsymbol{A^t \lambda}^*)_j} \tag{14}$$

Do not forget that above, $\boldsymbol{A^t \lambda}_j = \sum_{i=0}^{8} \lambda_i A_{i,j}$. Recall as well that $A_{0,j} = 1$ for $j = 1, ..., M$. As $\sum_{j=0}^{M} x^*(j) = 1$, we can rewrite (14) as

$$x_j^* = \frac{e^{-(\hat{\boldsymbol{A}}^{\boldsymbol{t}} \hat{\boldsymbol{\lambda}}^*)_j}}{z(\hat{\boldsymbol{\lambda}}^*)} \tag{15}$$

where we redefined $\hat{\mathbf{A}}$ as the matrix obtained from $\mathbf{A}$ by deleting the zero-th row, and $\hat{\boldsymbol{\lambda}}$ as the 8−dimensional vector obtained by deleting λ_0. To complete,

$$z(\hat{\boldsymbol{\lambda}}^*) = e^{-\lambda_0} = \sum_{j=1}^{M} e^{-(\hat{\boldsymbol{A}}^t \hat{\boldsymbol{\lambda}}^*)_j}.$$

Clearly, since $x_j^* = (1/M) f_Y^*((j-1)/M)$, from this, we have just one more procedure to determine $f_Y^*(y)$ approximately.

3 Quality of the reconstructions

Once a density has been determined, it is necessary to test if is consistent with the data. The evaluation process is inherently a statistical problem, which involves exploring, describing, and making inferences about data sets containing observed and estimated values. Here we describe a variety of tests, and in the Appendix, we add further detail about them.

Exploratory alternatives for model checking include visual comparisons through the use of graphical tools like reliability and calibration plots which measure the agreement between the estimation and the observed data. The reliability plot, popularly known as QQ-plot serves to determinate the quality of a fit by the proximity of the quantiles of the maxentropic (reconstructed) to the diagonal, the closer the better the approximation.

A similar tool is the marginal calibration plot which is the graph of $F^*(s_j) - F_n(s_j)$ versus x_t, where $F^*(s_j)$ is the (cumulative) distribution function of the reconstructed density, and $F_n(s_j)$ is the observed or empirical distribution function of the observed losses s_j. Here, observing small fluctuations about zero means that the observations and the maxentropic estimations have the same (or nearly the same) marginal distribution (See Appendix).

Numerical comparisons are also considered for the evaluation of the results. Among them, we compute the L_1 and L_2 distances between the densities and the histogram of

the observed data. The needed computations are:

$$L_1 = \sum_{k=0}^{M-1} \int_{b_k}^{b_{k+1}} |f^*(s) - f_e(s)|ds + \int_{b_M}^{\infty} |f^*(s)|ds$$

$$L_2 = \sqrt{\sum_{k=0}^{M-1} \int_{b_k}^{b_{k+1}} (f^*(s) - f_e(s))^2 ds + \int_{b_M}^{\infty} (f^*(s)^2 ds}$$

where b_k and b_{k+1} are the limits of each of the bins in the histogram, f^* is the maxentropic (reconstructed) density and f_e is the density obtained from the histogram (i.e. frequency/size of the data set) this measure has the disadvantage of depending on the number and the position of the bins in the histogram.

Also, we consider the MAE and RMSE errors between the distribution functions obtained with the maxentropic methods vs. the observed data, these are calculated as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |F^*(s_i) - F_n(s_i)|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (F^*(s_i) - F_n(s_i))^2}$$

where $F^*(s)$ is the distribution function obtained applying the maxentropic procedure, and $F_n(x_t)$ is the distribution function of the observed data. We could think of these as empirical L_1 and L_2 distances between distribution functions. These distances, or measures of fit, have the advantage of not relying neither on the number nor on the location of the bins. Here we calculate F^* for at the points of the sample data.

A further idea, consists of testing whether the estimated density function $\widehat{f}(\cdot)$, equals the true underlying density function $f(\cdot)$. One way to do this, is based on an integral

transform that date to Rosenblatt (1952), and popularized by Diebold et al.(1998), which consists in testing whether the probability integral transforms (PIT), that is, the p_j defined right below, are independent and uniformly distributed.

The probability integral transform (PIT) it is defined as

$$\int_{-\infty}^{s_j} f^*(s)ds = F^*(s_j) = p_j \tag{16}$$

where s_j is the $j-$th sample of the variable of interest, which in our case is the observed (simulated) total loss. It comes from a population described by an unknown density $f_S(s)$. In the definition, $f^*(s)$ denotes the reconstructed (maxentropic) density, and p_j, the probability integral transform (PIT) (of s_j). These variables should follow a uniform distribution, ($p_j \sim Unif(0,1)$) if $f^*(s)$ were a good reconstruction of $f_S(s)$.

Deviation from uniformity will indicate that the reconstruction may have failed to capture some aspect of the underlying data generation or data collection process. To test uniformity and independence in the PIT test, a visual inspection of a PIT-histogram and autocorrelation plots is used along with additional tests like the KS-test, the Anderson-Darling test and the Cramer-Von test. Additionally, we also consider the Berkowitz back test approach, which consists of taking the inverse normal transformation of the PIT, and then applying a joint test for normality and autocorrelation, that sometimes is combined with the test Jarque-Bera of Normality.

The calculation of the VaR and the TVaR using densities obtained applying the SME and MEM methods, may be considered to be an additional way to evaluate the quality of the reconstruction. This is done by comparing these values with the VaR and TVaR of the observed data, which in our case is rather abundant.

All the comparisons that we make in this paper are always made with respect to the observed data, because the true density of a compound sum of lognormal variables, is far from being known.

Additionally, for the evaluation of the results we use two independent data sets that

come from the same population. One, which we call the *observed data*, is used as input to the maxentropic methods, and a validation data set, which we call the *test data*, is used to audit the results. This is done in order to avoid the problem of overfitting, which is defined as the situation in which the model gives better results for the observed set than for other data set which comes from the same population. Besides this helps us to evaluate the ability of the maxentropic density to perform well on unobserved data.

4 Numerical essays

The example that we consider consists of a compound process in which the frequency N of events in some time interval, is a Poisson variable of parameter ℓ, and the individual losses are distributed according a lognormal distribution ($\sim logN(\mu, \sigma^2)$). We have already mentioned that this case the Laplace transform of $S(N)$ can not be computed, but can be estimated numerically when the sample is large. We shall consider several values of ℓ and several values of (μ, σ)

To generate the data we proceed as follows:

1. We generate a sample of size 8.000 from a compound distribution $S(k) = \sum_{1\leq j}^{n_k} X_n$, where $\{n_k | 0 \leq k \leq N\}$, are the number of samples from the lognormal entering in each random sum, and N the total number of samples n_k, that follow a Poisson distribution with parameter ℓ. This usage of the letter N should cause no confusion. Using this sample we calculate the moments $\mu(\alpha_k)$ which is the input to obtain the Maxentropic distribution.

2. We apply the SME and MEM approaches of Sections (2.1) and (2.2), using 8 fractional moments ($\mu(\alpha_k)$) of the exponential of the variable, where $\alpha_i = \frac{1.5}{i}$ with $i = 1, \ldots, 8$. For each maxentropic approach we consider five cases. One is used to exemplify all the procedures in full detail. This shall be called Case (1), and is characterized by the parameters $\ell = 3$, $\mu = 0$, $\sigma = 0.25$, and the other four

(Cases 2-5) are used to prove the robustness of the method with respect to changes in the parameters. The differences between the cases consist of keeping one of the parameters fixed and change the others. Thus we consider cases with the same parameters $\mu = 0, \sigma^2 = 0.25$ of the lognormal, but different Poisson parameters (Cases 2-3), and other two cases with the same Poisson parameter $\ell = 3$ and different lognormal parameters (Cases 4-5). The minimization procedure is carried out using Barzilai-Borwein's (1988) algorithm.

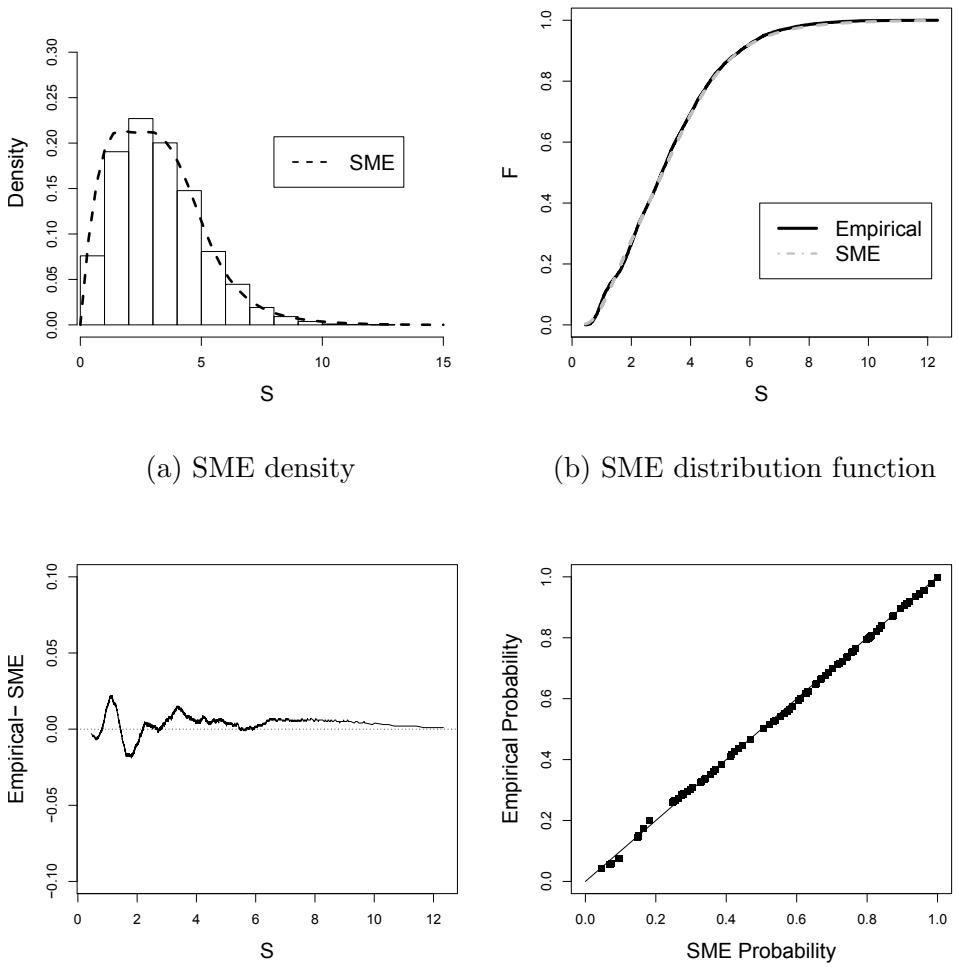
3. At the end we evaluate the quality of the reconstructions using a variety of criteria (described in Section (3) and in the Appendix) with two different data sets: the observed and a test set.

5 Numerical reconstructions with SME

We start with a compound process where the frequency N is a Poisson variable of parameter $\ell = 3$ and the individual losses are described by a lognormal distribution with parameters $\mu = 0$, $\sigma = 0.25$, the resulting compound sum, called S, represents the total losses that can affect a given business.

In Figure (1) we display the density, the distribution function, the marginal calibration and reliability diagram of the SME based reconstruction and the observed data constructed with the observed data set. These allow us to observe the performance of the SME method with the observed data as input, by applying different quality of reconstruction criteria to the same result, which gives us a visual idea of how good the obtained reconstruction is.

Notice in Figure (1a) that the SME density and the histogram of the observed data seem to look close. The same can be seen in Figure (1b) with the distribution function of the SME and the observed data. In Figure (1c) we display the marginal calibration diagram that allows us to observe the differences between the SME reconstruction and the



(a) SME density

(b) SME distribution function

(c) SME Marginal calibration

(d) SME Reliability Diagram

Figure 1: SME results of the Case (1): $\ell = 3$, $\mu = 0$, $\sigma = 0.25$

observed data. This figure shows that the differences are not larger than 0.022. Such small fluctuations about zero are indicators of the good quality of the reconstruction. Additionally, in Figure (1d) we have the reliability diagram of the observed frequency against the SME density. That diagram measures the agreement between estimated probabilities and the observed frequencies as shown by the proximity of the plotted curve to the diagonal. Here we can see that there is only a very small deviation at the beginning of the graph.

Even though the results of Figure (1) seems to indicate a good reconstruction, as measure of closeness we may also consider the L_1 and L_2 norms of the distances between

the reconstructed density and the empirical density, as well as the MAE and RMSE distances detailed in section (3). In Table (1) we show the results of the computations. These confirm the plausibility of the good quality of the reconstruction displayed in Figure (1).

Approach	L1	L2	MAE	RMSE
SME	0.1225	0.0598	0.0071	0.0089

Table 1: Errors SME approach,
Case (1): $\ell = 3$, $\mu = 0$, $\sigma = 0.25$

It is convenient to test density reconstructions for correct calibration. This consists of testing whether the inverse probability transforms (PIT) is independent and uniformly distributed. Deviations from uniformity may indicate a poor reconstruction. In Figure (2) we display the PIT transformation and correlograms of different powers for Case (1). As we can see the PIT histogram seems to be uniform and the correlation plots do not show any sign of dependence.

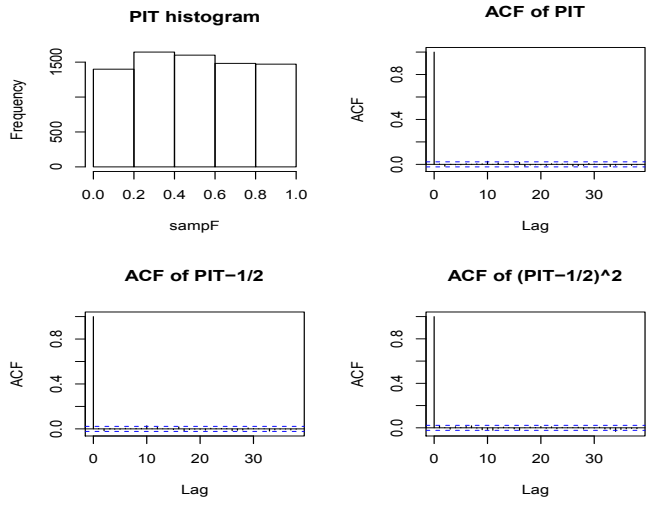


Figure 2: Probability integral transform (PIT) histogram and sample autocorrelation functions for the SME approach. Case (1) $\ell = 3$, $\mu = 0$, $\sigma = 0.25$.

To test robustness of the SME method, we performed reconstructions with simulated

data for other values of the parameters. In each of the cases we went through the same routine: we generated the data, computed the moments and carried out the maxentropic procedure. In Figure (3) we display the different SME reconstructions, along with the histogram of the observed data. A glance at the figures should convince us that the different reconstruction seem to fit the observed data reasonably. The same criteria that we used above to measure quality of the running case (case 1), yield consistent results in these cases as well.

For example, the distances between the estimated and the empirical densities in the L_1 and the L_2 norms, and MAE and RMSE distances, for cases (2) to (5) considered in Figure (2), are listed in Table (2). They suggest that the reconstructions are reasonably good.

Error	Case (2)	Case (3)	Case (4)	Case (5)
L1-norm	0.2649	0.0947	0.1145	0.1400
L2-norm	0.2099	0.0399	0.0464	0.0698
MAE	0.0216	0.0038	0.0069	0.0235
RMSE	0.0257	0.0047	0.0087	0.0297

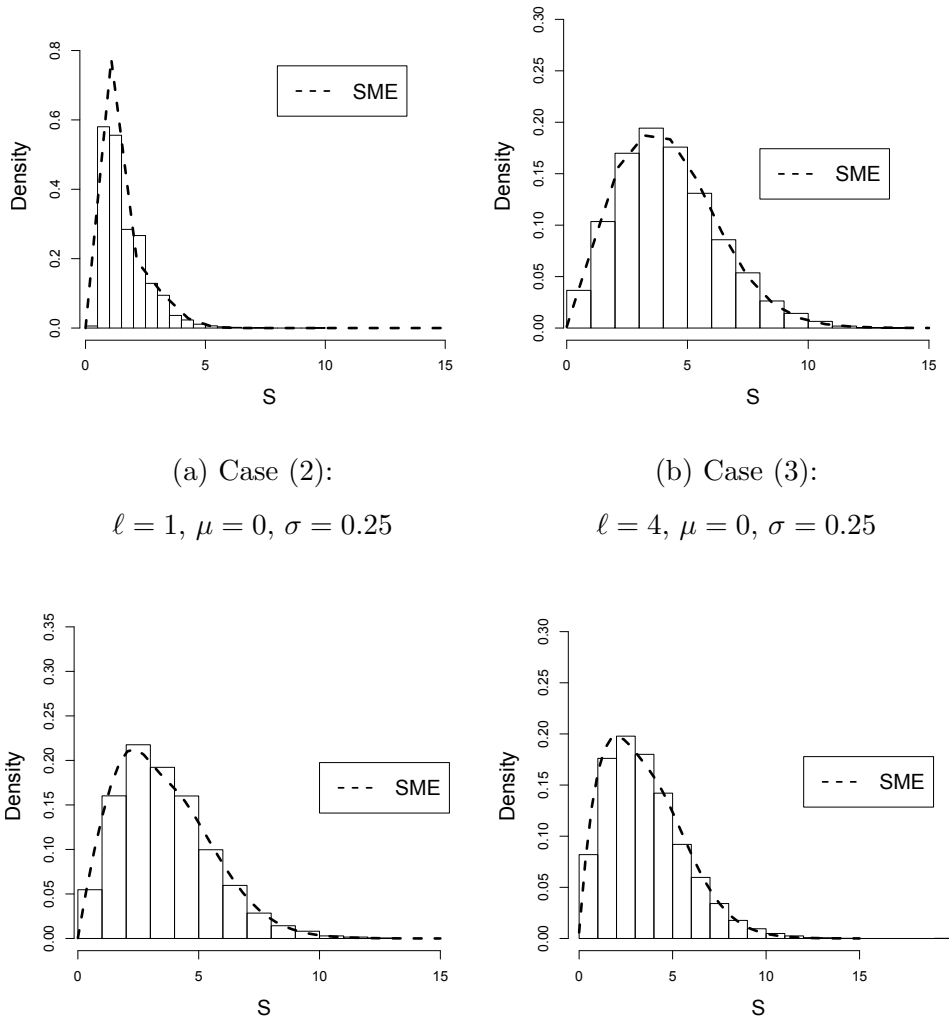
Table 2: Errors SME approach, Cases (2)-(5)

In order to evaluate the ability of the SME method to perform well on unobserved data, we calculate the error and distances of the SME densities with a test data set, that is, an independent sample which comes from the same population. The corresponding L_1, L_2, MAE and RMSE results for all the cases consider in this paper are showed in Table (3). They seem to show similar results to the obtained with the observed data set.

Error	Case (1)	Case (2)	Case (3)	Case (4)	Case (5)
L1-norm	0.1223	0.2103	0.0960	0.0865	0.1096
L2-norm	0.0649	0.1847	0.0408	0.0421	0.0556
MAE	0.0109	0.0216	0.0126	0.0110	0.0256
RMSE	0.0147	0.0259	0.0140	0.0142	0.0310

Table 3: Errors SME approach (test set)

For each case, that is for each choice of parameters, we applied the quality of recon-



(a) Case (2): $\ell = 1$, $\mu = 0$, $\sigma = 0.25$

(b) Case (3): $\ell = 4$, $\mu = 0$, $\sigma = 0.25$

(c) Case (4): $\ell = 3$, $\mu = 0.1$, $\sigma = 0.25$

(d) Case (5): $\ell = 3$, $\mu = 0$, $\sigma = 0.5$

Figure 3: Combined plot of densities of SME approach for compound sums S, for different parameters

struction criteria described in section (3) and the appendix. The results of the different criteria are displayed in Table (4). Three stars means that we do not reject the null hypothesis of no differences between the empirical and maxent density, at 1% of significance and two stars means that we do not reject the null hypothesis at 5% of significance. All the cases seems to pass all the tests. The critical values used for all the tests are indicated in the Appendix.

Criterion	Case (1)	Case (2)	Case (3)	Case (4)	Case (5)
KS test of uniformity	1.51***	1.28**	0.87**	1.17**	1.24**
Anderson-Darling test:	1.95**	2.44**	1.46**	1.51**	1.39**
Cramer- Vom test:	0.30**	0.44**	0.19**	0.13**	0.29**
Berkowitz Test: :	5.74**	2.17**	4.43**	4.55**	1.19**
Jarque-Bera Test :	1.34**	8.98***	2.93**	4.27**	3.67**

Table 4: *Statistics* of SME approach (test set)

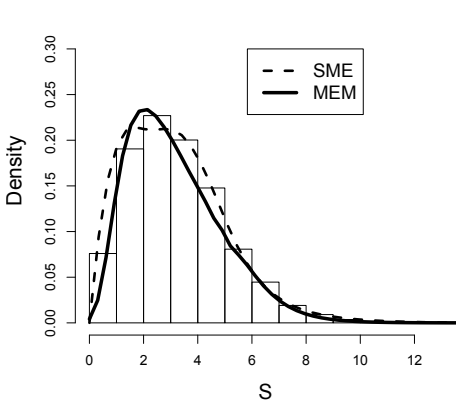
5.1 Reconstruction with the MEM method

In this section we describe the results of the implementation of the procedure described in section (2.2). We might regard this as the first step towards the process of incorporating errors in the data related to the size of the sample.

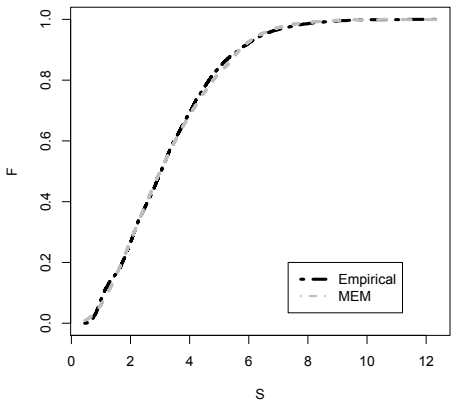
Again we start by considering the compound sum of a Poisson and simple lognormals with parameters $\ell = 3$, $\mu = 0$ and $\sigma = 0.25$. Here we use the Method de Maximum Entropy with a Poisson reference measure and parameters $\eta = 2$, and a partition of $[0, 1]$ od size $M = 200$, as described in section (2.2).

In Figure (4) we display the density, the distribution function, the marginal calibration and reliability diagram of the MEM reconstruction and the observed data set. In Figure (4a) we show the SME and MEM reconstructions along with the histogram of the observed data (observed data set). As in section (4.1) we show in Figures (4b), (4c) and (4d), the distribution function, the Marginal Calibration diagram and reliability diagram respectively, here we only show MEM reconstruction along with their empirical counterpart. In Figure (4c) we observe minors fluctuations about zero, with values not greater than 0.027 in absolute value, this minors fluctuations indicate a good fit.

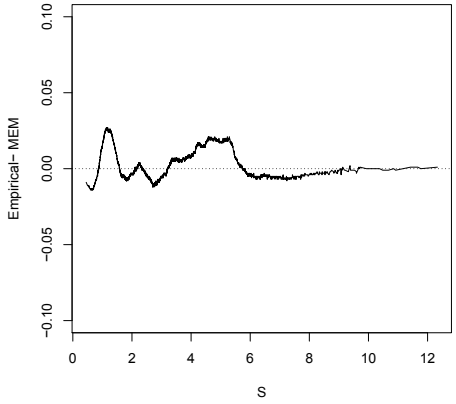
We also show in Table (5) the numerical distances between the reconstructions (SME & MEM) and the errors to see how different they are in relation with the observed data



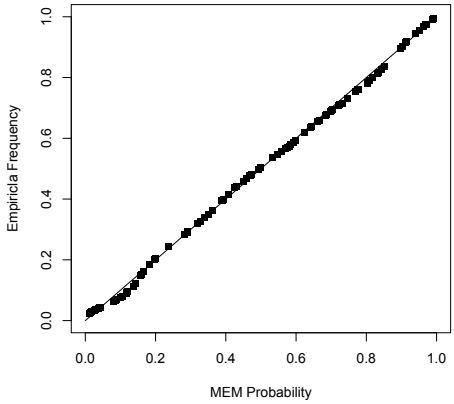
(a) MEM density



(b) MEM distribution function



(c) MEM Marginal calibration



(d) MEM Reliability Diagram

Figure 4: MEM ($\eta = 2$, $M = 200$) results of the case $\ell = 3$, $\mu = 0$, $\sigma = 0.25$

set. Clearly, in table (5) the SME reconstruction seems to be a little better than the MEM reconstruction.

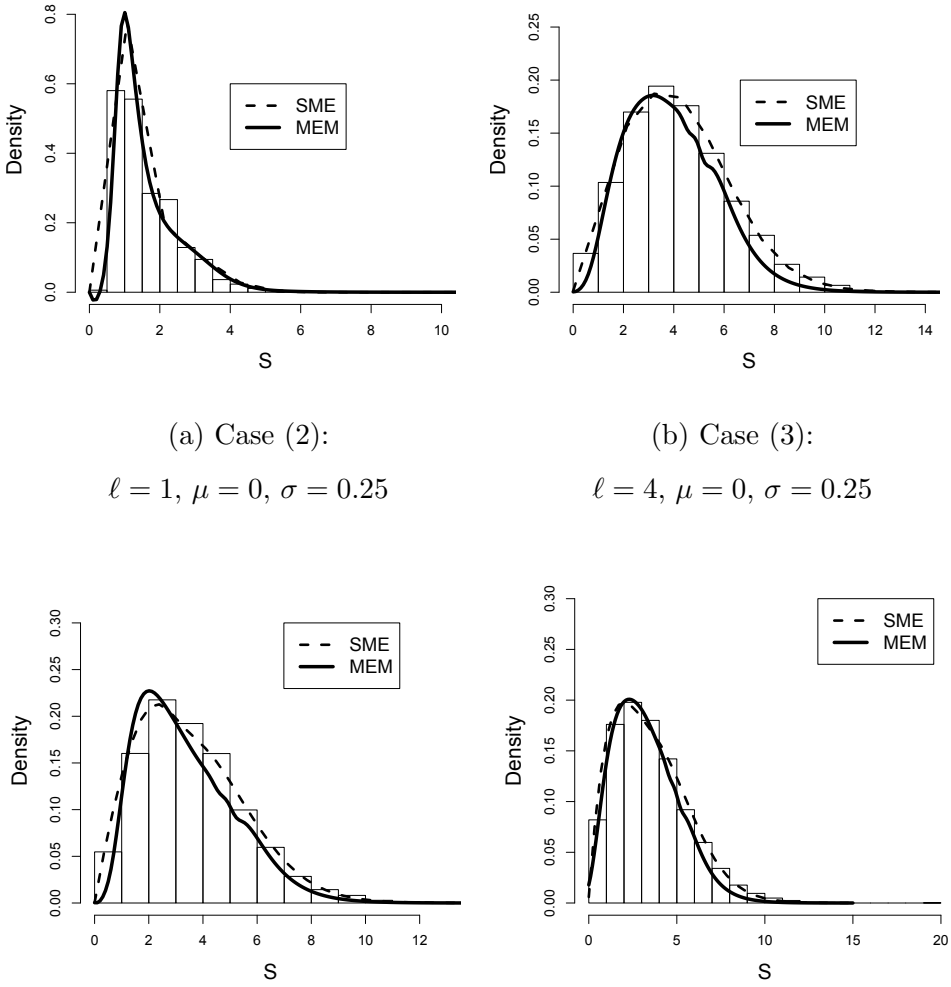
Approach	L1	L2	MAE	RMSE
SME	0.1225	0.0598	0.0071	0.0089
MEM	0.1279	0.0609	0.0086	0.0109

Table 5: Errors of SME and MEM reconstructions

Case (1): $\ell = 3$, $\mu = 0$, $\sigma = 0.25$

In Figure (5) we display the results of MEM and the SME along with the empirical histogram (observed set) of the compound sum for different values of the parameters ℓ,

μ & σ, in order to see the differences between the reconstructions and the observed data. The reconstructions seems to fit the histograms, although is clear there is little difference between the reconstruction. We point out that the wiggles in the MEM density result from the interpolation process (remember that MEM involves a discretization, and the plotting procedure interpolates between the resulting points).



(a) Case (2):
$\ell = 1$, $\mu = 0$, $\sigma = 0.25$

(b) Case (3):
$\ell = 4$, $\mu = 0$, $\sigma = 0.25$

(c) Case (4):
$\ell = 3$, $\mu = 0.1$, $\sigma = 0.25$

(d) Case (5):
$\ell = 3$, $\mu = 0$, $\sigma = 0.5$

Figure 5: Combined plot of SME & MEM densities

In Table (6) we list the L_1 and L_2 distances between reconstructed and empirical densities, along with the MAE and RMSE distances between the MEM-distribution func-

tion and the empirical distribution function. These errors are a little larger than those obtained for the SME reconstruction.

ERROR	**Case (2)**		**Case (3)**		**Case (4)**		**Case (5)**	
	SME	MEM	SME	MEM	SME	MEM	SME	MEM
L1-norm	0.2649	0.2560	0.0947	0.1952	0.1196	0.1652	0.1105	0.1498
L2-norm	0.2099	0.2091	0.0399	0.0857	0.0563	0.0770	0.0516	0.0605
MAE	0.0216	0.0182	0.0038	0.0172	0.0074	0.0123	0.0058	0.0114
RMSE	0.0257	0.0221	0.0047	0.0248	0.0094	0.0145	0.0064	0.0166

Table 6: Errors SME and MEM approaches, Cases (2)-(5)

In Figure (6) we display the PIT transformation and correlograms of different powers of Case (1). As we can see, the PIT histogram seems to be uniform and the correlation plots do not show any sign of dependence.

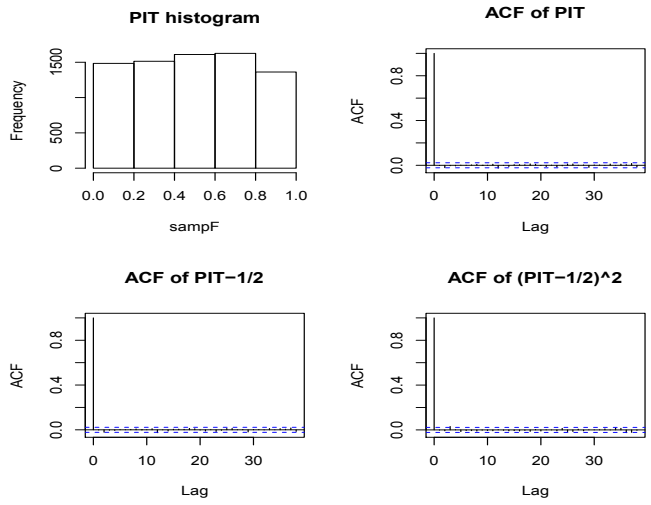


Figure 6: Probability integral transform (PIT) histogram and sample autocorrelation functions for the MEM approach. Case (1) $\ell = 3$, $\mu = 0$, $\sigma = 0.25$.

Using the test set we can observe how does the MEM reconstruction perform on unobserved data. In Table (7) where we display the values of the L_1, L_2, the MAE and the RMSE measures of distance. We may call that a good performance.

In Table (**??**) we display the results of the quality measures for the test data set. As

Error	Case (1)	Case (2)	Case (3)	Case (4)	Case (5)
L1-norm	0.1896	0.2057	0.1580	0.1598	0.1751
L2-norm	0.1370	0.1866	0.0781	0.0763	0.0704
MAE	0.0131	0.0186	0.0201	0.0170	0.0161
RMSE	0.0150	0.0225	0.0223	0.0201	0.0198

Table 7: Errors MEM approach (test set)

Criterion	Case (1)	Case (2)	Case (3)	Case (4)	Case (5)
KS test of uniformity	1.07**	1.11**	1.22**	1.53***	1.23**
Anderson-Darling test:	2.28**	1.68**	3.71***	3.31***	2.99***
Cramer- Vom test:	0.30**	0.31**	0.41**	0.25**	0.39**
Berkowitz Test:	7.74**	5.27**	10.04***	6.94**	1.54**
Jarque-Bera Test:	10.25	7.78***	3.75**	2.28**	9.20***

Table 8: *Critical Values* of MEM approach (test set)

in Section (4.1) those market with asterisks do no reject the null hypothesis of equality between distributions.

6 Risk measures

In this section we present the computation of the VaR and $TVaR$ of the total loss S corresponding to the parameters $\ell = 3$, $\mu = 0$ and $\sigma = 0.25$ (i.e., Case (1)), at various confidence levels. This results also serves to test the potential of the SME and MEM approaches. In this example we compare the quantiles of the reconstruction against the sample quantiles.

For the calculation of VaR and $TVaR$ at the confidence level γ of the SME and MEM reconstructions, we use the theoretical definition of VaR and TVaR simplified in

the following lemma taken from Rockafellar and Uryasev (2000).

Lemma 1. *The function*

$$a \to U(a) = a + \frac{1}{1-\gamma} \int_a^{\infty} (t-a) f^*(t) dt$$

defined on $(0, \infty)$ *is convex in* a*, achieves its minimum at* VaR_γ *and its minimum value is* $TVaR_\gamma(S) = E^*[S|S > VaR_\gamma]$*. Above,* f^* *denotes the maxentropic density.*

In Tables (9) and (10) we display the resulting computations for a collection of γ's. The lasts columns are corresponding to VaR and $TVaR$ of the simulated sample (observed data set) along with their confidence levels at 95%. Additionally we include the absolute difference between the VaR and TVaR of the observed data with those from the VaR and TVaR obtained from the reconstructions.

In order to calculate the empirical VaR and TVaR we consider $S > 0$ ordered in increasing size ($s_1 \leq s_2 \leq s_n$), and then estimated VaR as $\widehat{VaR}_\gamma(S) \approx x([N(\gamma)])$, where $[a]$ denotes the integer part of the real number a.. The estimate of the TVaR is obtained from the same ordered list of values as

$$\widehat{TVaR}_\gamma = \frac{1}{N - [N\gamma] + 1} \sum_{j=[N(\gamma)]}^{N} s_j$$

. The confidence intervals for the VaR and TVaR were calculated by resampling without replacement using subsamples of size 90% of the data size.

In tables (9) and (10) the asterisks indicate that the calculated VaR and TVaR for the reconstructions belongs to the confidence interval of 95% showed in the same tables.

7 Decompounding

It may not be always possible to observe frequency and severity separately, even though the frequency of events were recorded, only the total or aggregated loss data is available.

Approach				Errors		Confidence Interval	
γ	SME	MEM	Empirical	SME error	MEM error	VaR_{inf}	VaR_{sup}
0.900	5.657*	5.657*	5.672	0.015	0.015	5.587	5.763
0.910	5.798*	5.818*	5.809	0.011	0.009	5.725	5.920
0.920	5.939*	5.980*	5.968	0.029	0.012	5.872	6.055
0.930	6.081*	6.141*	6.118	0.037	0.023	6.021	6.227
0.940	6.222*	6.303*	6.299	0.077	0.004	6.202	6.379
0.950	6.505*	6.465*	6.474	0.031	0.009	6.377	6.614
0.960	6.788*	6.788*	6.759	0.029	0.029	6.617	6.908
0.970	7.071*	7.273*	7.122	0.051	0.151	6.955	7.334
0.980	7.495*	7.919	7.583	0.088	0.336	7.428	7.767
0.990	8.485*	8.566*	8.384	0.101	0.182	8.078	8.593
0.995	9.051*	9.061*	9.016	0.035	0.045	8.747	9.210
0.999	9.192	10.182*	10.34	1.148	0.158	9.686	11.43

Table 9: Comparison of VaR for the SME and MEM reconstructions,
Case (1): $\ell = 3$, $\mu = 0$, $\sigma = 0.25$

Nevertheless, the risk analyst may want to know the distribution of individual losses, because it is at that level where loss prevention or mitigation are applied. The tools that we have developed allow us to determine the individual loss distribution as well.

In our case, we either know how to compute or we can estimate numerically, the Laplace transform $\psi(\alpha)$ of the total losses S and we also know how to compute the generating function $G(z)$ of the frequency of the events N. From these ingredients, we can compute the Laplace transform $\phi(\alpha)$ of the individual losses, which we can use as starting point to determine the probability distribution of individual losses. The relationship between the various Laplace transforms is contained in (3, from which we obtain

$$\phi(\alpha_k) = \tfrac{1}{l} ln(\psi(\alpha_k)) + 1,$$

Approach				Errors		Confidence Interval	
γ	SME	MEM	Empirical	SME error	MEM error	$TVaR_{inf}$	$TVaR_{sup}$
0.900	6.817*	6.833*	6.839	0.022	0.006	6.721	6.948
0.910	6.930*	6.967*	6.961	0.031	0.006	6.846	7.079
0.920	7.055*	7.041*	7.095	0.040	0.054	6.974	7.232
0.930	7.192*	7.121*	7.245	0.053	0.124	7.109	7.389
0.940	7.344*	7.303*	7.417	0.073	0.114	7.276	7.562
0.950	7.513*	7.528*	7.622	0.109	0.094	7.459	7.781
0.960	7.935*	7.814*	7.874	0.061	0.060	7.700	8.045
0.970	8.205*	8.220*	8.188	0.017	0.032	7.989	8.373
0.980	8.533*	8.516*	8.601	0.068	0.085	8.405	8.817
0.990	8.959*	8.912*	9.262	0.303	0.350	8.968	9.555
0.995	9.556*	9.606*	9.843	0.287	0.237	9.465	10.222
0.999	10.543*	11.215*	11.167	0.624	0.048	10.341	11.848

Table 10: Comparison of TVaR for the SEM and MEM reconstructions,
Case (1): $\ell = 3$, $\mu = 0$, $\sigma = 0.25$

where, to use what we have developed above, we may write $\psi(\alpha_k) = e^{-\ell} + (1 - e^{\ell}) \int_0^1 y^{\alpha_k} f_Y^*(y) dy$, for the Laplace transform of the aggregate losses, ℓ being the parameter of the observed Poisson frequency, which is supposed to be known, and f_Y^* is the maxentropic probability density of the total losses , which we have already determined.

To exemplify our procedure, we shall use Case (1), characterized by the following parameters: $\ell = 3$, $\mu = 0$, $\sigma = 0.25$. Having determined f^* and computed $\phi(\alpha)$ as mentioned above, we can apply the SME and MEM procedures to obtain the probability density of individual losses. The resulting individual densities are shown in Figure (7). The comparison with the known probability density that was used to generate the total losses is another consistency test of the procedure. The true distributions of individual losses is a lognormal density with parameters $\mu = 0$, $\sigma = 0.25$, and is included is include

in Figure (7).

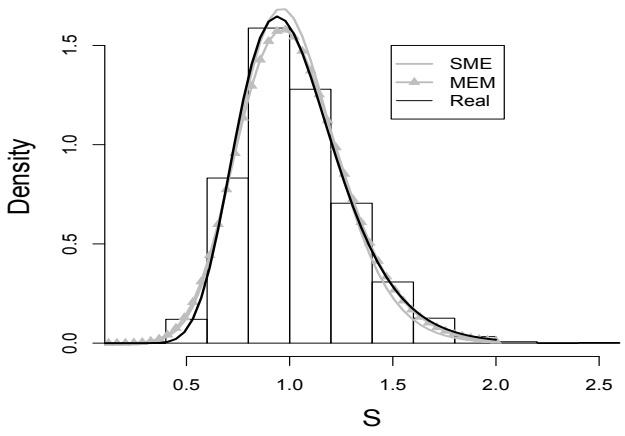


Figure 7: Density of the individual losses obtained by SME & MEM approaches
Case (1): $\ell = 3$, $\mu = 0$, $\sigma = 0.25$

In Tables (11) and (12) we present the results of simple measures of quality of reconstruction. The MAE and RMSE distances along with the L_1 and L_2 distances are shown in Tables (11) and (12). Here the SME and MEM reconstructions are compared with the true lognormal density and with the observed (simulated) data. We include the comparison between the reconstructed and true density to the data obtained by simulation, but keep in mind that such data might not be available. The true density is to be used as a benchmark to test the quality of the maxentropic procedures. Note that the best results are obtained for the reconstruction with MEM.

Approach	Hist. vs. Real density		Hist. vs. Maxent		Real density vs. Maxent	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
SME	0.0042	0.0047	0.0127	0.0257	0.0143	0.0232
MEM	0.0042	0.0047	0.0087	0.0105	0.0096	0.0113

Table 11: MAE and RMSE values of the individual losses calculated by SME & MEM

Approach	Hist. vs. Real density		Hist. vs. Maxent		Real density vs. Maxent	
	L1-norm	L2-norm	L1-norm	L2-norm	L1-norm	L2-norm
SME	0.164	0.1865	0.1886	0.1992	0.0679	0.0574
MEM	0.164	0.1865	0.1659	0.1818	0.0621	0.0624

Table 12: L_1 and L_2 distances of the individual losses calculated by SME & MEM

8 Appendix

GOODNESS-OF-FIT TESTS

There are a few test that help us determine whether the fitted model is appropriate. Here we recall briefly the basic facts about the tests that we mentioned above.

8.1 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test is a test of uniformity which verifies the differences in fit between the empirical distribution function (EDF) and the estimated (reconstructed) distribution function, using the largest absolute observed distance between them,

$$D_n = \underset{x}{sup}|F_n(x) - F(x)|$$

where: n: is the number of data points; F_n: is the empirical distribution function; $F_S(x)$: is the estimated distribution function.

Closely related to the KS statistic are the statistics

$$D_n^+ = \underset{x}{sup}|F_n(x) - F(x)| = \underset{1\leq i\leq n}{max}\left(\frac{i}{n} - F(x)\right)$$

$$D_n^- = \underset{x}{sup}|F(x) - F_n(x)| = \underset{1\leq i\leq n}{max}\left(F(x) - \frac{i-1}{n}\right)$$

They represent the largest positive D_n^+ and the largest negative D_n^- deviation. The KS statistic may be also be defined as $D_n = max(D_n^+, D_n^-)$. To eliminate the sample size dependence of the theoretical D distribution one may use $\sqrt{n}D_n$

The null hypothesis H_o of no difference between distributions, has to be rejected at the chosen significance level α of 0.1, 0.05 or 0.01 whenever $\sqrt{n}D_n > \sqrt{n}d_{\alpha,n}$ or if $p-value < \alpha$, where the critical value $d_{\alpha,n}$ and the p-value are calculated from the distribution of the K-S statistic when the null hypothesis is true. This is not a straightforward distribution, but could be obtained asymptotically or by simulation. In summary, the null hypothesis is rejected if $\sqrt{n}D_n$ is greater than 1.22, 1.36, 1.63 at the 90%, 95% and 99% confidence levels respectively.

A problem with this test is that KS statistic depends on the maximum difference, without considering the whole estimated distribution. This is important when the differences between distributions are suspected to occur only at the upper or lower end of their range. This may be problematic in small samples. Besides little is known about the impact of the departures from independence in the sample on the distribution of D_n, that is, if we are not sure about the independence of the sample, we would not be sure of the results of the test. In our data generation process, the independence is guaranteed. There exist other EDF tests, which in most situations are more effective than the simple KolmogorovSmirnov test. To finish, we mention that MAE is a tighter measure f distance between distribution functions that the D_n.

8.2 Anderson-Darling test

This is a more sophisticated version of the KS approach and it is based on the quadratic difference between $F_n(s)$ and $F^*(s)$. Here the AD statistic is computed as follows:

$$A_n^2 = n \int_{-\infty}^{\infty} |F_n(s) - F^*(s)|^2 \Psi(s) f(s) ds$$

where: $\Psi(x) = \frac{1}{F^*(x)(1-F^*(s))}$ is a weighting function; n is the number of data points; F_n is the empirical distribution function and $F^*(s)$ is the fitted distribution function;

When $\Psi(x) = 1$ the Anderson-Darling (AD) statistic reduces to the statistic which its today known as the Cramér-von Mises statistic. The test make maximum use of the observed data by integrating the vertical distances over all values of x, increasing its power balancing the variance between distributions through the use of $\Psi(x)$. The test of Anderson emphasizes more the tails of the distribution than the KS-test does.

The test statistic may be assessed against critical values in order to reject or not the H_o of uniformity. The null hypothesis rejects if A_n^2 is greater than a critical value of 2.492 and 3.857 at the 95% and 99% confidence levels respectively. For the case of Cramér-von Mises test, we reject the null hypothesis when the statistic is greater than 0.461 and 0.743 at the 95% and 99% confidence level respectively.

The Anderson-Darling (AD) statistic behaves similarly to the Cramér-von Mises statistic, but is more powerful to test whether $F^*(x)$ departs from the true distribution in the tails, especially when there appear to be many outlying values. For goodness-of-fit testing, departure in the tails is often important to detect, and A_n^2 is the recommended statistic.

8.3 Berkowitz back test

Berkowitz(2001) proposed the transformation $z_n = \Phi^{-1}(\int_{-\infty}^{s_n} f^*(s)ds) = \Phi^{-1}(F(s_n))$, to make the data iid standard normal under the null hypothesis. This allows to make use of powerful battery of available normality tests, instead of relying on uniformity tests. Besides that, Berkowitz back test provides a joint test of normality and independence.

The procedure consists of testing the null hypothesis of $\rho = \mu = 0$, $\hat{\sigma} = 1$, against a first-order autoregressive alternative ($z_t - \mu = \rho(z_{t-1} - \mu) + \varepsilon_t$) with mean and variance possibly different from (0,1). The LR test can be formulated as

$$LR_3 = -2(L(0,1,0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})) \tag{17}$$

where $L(\hat{\mu}, \sigma^2, \hat{\rho})$ is the likelihood as a function only of the unknown parameters of the model, the hats denote estimated values. The exact function associated with the first order autoregressive alternative is reproduced here by convenience.

$$\begin{aligned} L(\mu, \sigma^2, \rho) = &-\frac{1}{2}log(2\pi) - \frac{1}{2}log[\sigma^2/(1-\rho^2)] - \frac{(z_1 - \mu/(1-\rho))^2}{2\sigma^2/(1-\rho^2)} \\ &-\frac{T-1}{2}log(2\pi) - \frac{T-1}{2}log(\sigma^2) - \sum_{t=2}^{T}\left(\frac{(z_n - \mu - \rho z_{n-1})^2}{2\sigma^2}\right) \end{aligned} \tag{18}$$

where $\sigma^2 = VAR(\varepsilon_t)$. Under the null hypothesis, the test statistic is distributed as a $\chi^2(3)$. This means that we reject the null hypothesis when the statistic is greater than 7.815 and 11.34 at the 95% and 99% confidence levels, respectively.

It is usually recommended to supplement the Berkowitz test with at least one additional test for normality, for example Bera-Jarque test. This extra test ensures that we test for the predicted normal distribution, and not just for the predicted values of the parameters ρ, μ and σ.

8.4 Jarque-Bera test

The Jarque-Bera-test is a test of normality, whose statistic JB is defined by

$$JB = \left(\frac{n}{6}\right)\left(S^2 + \frac{(K-3)^2}{4}\right)$$

where the sample skewness is S=$\hat{\mu}_3/\hat{\mu}_2^{3/2}$ and the sample kurtosis is $K = \hat{\mu}_4/\hat{\mu}_2^2$, for $\{\hat{\mu}_j : j = 2, 3, 4\}$ respectively, are the second, third and fourth central moments, estimated by $\hat{\mu}_j = (1/n)\sum(x_i - \bar{x})^j$, $j = 2, 3, 4$, where n is the sample size.

JB is asymptotically chi-squared distributed with two degrees of freedom because JB is just the sum of squares of two asymptotically independent standardized normals. That means that H_o has to be rejected at level α if $JB \geq \chi^2_{1-\alpha,2}$, being H_o the null hypothesis which checks if the sample follows a normal random variable with unknown mean and variance. The critical values for reject the null hypothesis are 5.991 and 9.21 at the 95% and 99% confidence levels, respectively.

Unfortunately, for small samples the chi-squared approximation is too sensitive, often rejecting the null hypothesis when it is in fact true. A reason for this is that skewness and kurtosis are not independently distributed, and the sample kurtosis especially approaches normality very slowly. That is, since the test statistic converges too slowly to its limiting distribution, the test behave erratically in small and even in reasonably large samples.

One of the ways of dealing with, has been to use the bootstrap technique. By bootstrapping under the null hypothesis it is possible to approximate the distribution of the test statistic, thereby generating more robust critical values and p-values for the test statistic.

8.5 Correlograms

Tests like KS and AD does not prove independence, so to asses whether p_t is iid, we use a graphical tool, the correlogram. As we are interested not only in linear dependence but also in other forms of nonlinear dependence such as conditional heteroskedasticity, we examine the correlogram of $(p_t - \bar{p}_t)$ and the correlograms of $(p_t - \bar{p}_t)^2$ and $(p_t - \bar{p}_t)^3$. This

test is sometimes complemented with the LjungBox test, it tests the overall randomness based on a number of lags.

8.6 Reliability Diagram or QQ-plots

This plot serves to determine the quality of a fit by the proximity of the fitted curve to the diagonal, the closer the better the approximation, deviations from the diagonal gives the conditional bias. Additionally, this plot could indicates problems as overfitting, when the fitted curve lies below the diagonal line and underfitting when the fitted curve lies above the line.

To obtain the reliability diagram we need to find the fitted CDF (F^*) and the respective quantiles against the empirical one through a plotting formula, such as $F_n = (n - k + 0.5)/n$, here the n observations are ranked in descending order, being k the order of each data point. If the model is good, the points will lie very close to the line from 0 to 1.

8.7 Marginal Calibration Plot

Another useful and similar tool is the marginal calibration plot and is based in the idea that a system is marginally calibrated if estimations and observations have the same (or nearly the same) marginal distribution. Then, the graphical device is a plot of $F^* - F_n$ versus s_n, where F^* is the fitted or maxentropic CDF, and F_n is the empirical CDF of the observations. Under the hypothesis of marginal calibration, we expect minor fluctuations about zero. The same information can be visualized in terms of quantiles,

$$Q(F*, q) - Q(F_n, q), \quad q \in (0, 1)$$

of the functions $F*$ and F_n.

References

[1] Barzilai, J., & Borwein, J. M. (1988). *Two-point step size gradient methods*. IMA Journal of Numerical Analysis, **8**, 141-148.

[2] Berkowitz, J. (2001). *Testing density forecasts, with applications to risk management*. Journal of Business & Economic Statistics, **19**, 465-474.

[3] Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998) *Evaluating density forecasts with applications to financial risk management*, International Economic Review, **39**, 863-883.

[4] Gzyl, H., Novi-Inverardi, P.L. and Tagliani, A. (2013) *A comparison of numerical approaches to determine the severity of losses* Journal of Operational Risk, **8** pp.3-15.

[5] Gzyl, H., & Tagliani, A. (2012). Determination of the distribution of total loss from the fractional moments of its exponential. Applied Mathematics and Computation.

[6] Leipnik, R.B. (1991) *On Lognormal Random Variables: I The Characteristic Function*, Journal of the Australian Mathematical Society Series B, **32** pp.327347.

[7] Gzyl, H. and Velásquez, Y. *Linear Inverse Problems: The maximum entropy connection*, World Scientific Pubs, Singapore, (2011).

[8] Kapur, N. (199X) *Maximum Entropy Models in Science and Engineering* Wiley Interscience, New York.

[9] Thas, O. (2010). Comparing distributions. Berlin, Germany: Springer.

[10] Rockafellar, R.T, and Uryasev, S. (2000) *Optimization of conditional value at risk*, Journal of Risk, **2**, 21-41.

[11] Tay, A. S., and Wallis, K. F. *Density forecasting: A survey*. Journal of Forecasting, *19*, (2000): 235-254. Reprinted in Clements, M. P. and Hendry, D. F. (eds.) A Companion to Economic Forecasting,28 pp.45 - 68, Oxford: Blackwells (2002).

[12] Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. Journal of Business & Economic Statistics, 19(4), 465-474.

[13] Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts.

[14] Raydan M., On the Barzilai and Borwein choice of steplength for the gradient method, IMA Journal of Numerical Analysis, 13,321-326, (1993).