Bayesian Estimation of Lévy Copulas for Multivariate Operational Risks

Philipp Gebhard[1], Gernot Müller[2], Klaus Böcker[3]

June 11, 2010

[1]Centre of Mathematical Sciences, Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany, email: philipp.gebhard@mytum.de

[2]Centre of Mathematical Sciences, Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany, email: mueller@ma.tum.de

[3]Risk Integration & Reporting - Risk Analytics and Methods - UniCredit Group, Munich Branch, email: klaus.boecker@unicreditgroup.de

1 The need for structural dependence modelling in operational risk

Among the advanced measurement approaches (AMA) of Basel II [1] for calculating capital charges for operational risks, perhaps the most popular in the financial service industry is the loss distribution approach (LDA).

Let us briefly focus on one-dimensional operational risk modelling before we move on to the multivariate problem. Consider a single cell (business line/ loss event type) for which the total operational loss up to time horizon t is given by the aggregate loss process

$$S(t) = \sum_{k=1}^{N(t)} X_k \,, \quad t \geq 0 \,, \tag{1.1}$$

where $(X_k)_{k\in\mathbb{N}}$ is a sequence of independent and identically distributed (i.i.d.) positive random variables and $(N(t))_{t\geq 0}$ is a counting process in the time interval $[0,t]$, independent of $(X_k)_{k\in\mathbb{N}}$. The success of LDA models is, at least partially, owing to the fact that they allow for a very intuitive interpretation. Basically, a bank's total operational loss, e.g. within a given year, is simply the sum of all loss events observed during the year where, of course, both the number of losses and the loss severity of a single loss event are random. Hence, (1.1) can be considered as a *structural* model for univariate operational risk since it provides insight about how a yearly loss arises, here, by the accumulation of several random i.i.d. loss events[4]. For most practical applications, model (1.1) is even too general because in most cases the number of losses (or the loss frequency) is described by a homogeneous Poisson process with intensity $\lambda > 0$, so that (1.1) becomes a *compound Poisson process* (CPP). Interestingly, as pointed out by Böcker and Klüppelberg [2], the choice of the frequency process $(N(t))_{t\geq 0}$ has only very little impact on operational Value-at-Risk (VaR) as long as one is interested in extreme losses and high confidence levels (as it is typically the case). Therefore, one can correctly assume that univariate operational VaR figures calculated via models like (1.1) are *effectually all* equivalent to a one-dimensional compound Poisson model.

However, the world is more complex. According to Basel II, banks should distinguish between business lines (usually eight) and loss event types (usually seven), and this is exactly where the trouble comes in. For each cell $i = 1, \ldots, d$, the cumulated operational loss $S_i(\cdot)$ is described by a model of the form (1.1) with frequency and severity processes $N_i(\cdot)$ and $(X_k^i)_{k\in\mathbb{N}}$, respectively. Then these univariate losses $S_i(\cdot)$ have to be combined to compute the total operational risk, thereby trying to take the dependence structure between different cells into consideration. It is exactly this step – the modelling of the dependence structure – where most LDA approaches lose their appeal. The reason therefore is not that there are no possibilities of how the dependence structure can be introduced in

[4]It should be clear that we do not claim that model (1.1) is the only possible structural model for operational risk nor that its usage is particularly supported by empirical data. We only observe that these kind of models are very popular and have a long history in the finance industry. As a matter of fact, it is worth mentioning that (1.1) has neither been invented by the Basel committee nor the banking industry, instead it goes back to Filip Lundberg in 1903 in the context of insurance risk theory.

principle; quite the opposite is true. There are (too) many ways for doing this, e.g. by modelling the association between the number of losses occurring in different cells, between the severity distributions of different cells, or between the different cells' yearly aggregate loss distributions, see Böcker and Klüppelberg [3] and references therein. In most of these models the complexity increases dramatically (and thus the number of parameters to be estimated, which is particularly problematic when data are scarce), thereby destroying, or at least reducing, the benefits of the structural compound Poisson model presented in (1.1). The reason for this is that in such models dependence is understood as a measure of association between the severity and/or loss numbers and/or total loss random variables. For instance, by coupling (for a fixed $t \geq 0$, e.g. one year) the aggregated loss distributions $S_i(t)$ and $S_j(t)$ of two different cells by means of a distributional copula we may describe the statistical *phenomenon* of their association, however, such an approach does not provide any information that helps to explain the observed dependence.

In order to generalize the structural LDA model (1.1) and all its benefits to several dimensions, one has to switch from the usual distributional dependence model (no matter if number of losses, severities or aggregate losses are considered) to the *frequency domain*. This leads to the concept of *Lévy copulas*, see e.g. Cont and Tankov [4] and references therein.

2 Modelling multivariate operational risk with Lévy copulas

In this section we recall some definitions and properties of the multivariate operational risk model using Lévy copulas, see also Böcker and Klüppelberg [3] for a more detailed description of this approach. We assume that the cumulated operational loss $S_i(\cdot)$ in each cell $i = 1, \ldots, d$ is described by a CPP

$$S_i(t) = \sum_{k=1}^{N_i(t)} X_k^i \,, \quad t \geq 0 \,, \tag{2.1}$$

where $(X_k^i)_{k\in\mathbb{N}}$ for all i are i.i.d. and $(N_i(t))_{t\geq 0}$ are independent of $(X_k^i)_{k\in\mathbb{N}}$ with

$$P(N_i(t) = n) = e^{-\lambda_i t} \, \frac{(\lambda_i t)^n}{n!}, \quad n \in \mathbb{N}_0 \,,$$

where the rate $\lambda_i > 0$ is also referred to as frequency parameter. We further assume that $\mathbf{S}(t) := (S_1(t), \ldots, S_d(t))_{t\geq 0}$ constitutes a d-dimensional compound Poisson process, which implies that also the bank's total operational risk

$$S^+(t) := S_1(t) + S_2(t) + \cdots + S_d(t) \,, \quad t \geq 0 \,, \tag{2.2}$$

constitutes a one-dimensional CPP. CPP belong to a class of more general stochastic processes, the so called *Lévy processes*, which are processes in $\mathbb{R}^d$ with independent and stationary increments. If a Lévy process has only positive jumps, it is called a spectrally

positive Lévy process. Since operational losses are assumed to be positive, our CPP $\mathbf{S}(\cdot)$ is a special spectrally positive Lévy process.

The dependence structure between spectrally positive Lévy processes can be modelled by means of Lévy copulas. In contrast to distributional copulas, which are defined in the domain of probability measures, Lévy copulas are defined in the domain of *Lévy measures* and *tail integrals*. Basically, a Lévy measure controls the jump behaviour of a Lévy process. For instance, the Lévy measure Π_i of a univariate CPP $S_i(\cdot)$ can be written in terms of the rate (frequency) $\lambda_i > 0$ and the jump size distribution function (loss severity), namely $\Pi_i([0,x)) := \lambda_i P(X^i \leq x) = \lambda_i F_i(x)$ for $x \in [0,\infty)$. Hence, the Lévy measure of a single operational risk cell measures the expected number of losses per unit time with a loss amount in a pre-specified interval. Related to the concept of a Lévy measure is that of a tail integral. A one-dimensional tail integral is simply the expected number of losses, per unit time, that are above a given threshold x,

$$\overline{\Pi}_i(x) := \Pi_i([x,\infty)) = \lambda_i P(X^i > x) = \lambda_i \overline{F}_i(x)\,, \quad x \in [0,\infty)\,. \tag{2.3}$$

Similarly, the multivariate Lévy measure controls the joint jump behaviour (per unit time) of all univariate components and contains all information of dependence between the components. Finally, the multivariate tail integral is basically given by

$$\overline{\Pi}(x_1,\ldots,x_d) = \Pi([x_1,\infty)\times\cdots\times[x_d,\infty))\,, \quad (x_1,\ldots,x_d) \in [0,\infty]^d\,. \tag{2.4}$$

As a multivariate distribution can be built from marginal distributions via a distributional copula, a multivariate tail integral (2.4) can be constructed from the marginal tail integrals (2.3) by means of a Lévy copula. This is the content of Sklar's theorem for Lévy processes with positive jumps, which basically says that every multivariate tail integral $\overline{\Pi}$ can be decomposed into its marginal tail integrals and a Lévy copula $\mathfrak{C}$ according to

$$\overline{\Pi}(x_1,\ldots,x_d) = \mathfrak{C}(\overline{\Pi}_1(x_1),\ldots,\overline{\Pi}_d(x_d))\,, \qquad (x_1,\ldots,x_d) \in [0,\infty]^d\,. \tag{2.5}$$

For a precise formulation of this theorem we refer to Cont and Tankov [4], Theorem 5.6.

We can therefore conclude that the multivariate compound Poisson model based on Lévy copulas is a very natural and straight-forward extension of the well-known univariate CPP model (1.1) to several dimensions. The loss frequency *and* the severity (both encoded in the tail integral of the CPP) between different cells are modelled by the same concept, namely that of a Lévy copula. This results in a model with comparably few parameters, making it particularly advantageous in case of rare data. Furthermore, we want to stress that Lévy copulas allow for an intuitive and structural explanation of what 'dependence' in operational risk actually can be thought of: Dependence means that losses in different cells occur at the same time. More precisely, independence means that losses in different cells never occur at the same time and that their loss severity variables are also independent. Complete positive dependence, on the other hand, means that losses always occur at the same points in time and that the loss severity variables also have a perfect positive dependence structure (comonotonicity).

Finally, we should remark that Lévy copulas give a *dynamic* description of the dependence structure of a Lévy process $\mathbf{S}(\cdot)$ in contrast to static models where the *distributional*

dependence between the marginals of $\mathbf{S}(\cdot)$ for a predetermined and fixed $t \geq 0$ is considered. Since operational losses occur in time, a statical dependence model can never reflect coincidence of losses in different cells caused e.g. by the same catastrophic event such as Hurricane Katrina or the break-down of a bank wide computer system. This is also acknowledged by the regulators who, by assuming a statical dependence model, demand that losses which affect different cells but which are caused by one and the same event are NOT counted as several small losses (simultaneously happening in different cells) but rather as one single big loss impacting only a single cell. The reason is that a statical model would 'forget' that these losses actually have the same origin, and it would falsely treat them as independent events instead of one single, perhaps disastrous, incident. Of course, in the framework of Lévy copulas suggested here, this artificial correction is not necessary, because coincident losses are properly reflected in the dependence model itself.

3 The bivariate Clayton model

A bivariate model is particularly useful to illustrate how Bayesian statistics can be used to estimate the parameters of CPP, especially those of the Lévy copula. Therefore, we now focus on two operational risk cells (index $i = 1, 2$) with frequency parameters λ_i and severity distributions F_i, so that the marginal tail integrals are given by $\overline{\Pi}_i(\cdot) = \lambda_i \overline{F}_i(\cdot)$ as explained in (2.3).

3.1 The general Clayton model

In principle, the functions F_i, $i = 1, 2$, can be any parametric distribution function that is considered as an appropriate choice for modelling loss severities. In operational risk these are typically (one-sided) heavy-tailed distributions that can handle big losses.

In the following, we model the dependence structure between the two processes $S_1(\cdot)$ and $S_2(\cdot)$ using the *Clayton Lévy copula* with parameter $\delta > 0$,

$$\mathfrak{C}(u, v) = \left(u^{-\delta} + v^{-\delta}\right)^{-1/\delta}, \qquad u, v > 0\,.$$

This copula covers the whole range of positive dependence: For $\delta \to 0$ we obtain independence of the marginal processes given by $\mathfrak{C}_\perp(u, v) = u1_{v=\infty} + v1_{u=\infty}$, and losses in different cells never occur at the same time. For $\delta \to \infty$ we get the complete positive dependence Lévy copula given by $\mathfrak{C}_\parallel(u, v) = \min(u, v)$, and losses always occur at the same points in time. By varying δ, the cell dependence changes smoothly between these two extremes.

For a Clayton model where the jump sizes are described by the distributions F_1 and F_2, we will write F_1 - F_2 - Clayton model in the following. If $F_1 = F_2 = F$, we will use the notation $(F)^2$ - Clayton model.

Our goal is to determine the likelihood function for a bivariate CCP $(S_1(\cdot), S_2(\cdot))$ based on the observed loss times and loss severities in both components. In doing so, it is useful to decompose the marginal components S_i (representing the cells' aggregate loss processes) into a jump dependent part and an independent part (see Böcker and Klüppelberg [3],

and Esmaeili and Klüppelberg [6] for details about the subsequent results),

$$S_1 = \sum_{i=1}^{N_1} X_i = S_1^{\perp} + S_1^{\parallel} = \sum_{k=1}^{N_1^{\perp}} X_k^{\perp} + \sum_{m=1}^{N^{\parallel}} X_m^{\parallel}\,, \tag{3.1}$$

$$S_2 = \sum_{j=1}^{N_2} Y_j = S_2^{\perp} + S_2^{\parallel} = \sum_{l=1}^{N_2^{\perp}} Y_l^{\perp} + \sum_{m=1}^{N^{\parallel}} Y_m^{\parallel}\,, \tag{3.2}$$

where $S_1^{\perp}$ and $S_2^{\perp}$ are independent (no joint losses) and $(S_1^{\parallel}, S_2^{\parallel})$ describe aggregate losses of both cells that are caused by the same event and thus always happen together. Moreover, $N_1^{\perp}(\cdot)$, $N_2^{\perp}(\cdot)$ and $N^{\parallel}(\cdot)$ are Poisson processes with intensities $\lambda_1^{\perp}, \lambda_2^{\perp}$ and $\lambda^{\parallel} > 0$, respectively, for which it can be shown that

$$\lambda^{\parallel} = \mathfrak{C}\left(\lambda_1, \lambda_2\right), \qquad \lambda_i^{\perp} = \lambda_i - \lambda^{\parallel}\,, \qquad i = 1, 2\,.$$

In particular, we obtain for the Clayton Lévy copula

$$\lambda^{\parallel} = (\lambda_1^{-\delta} + \lambda_2^{-\delta})^{-1/\delta}\,, \tag{3.3}$$

i.e. the frequency of simultaneous losses is a simple function of the Clayton copula parameter δ. This could be used for expert judgement about δ by eliciting $\lambda^{\parallel}$, or to compute an empirical estimate for δ based on observations of λ_1, λ_2 and $\lambda^{\parallel}$.

Due to (3.1) and (3.2) one can write the likelihood function for the bivariate process (S_1, S_2) as the product of the likelihoods for the processes $S_1^{\perp}$, $S_2^{\perp}$ and $(S_1^{\parallel}, S_2^{\parallel})$. For the convenience of the reader, we briefly describe the likelihood of a univariate CPP. Assume that within the time interval $0 < t < T$ one observes n jumps at times $t_1, \ldots, t_n$, each with jump size $x_i \sim f(\cdot|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector. Defining the inter-arrival times $T_i = t_i - t_{i-1}$ for $i = 1, \ldots, n$ with $t_0 = 0$ and recalling that the T_i are i.i.d. exponential random variables with parameter λ, we can write the likelihood as

$$\begin{aligned} f(x_i, n|\lambda, \boldsymbol{\theta}) &= e^{-\lambda(T-t_n)} \prod_{i=1}^{n} \lambda e^{-\lambda T_i} \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \\ &= \lambda^n e^{-\lambda T} \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})\,. \end{aligned} \tag{3.4}$$

The last term in (3.4) is the likelihood for the observed jump sizes, the part in the middle is the likelihood for the observed inter-arrival times, and the first factor is simply the probability that there are no jumps within $[t_n, T]$. Note that the jump size density for the jump dependent process $(S_1^{\parallel}, S_2^{\parallel})$ not only depends on the marginal jump size densities f_1 and f_2 but also on the Lévy copula and its parametrisation. Having said that, the entire likelihood for the bivariate CPP can be set-up straight forwardly. Let $\boldsymbol{\theta}_1$ be the parameter vector of the marginal density f_1 and $\boldsymbol{\theta}_2$ the parameter vector of the marginal density f_2, then the general result is given in Esmaeili and Klüppelberg [6],

$$
\begin{aligned}
& f\left(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{x}, \mathbf{y}, n_1^{\perp}, n_2^{\perp}, n^{\parallel} \mid \lambda_1, \boldsymbol{\theta}_1, \lambda_2, \boldsymbol{\theta}_2, \delta\right) \\
&= (\lambda_1)^{n_1^{\perp}} e^{-\lambda_1^{\perp} T} \prod_{i=1}^{n_1^{\perp}} \left[f_1\left(\tilde{x}_i; \boldsymbol{\theta}_1\right) \left(1 - \frac{\partial}{\partial u} \mathfrak{C}\left(u, \lambda_2; \delta\right) \Big|_{u=\lambda_1 \overline{F}_1(\tilde{x}_i; \boldsymbol{\theta}_1)}\right)\right] \\
&\times (\lambda_2)^{n_2^{\perp}} e^{-\lambda_2^{\perp} T} \prod_{i=1}^{n_2^{\perp}} \left[f_2\left(\tilde{y}_i; \boldsymbol{\theta}_2\right) \left(1 - \frac{\partial}{\partial v} \mathfrak{C}\left(\lambda_1, v; \delta\right) \Big|_{u=\lambda_2 \overline{F}_2(\tilde{y}_i; \boldsymbol{\theta}_2)}\right)\right] \qquad (3.5) \\
&\times (\lambda_1 \lambda_2)^{n^{\parallel}} e^{-\lambda^{\parallel} T} \prod_{i=1}^{n^{\parallel}} \left[f_1\left(x_i; \boldsymbol{\theta}_1\right) f_2\left(y_i; \boldsymbol{\theta}_2\right) \frac{\partial^2}{\partial u \partial v} \mathfrak{C}\left(u, v; \delta\right) \Big|_{u=\lambda_1 \overline{F}_1(x_i; \boldsymbol{\theta}_1), v=\lambda_2 \overline{F}_2(y_i; \boldsymbol{\theta}_2)}\right],
\end{aligned}
$$

where $\tilde{\mathbf{x}}$, $\tilde{\mathbf{y}}$ are the observed independent jump sizes of S_1, S_2, respectively, and $(\mathbf{x}, \mathbf{y})$ are the joint jump sizes of $(S_1^{\parallel}, S_2^{\parallel})$. Furthermore, $n_1^{\perp}$, $n_2^{\perp}$ are the number of the independent jumps in S_1, S_2, respectively, and $n^{\parallel}$ is the number of joint jumps of $(S_1^{\parallel}, S_2^{\parallel})$.

In case of the Clayton Lévy copula we can further specify

$$
\begin{aligned}
\frac{\partial}{\partial u} \mathfrak{C}\left(u, v\right) &= \left(1 + \left(\frac{u}{v}\right)^{\delta}\right)^{-1/\delta - 1}, \qquad (3.6) \\
\frac{\partial^2}{\partial u \partial v} \mathfrak{C}\left(u, v\right) &= (\delta + 1) (uv)^{\delta} \left(u^{\delta} + v^{\delta}\right)^{-1/\delta - 2}, \qquad u, v > 0\,.
\end{aligned}
$$

For notational convenience we set $\mathbf{z} := \left(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{x}, \mathbf{y}, n_1^{\perp}, n_2^{\perp}, n^{\parallel}\right)$, $\boldsymbol{\psi} := (\lambda_1, \boldsymbol{\theta}_1, \lambda_2, \boldsymbol{\theta}_2, \delta)$. Given the likelihood function $f(\mathbf{z}|\boldsymbol{\psi})$ for a realization $\mathbf{z}$ of the process, it is straightforward to compute maximum likelihood estimates of the parameters $\boldsymbol{\psi}$, see again Esmaeili and Klüppelberg [6]. In the following, however, we use a Bayesian setting which enables us to derive uncertainty bounds for the estimates in a very natural way. A brief introduction to Bayesian statistics is given in Chapter 1 by Robert and Rousseau in this book.

3.2 The (Burr/GPD)^2 - Clayton model

To find a Clayton model that appropriately describes the fire insurance data, which we will analyze later in this Chapter, we need to select suitable classes of distributions for the jump sizes in the bivariate process. Since the losses are naturally bounded below by zero and since there are several big losses in both components, right-sided heavy-tailed distributions seem to be a reasonable choice. After having fitted several Clayton models to the data set — including Burr, Gamma, Loggamma, Lognormal, truncated Normal, Pareto and Weibull models as well as combinations of them — it turned out that, following a Bayes factor analysis described at the end of this Chapter, the (Burr/GPD)^2 - Clayton model is the best choice. According to the notation introduced in the previous Subsection, the (Burr/GPD)^2 - Clayton model is a bivariate Clayton model where the jump sizes of both components are modelled by a sliced Burr/GPD distribution, each. For $i = 1, 2$, it is composed by a truncated Burr distribution with parameters $c_i, k_i > 0$ and density

$$
f(z_i) \propto c_i k_i z_i^{c_i - 1} \left(1 + z_i^{c_i}\right)^{-(k_i + 1)}, \qquad 0 < z_i \leq u_i\,,
$$

and a truncated generalized Pareto distribution (GPD) with parameters $h_i, \beta_i, \xi_i > 0$ and density

$$f(z_i) \propto \frac{1}{\beta_i}\left(1+\xi_i\frac{z_i+h_i-u_i}{\beta_i}\right)^{-1/\xi_i-1}, \qquad z_i > u_i\,,$$

for a fixed $u_i > 0$. For $z_i > 0$ the sliced density is then given by

$$f_i\left(z_i;\boldsymbol{\theta}_i\right) = A_i\left(\mathbf{1}_{(0,u_i]}(z_i)\, c_i k_i z_i^{c_i-1}\left(1+z_i^{c_i}\right)^{-(k_i+1)} + \mathbf{1}_{(u_i,\infty)}(z_i)\,\frac{1}{\beta_i}\left(1+\xi_i\frac{z_i+h_i-u_i}{\beta_i}\right)^{-1/\xi_i-1}\right),$$

where $\boldsymbol{\theta}_i = (c_i, k_i, h_i, \beta_i, \xi_i)$ and the normalization constant calculates as

$$A_i := \left(1-\left(1+u_i^{c_i}\right)^{-k_i}+\left(1+\xi_i\frac{h_i}{\beta_i}\right)^{-1/\xi_i}\right)^{-1},$$

due to the truncation.

We see that the transition between the distributions is executed at the fixed thresholds $u_i > 0$, $i = 1, 2$. That is to say, we apply the Burr distribution for $z_i \leq u_i$, whereas for $z_i > u_i$ the GPD is used. We choose these transition values to be the empirical 90% quantiles (e.g. for the Danish fire insurance data we have $u_1 = 3.96$ and $u_2 = 3.90$). More details on how to determine these thresholds appropriately can be found in Embrechts et al. [5], Section 6.5.

Given the densities from above, the tail distribution for $z_i > 0$, $i = 1, 2$, can be calculated as

$$\overline{F}_i(z_i;\boldsymbol{\theta}_i) = A_i\left(\mathbf{1}_{(0,u_i]}(z_i)\left(A_i^{-1}-1+\left(1+z_i^{c_i}\right)^{-k_i}\right)+\mathbf{1}_{(u_i,\infty)}(z_i)(1+\xi_i\frac{z_i+h_i-u_i}{\beta_i})^{-1/\xi_i}\right).$$

Together with Equations (3.6) the corresponding likelihood function for the bivariate CPP can now be written as in (3.5).

According to the Bayesian method, this likelihood function $f(\mathbf{z}|\boldsymbol{\psi})$ has then to be combined with an appropriate prior distribution $\pi(\boldsymbol{\psi})$ to obtain the posterior distribution,

$$f(\boldsymbol{\psi}|\mathbf{z}) \propto f(\mathbf{z}|\boldsymbol{\psi})\pi(\boldsymbol{\psi})\,.$$

The prior distribution may reflect available prior knowledge about the model parameters and we will give more details about their specification later in this Chapter when we are analyzing the empirical data set. In the following we describe the Markov chain Monte Carlo (MCMC) method used for sampling from the posterior distribution $f(\boldsymbol{\psi}|\mathbf{z})$.

4 The MCMC algorithm

In this section we develop a MCMC method to fit the Clayton models (particularly the (Burr/GPD)^2 - Clayton model) to an empirical data set, which will enable us to approximate the marginal posterior distributions for the thirteen parameters of interest.

4.1 Gibbs sampling

MCMC algorithms can achieve higher levels of efficiency when they take into account the specifics of the target density. Since this can be done more precisely when determining multiple one-dimensional proposal densities instead of one multi-dimensional proposal density, we employ the Gibbs sampler.

Denoting by I the total number of iterations, the structure of our MCMC sampler is classical, using thirteen univariate updates from the full conditional distributions:

1. Specify initial values $\lambda_1^{(0)}$, $\boldsymbol{\theta}_1^{(0)}$, $\lambda_2^{(0)}$, $\boldsymbol{\theta}_2^{(0)}$, $\delta^{(0)}$.

2. Repeat for $i = 1, ..., I$:

   - Sample $\lambda_1^{(i)}$ from $f(\lambda_1|\mathbf{z}, \boldsymbol{\theta}_1, \lambda_2, \boldsymbol{\theta}_2, \delta)$.
   - Sample $c_1^{(i)}$ from $f(c_1|\mathbf{z}, \lambda_1, k_1, h_1, \beta_1, \xi_1, \lambda_2, \boldsymbol{\theta}_2, \delta)$.
   - Sample $k_1^{(i)}$ from $f(k_1|\mathbf{z}, \lambda_1, c_1, h_1, \beta_1, \xi_1, \lambda_2, \boldsymbol{\theta}_2, \delta)$.

     $\vdots$

   - Sample $\xi_2^{(i)}$ from $f(\xi_2|\mathbf{z}, \lambda_1, \boldsymbol{\theta}_1, \lambda_2, c_2, k_2, h_2, \beta_2, \delta)$.
   - Sample $\delta^{(i)}$ from $f(\delta|\mathbf{z}, \lambda_1, \boldsymbol{\theta}_1, \lambda_2, \boldsymbol{\theta}_2)$.

The vectors $\boldsymbol{\psi}^{(i)} := (\lambda_1^{(i)}, \boldsymbol{\theta}_1^{(i)}, \lambda_2^{(i)}, \boldsymbol{\theta}_2^{(i)}, \delta^{(i)})$ are then considered to be samples from the posterior distribution. Looking at Equation (3.5) it gets clear immediately that all full conditional distributions appearing above are non-standard. Hence, Metropolis-Hastings (MH) steps have to be used and the question arises how to choose the required proposal densities. Further information about useful computational methods in the context of Bayesian statistics can be found in the second Chapter by Robert and Marin in this volume.

4.2 Choice of the proposal densities

As usual, the selection of good proposal densities for the MH steps is one of the most important choices one has to make in the setup of a MCMC procedure. The better a proposal density mimics the target density, the better the mixing of the produced MCMC chains will be.

Since we use thirteen univariate updates, we have to select thirteen proposal densities for the full conditionals $f\left(\psi_k \mid \boldsymbol{\psi}_{-k}, \mathbf{z}\right)$, $k = 1, \ldots, 13$. To get an idea of the shape of the full conditionals, it is useful to simulate data from the model and to plot the likelihood function (3.5) given the data, where twelve of the thirteen parameters are fixed to the known values from the simulation. For various parameter sets and simulations, these marginal densities looked unimodal and quite symmetric so that we use for all thirteen parameters normal proposals with different means μ_k and variances σ_k^2, $k = 1, \ldots, 13$.

To achieve good acceptance rates for the MH steps, μ_k and σ_k^2 are adapted several times during the sampling procedure. The parameters μ_k, $k = 1, \ldots, 13$, are set to the values which maximize the current full conditional $f\left(\psi_k \mid \boldsymbol{\psi}_{-k}, \mathbf{z}\right)$, each. Finding these maxima is a univariate optimization problem each and requires only few computation

time. The standard deviations σ_k, $k = 1, \ldots, 13$, are specified using a special property of the normal distribution. Denoting by $g(x; \mu, \sigma^2)$ the density of the normal distribution with mean μ and variance σ^2 evaluated at x, one easily derives that

$$g''\left(x; \mu, \sigma^2\right) = \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\left(x-\mu\right)^2 - 1\right) g\left(x; \mu, \sigma^2\right).$$

Hence, $\sigma^2 = -g(\mu; \mu, \sigma^2)/g''(\mu; \mu, \sigma^2)$, and a sensible choice for the variance of the proposal density is $\sigma_k^2 = -f(\psi_k|\boldsymbol{\psi}_{-k}, \mathbf{z})|_{\mu_k}/f''(\psi_k|\boldsymbol{\psi}_{-k}, \mathbf{z})|_{\mu_k}$ which can be evaluated easily using numerical approximations of the second derivative of $f(\cdot|\boldsymbol{\psi}_{-k}, \mathbf{z})$ at μ_k.

It is not necessary to adapt μ_k and σ_k^2 in each iteration of the MCMC sampler, except of the burn-in phase, where adaption in each iteration improves the mixing of the sampler significantly.

4.3 Initial values, burn-in, number of iterations, subsampling

Besides the choice of the proposal densities the choice of good starting values plays an important role. In our case, we can use maximum likelihood estimates since the maximization of the likelihood given in Equation (3.5) is a thirteen dimensional problem which can usually be handled by statistical software. Therefore, we will always use ML estimates as initial values, if not stated otherwise.

From several simulation studies we found that the convergence behaviour of the sampler is very satisfying. Even if the initial values are chosen quite far away from the parameters used as an input for simulation, the chain converges fast and a burn-in period of 1000 iterations is always sufficient.

The last aspect we want to address is the subsampling of simulated values. As we will see in the next section, the produced Markov chains for the thirteen parameters of interest show a significant autocorrelation over several time lags. One reason for this is obviously the MH algorithm itself, which allows to replicate the current value when the proposed value is rejected. To reduce the dependence between the samples from the MCMC algorithm we use subsampling. Since the computation time in our examples is only a few minutes for 1000 iterations, we will use only every 10th iteration. Therefore, in order to get 2000 samples from the posterior distribution we run the algorithm always for 21000 iterations (including the burn-in period) and then use iterations $1010, 1020, \ldots, 21000$.

5 Simulation study

We now assess the mixing and convergence behaviour of the chains and the quality of the posterior mean estimates using simulated data. To simulate paths of the CPP from the (Burr/GPD)^2 - Clayton model, we employ the algorithm proposed by Esmaeili and Klüppelberg in [6] which simulates the jump times and the jump sizes independently.

Throughout, we always use the parameters $\lambda_1 = 34$, $c_1 = 4.1$, $k_1 = 0.42$, $h_1 = 7.3$, $\beta_1 = 3.8$, $\xi_1 = 0.42$, $\lambda_2 = 26$, $c_2 = 1.2$, $k_2 = 1.9$, $h_2 = 16$, $\beta_2 = 8.4$, $\xi_2 = 0.18$ and $\delta = 1.8$ for simulation. These values are motivated by the analysis of the fire insurance data in the next Section. Similarly, we set $T = 12$ where a time unit corresponds to one month.

In general, the observed number of jumps in these simulations is random, but it is always around 480.

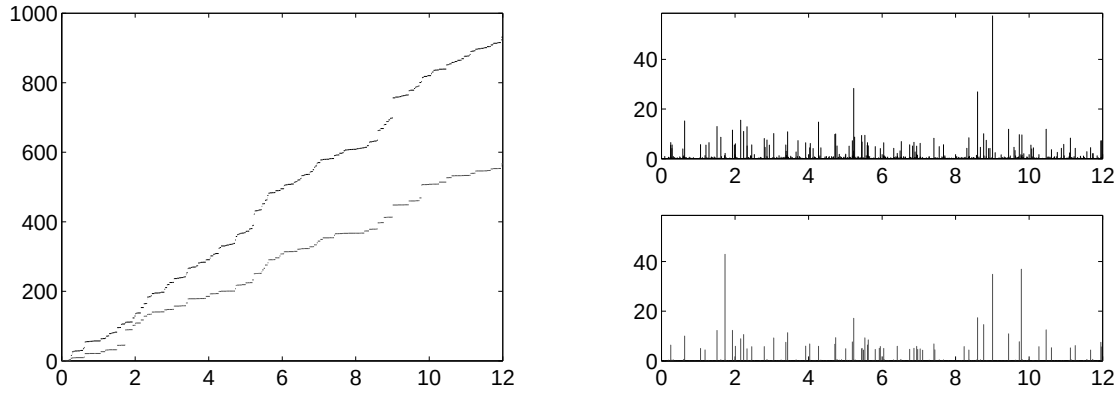
For all thirteen parameters we choose independent Gamma[5] prior distributions $\Gamma(p_k, b_k)$, $k = 1, \ldots, 13$. Note that we always use the parametrization where the Gamma density is given by

$$f_\Gamma(x; p, b) = \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} 1_{[0,\infty)}(x)\,, \qquad p,\, b > 0\,,$$

with mean p/b and variance p/b^2. In our analysis, the hyperparameters p_k and b_k are chosen in such a way that the prior means correspond to the initial ML estimates and the standard deviations are about one fourth of the means. It should be mentioned, however, that in a pure Bayesian approach the hyperparameters would typically be specified by means of expert elicitation as it is described in the Chapter by Oakley in this book.

5.1 Illustrative example

In this section we illustrate the MCMC algorithm using the simulated data set shown in Figure 1. Observable are the single jump sizes $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ of both components, the jump sizes $\mathbf{x}$ and $\mathbf{y}$ of the joint jumps, the numbers $n_1^\perp$ and $n_2^\perp$ of single jumps in both components and the number $n^\parallel$ of joint jumps.



(a) Sample paths

(b) Same paths as marked point processes

Figure 1: Simulation of a bivariate CPP in the (Burr/GPD)^2 - Clayton model over a time interval of 12 months.

Now we apply the Bayesian method using MCMC to recover the marginal posterior distributions of all parameters used for simulation. As stated above the ML estimates are a convenient choice for starting values; they help to shorten the burn-in period.

The means of the independent Gamma priors are calculated as 32, 4.2, 0.41, 4.7, 3.2, 0.72, 26, 1.3, 2.1, 16, 12, 0.17 and 2.3, respectively, for the parameters λ_1, c_1, k_1, h_1, β_1, ξ_1, λ_2, c_2, k_2, h_2, β_2, ξ_2 and δ. The standard deviations are chosen to be 8.0, 1.0, 0.10, 1.2,

[5]Let us stress that this explicit choice is exemplary. Any appropriate distribution can be used as prior instead.

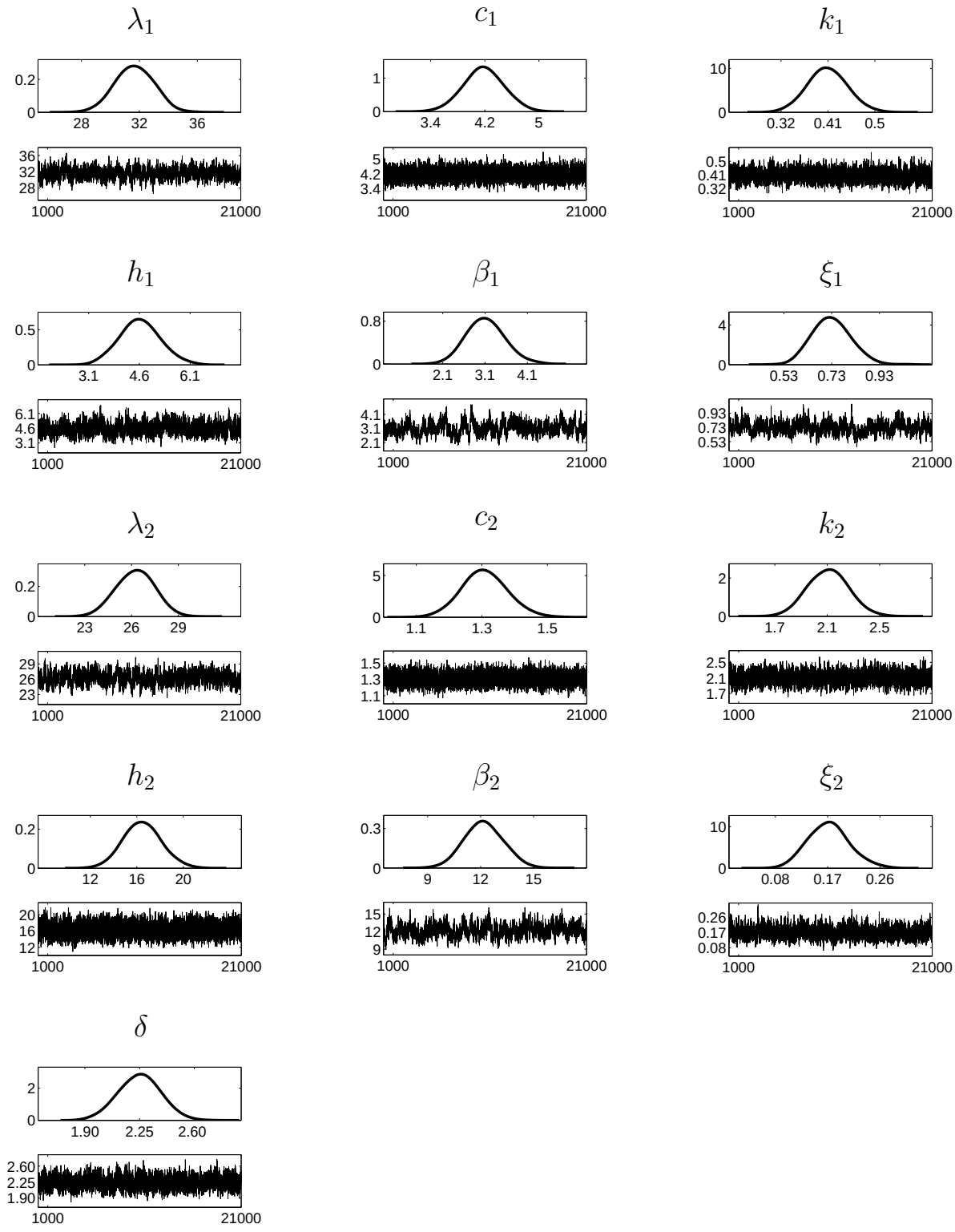


Figure 2: Density estimates of marginal posterior distributions and sample paths for the parameters of the (Burr/GPD)^2 model. Simulated data and independent Gamma priors are used.

0.80, 0.18, 6.5, 0.33, 0.53, 4.1, 3.0, 0.043 and 0.57, respectively. Figure 2 shows, for each parameter of the (Burr/GPD)^2 - Clayton model, the marginal posterior distribution and the produced sample path.

The marginal posterior distributions are all unimodal with peaks which are very close to the ML estimates and the distributions are all quite symmetric. Considering the sample paths we see that that the mixing behaviour of our MCMC sampler is very satisfying. We notice that there are small differences in the evolution of the paths for the different parameters. Particularly for β_i, ξ_i, $i = 1, 2$, which determine the tails of the distributions, the chain does not mix as fast as it does for the other parameters; in consequence the autocorrelations are higher. Hence, we decided to use subsampling and to take only every 10th iteration in order to reduce the autocorrelations.

Repeating the analysis with different initial values, also very poor ones, has shown that the produced Markov chains converge very fast, usually within 500 iterations. Thus, a burn-in period of 1000 iterations is sufficient, and we use iterations $1010, 1020, 1030, \ldots$ to derive samples from the posterior distribution.

5.2 Quality of the posterior mean estimates

To assess the quality of the posterior mean estimates we repeat the analysis of the previous section 100 times for uniform priors. More precisely, we simulate 100 data sets, using always the same fixed parameter set as before. We fit the (Burr/GPD)^2 - Clayton model to these 100 data sets by MCMC to get posterior mean estimates for the parameters for all 100 data sets.

Table 1 gives the means and the standard deviations of the posterior mean estimates for these 100 data sets, for all thirteen parameters. It shows that on average the posterior mean estimates match the simulation values very well and that the standard deviations are reasonably small, except for the GPD parameters (h_i, β_i, ξ_i, $i = 1, 2$). Of course, this is due to the small number of observations in the tail. When one would consider larger data sets, the posterior mean estimates would become more precise. Overall, however, we can conclude that the performance of our sampler is very satisfying.

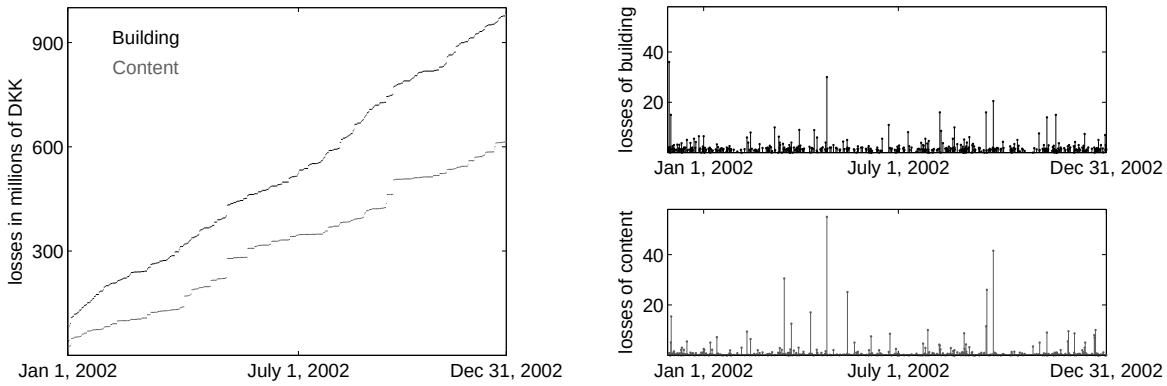
	λ_1	c_1	k_1	h_1	β_1	ξ_1	
True value	34	4.1	0.42	7.3	3.8	0.42	
Mean	34.2	4.08	0.409	7.38	3.93	0.438	
Std	1.72	0.320	0.0313	1.84	1.08	0.143	
	λ_2	c_2	k_2	h_2	β_2	ξ_2	δ
True value	26	1.2	1.9	16	8.4	0.18	1.8
Mean	26.1	1.26	1.97	16.5	8.68	0.173	1.88
Std	1.79	0.0726	0.150	3.18	1.99	0.0475	0.211

Table 1: Means and standard deviations of the posterior mean estimates in the (Burr/GPD)^2 - Clayton model for 100 simulations.

6 Analysis of fire insurance data

To further exemplify our approach, we now analyze the Danish fire insurance data. We choose this specific data set, because empirical data for operational risks is hardly available. However, analyzing a data set from operational risk is performed exactly as it is done here.

The data were collected at Copenhagen Reinsurance, an aggregated form of them appears for example in Embrechts et al. [5]. The losses of profit, which are contained in this set, are not considered, i.e. we restrict the data to the bivariate data set consisting of losses of buildings and losses of contents. The claims are reported in millions of Danish Kroner. Our analysis comprises in total 474 observations collected in the year 2002.



(a) Sample paths

(b) Marked point processes

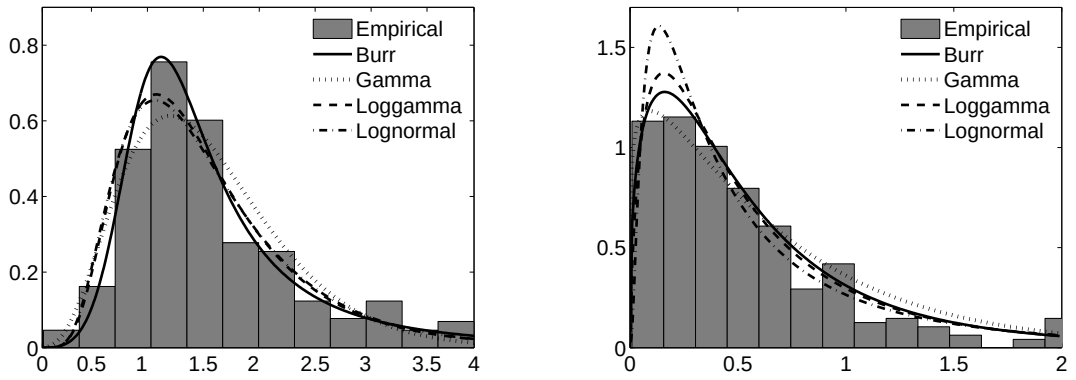
Figure 3: The Danish fire insurance data: Observed sample paths of the accumulated losses for buildings and contents, and their representation as marked point processes.

Figure 3 shows the observed sample paths of the accumulated losses for buildings and contents as well as their representation as marked point processes. Table 2 summarizes some important features.

To work with sliced distributions provides quite a flexibility in modelling data sets. Whereas for the tails both of the building and the content component a GPD seems to be a reasonable choice, it is not so clear a priori which distribution could be appropriate for the smaller losses. Hence, Figure 4 shows histograms of the small losses in the building and the content component, respectively. Ignoring all losses greater than 4 million and 2 million DKK, respectively, we now fit several standard distributions - Burr, Gamma, Loggamma, Lognormal - to these smaller losses, by maximum likelihood. As we can see from Figure 4, the fit both of the Burr distribution and the Loggamma distribution is quite satisfying. Hence, in the following we concentrate on these two families to model the body of the loss distributions. Note, that this initial investigation is not linked to Bayesian analysis. It just helps to select appropriate distributions. Of course, later on, the chosen models should be compared using model selection criteria as, e.g., Bayes factors.

	single jumps	common jumps	minimal loss	maximal loss	average loss
		period January to December 2002			
building	131	308	70,000	36,000,000	2,076,335
content	35	308	9,000	55,000,000	1,286,908
		period July to December 2002			
building	71	147	70,000	20,500,000	2,257,215
content	18	147	11,000	41,502,000	1,708,442

Table 2: Summary of the Danish fire insurance data: number of single and common jumps and minimal, maximal and average losses in DKK for insured buildings and contents.



(a) Building

(b) Content

Figure 4: Histograms of small losses together with different body distributions for the Danish fire insurance data, year 2002.

6.1 Fitting the (Burr/GPD)^2 - Clayton model to the data

Motivated by our initial analysis, we now fit the (Burr/GPD)^2 - Clayton model to the data using our MCMC algorithm. As initial values we choose ML estimates for the different parameters, and then we run the sampler for 21,000 iterations and discard the first 1,000 for burn-in. Furthermore, we thin out the samples to reduce the autocorrelation and take only every 10th sample into account, as in the simulation study.

Figure 5 shows estimated marginal posterior densities for the parameters $\lambda_1, \boldsymbol{\theta}_1, \lambda_2, \boldsymbol{\theta}_2, \delta$ of the (Burr/GPD)^2 - Clayton model for the Danish fire insurance data between January and December 2002. Three different prior distributions for the parameters are used: uniform (noninformative) priors, independent Gamma priors with large standard deviations (denoted by Γ_1), and independent Gamma priors with smaller standard deviations (denoted by Γ_2). In the first case the prior standard deviations are taken as the corresponding mean divided by 4, in the second case they are taken as the corresponding mean divided

by 20, hence smaller by factor 5.

We clearly see the impact of the prior distributions on the uncertainty in the parameters. Whereas for the uniform prior and the Γ_1 priors the results are quite similar, the marginal densities are much more concentrated for the more informative Γ_2 priors. Particularly the estimates for the parameters of the GPD distribution used to model the tails of the sliced distributions show significant uncertainties for the uniform and the Γ_1 prior. Recalling that we have only about 40 observations available to model both tails, this behaviour is not a surprise and can be expected.

Table 3 reports the posterior mean estimates together with 95% credible intervals for all parameters, given the three different priors. Summarizing the plots and the credible intervals it becomes apparent that the choice of the standard deviations in the prior distributions has a major impact on the simulated values and the estimated posterior distributions. When choosing large values there is no significant difference compared to an uniform prior. Maybe informative expert judgment from experienced risk managers could justify the usage of small standard deviations in the priors and, hence, to make them quite informative. From Figure 5, however, one can see that this comes along with the risk of underestimating the uncertainty about the parameter estimates.

In Table 4 we see how the estimates change when the shorter time period of six months (T=6) is considered instead of the one year period (T=12). As to expect, the 95% credible intervals for the shorter data set are bigger, since less observations are available. However, the posterior mean estimates are not significantly different.

In Figure 6 we finally plot the density of the fitted (Burr/GPD)^2 - Clayton model together with histograms of the original data set, for both the building and the content component. Note that the estimated densities are based on the uniform prior distributions. Using the posterior mean estimates as parameters for the marginal jump size distributions in both components yields the following explicit densities for $x, y > 0$:

$$
\begin{aligned}
f_1(x) &= \mathbf{1}_{(0,3.96]}\, 1.49\, x^{3.10} \left(1 + x^{4.10}\right)^{-1.42} + \mathbf{1}_{(3.96,\infty)}\, 0.226 \left(1 + 0.108 \quad (x + 3.36)\right)^{-3.40}, \\
f_2(y) &= \mathbf{1}_{(0,3.90]}\, 1.98\, y^{0.196} \left(1 + y^{1.20}\right)^{-2.94} + \mathbf{1}_{(3.90,\infty)}\, 0.101 \left(1 + 0.0217\, (y + 11.7)\right)^{-6.47}.
\end{aligned}
$$

6.2 Bayes factors

So far we just stated that the (Burr/GPD)^2 - Clayton model is best (at least among the models we tried) for fitting the Danish fire insurance data with the MCMC method. Exemplary, we now want to compare this model to two other Clayton models, namely the (Loggamma/GPD)^2 - Clayton model and the (Burr)^2 - Clayton model. To compare these three competing models, we employ the Bayes factors which is a powerful tool in Bayesian model selection.

Each of the models is described by a model-specific parameter vector $\boldsymbol{\psi}_l \in \Psi_l \subset \mathbb{R}^l$, $l = 1, 2, 3$. In particular we have $\boldsymbol{\psi}_1 = (\lambda_1, c_1, k_1, h_1, \beta_1, \xi_1, \lambda_2, c_2, k_2, h_2, \beta_2, \xi_2, \delta)$ for the (Burr/GPD)^2 - Clayton model, $\boldsymbol{\psi}_2 = (\lambda_1, a_1, b_1, h_1, \beta_1, \xi_1, \lambda_2, a_2, b_2, h_2, \beta_2, \xi_2, \delta)$ for the (Loggamma/GPD)^2 - Clayton model and $\boldsymbol{\psi}_3 = (\lambda_1, c_1, k_1, \lambda_2, c_2, k_2, \delta)$ for the (Burr)^2 - Clayton model. As explained in the Chapter by Robert and Marin earlier in this book,

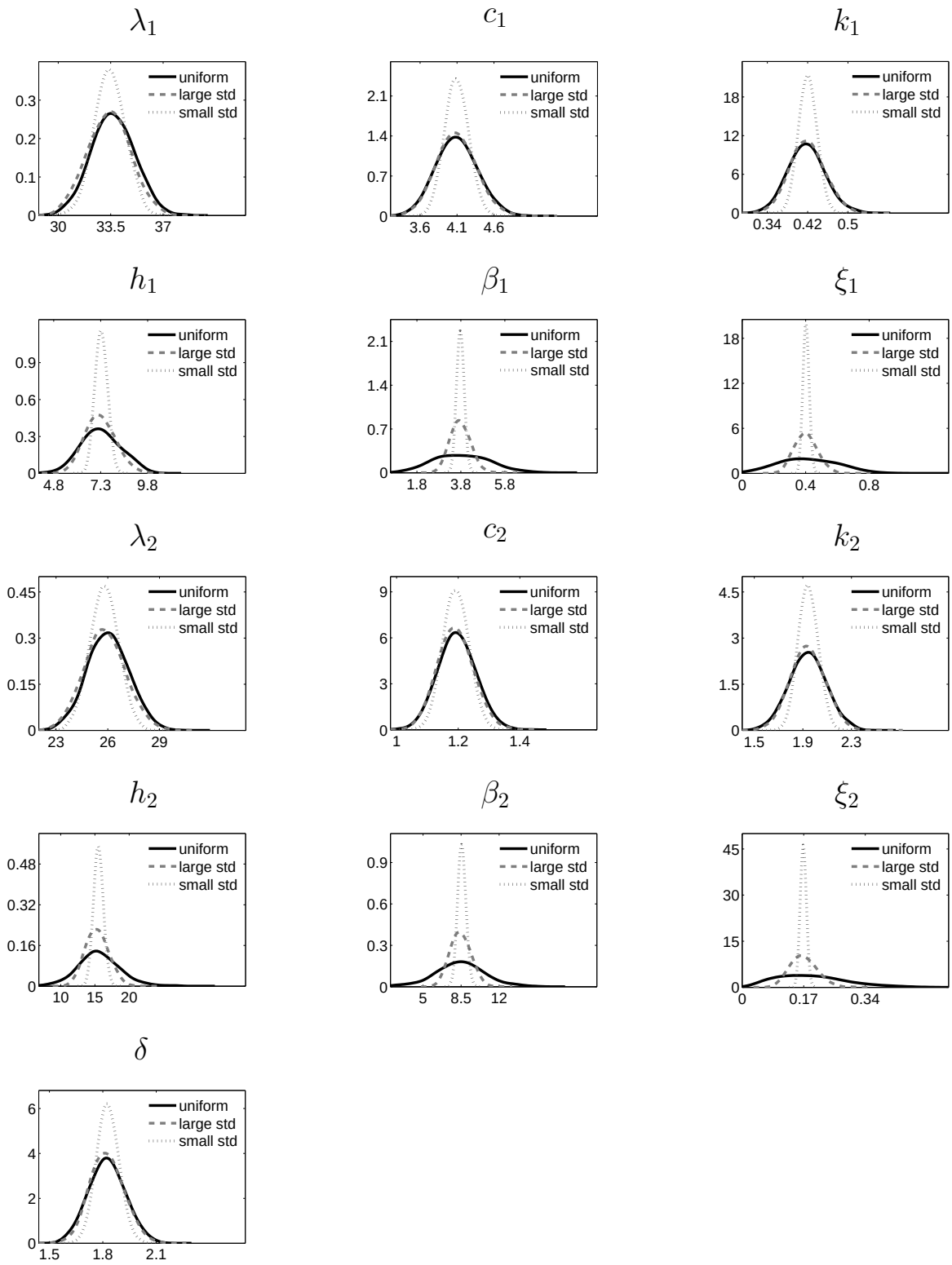


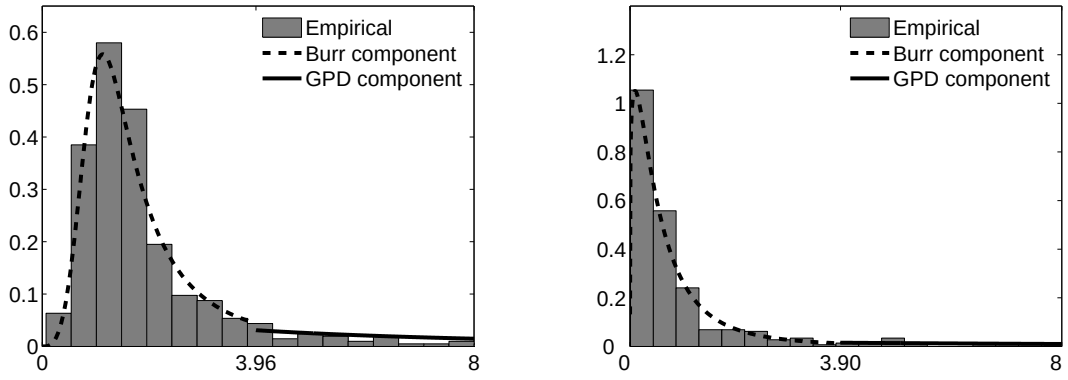
Figure 5: Density estimates of marginal posterior distributions for the parameters of the (Burr/GPD)^2 - Clayton model. Results for three different prior distributions.

	λ_1	c_1	k_1
(ML estimate)	33.49	4.094	0.4202
uniform prior	33.66 [30.90, 36.47]	4.096 [3.578, 4.658]	0.4178 [0.3507, 0.4926]
Γ_1 prior	33.37 [30.52, 36.25]	4.095 [3.568, 4.637]	0.4202 [0.3541, 0.4895]
Γ_2 prior	33.44 [31.44, 35.39]	4.091 [3.791, 4.396]	0.4200 [0.3865, 0.4558]
	h_1	β_1	ξ_1
(ML estimate)	7.325	3.776	0.4027
uniform prior	7.319 [5.506, 9.235]	3.846 [1.597, 6.356]	0.4170 [0.0885, 0.7748]
Γ_1 prior	7.291 [5.830, 8.973]	3.776 [2.928, 4.744]	0.4007 [0.2718, 0.5489]
Γ_2 prior	7.338 [6.707, 7.984]	3.778 [3.457, 4.116]	0.4023 [0.3646, 0.4412]
	λ_2	c_2	k_2
(ML estimate)	25.85	1.193	1.946
uniform prior	26.02 [23.69, 28.39]	1.196 [1.081, 1.318]	1.940 [1.647, 2.251]
Γ_1 prior	25.82 [23.49, 28.16]	1.191 [1.084, 1.306]	1.941 [1.680, 2.209]
Γ_2 prior	25.80 [24.14, 27.39]	1.192 [1.113, 1.273]	1.943 [1.790, 2.102]
	h_2	β_2	ξ_2
(ML estimate)	15.51	8.566	0.1690
uniform prior	15.60 [9.67, 22.23]	8.424 [3.629, 13.32]	0.1830 [0.0499, 0.4152]
Γ_1 prior	15.29 [12.16, 18.73]	8.460 [6.616, 10.33]	0.1700 [0.1049, 0.2504]
Γ_2 prior	15.52 [14.22, 16.93]	8.578 [7.840, 9.32]	0.1689 [0.1528, 0.1863]
	δ		
(ML estimate)	1.829		
uniform prior	1.824 [1.630, 2.025]		
Γ_1 prior	1.822 [1.650, 2.016]		
Γ_2 prior	1.826 [1.704, 1.950]		

Table 3: Posterior mean estimates together with 95% credible intervals, maximum likelihood estimates for comparison. The 95% credible intervals are based on the empirical 2.5% and 97.5% quantiles. Results for three different prior distributions: uniform, independent Gamma with large standard deviations (Γ_1), independent Gamma with small standard deviations (Γ_2).

	λ_1	c_1	k_1
$T = 12$	33.37 [30.52, 36.25]	4.095 [3.568, 4.637]	0.4202 [0.3541, 0.4895]
$T = 6$	33.20 [29.55, 37.07]	4.113 [3.441, 4.799]	0.4092 [0.3258, 0.4957]
	h_1	β_1	ξ_1
$T = 12$	7.291 [5.830, 8.973]	3.776 [2.928, 4.744]	0.4007 [0.2718, 0.5489]
$T = 6$	7.210 [5.424, 9.277]	3.839 [2.838, 4.961]	0.3900 [0.2447, 0.5597]
	λ_2	c_2	k_2
$T = 12$	25.82 [23.49, 28.16]	1.191 [1.084 , 1.306]	1.941 [1.680, 2.209]
$T = 6$	25.17 [22.00, 28.47]	1.132 [0.9830, 1.288]	1.881 [1.504, 2.301]
	h_2	β_2	ξ_2
$T = 12$	15.29 [12.16, 18.73]	8.460 [6.616, 10.33]	0.1700 [0.1049 , 0.2504]
$T = 6$	14.40 [10.56, 18.56]	7.645 [5.689, 9.789]	0.1630 [0.09763, 0.2462]
	δ		
$T = 12$	1.822 [1.650, 2.016]		
$T = 6$	1.778 [1.529, 2.044]		

Table 4: Comparison of posterior mean estimates (for Γ_1 prior) for the periods January to December 2002 to the estimates for the period July to December 2002. The estimates are given with 95% credible intervals.



(a) Building (b) Content

Figure 6: Fitted marginal jump size densities in the (Burr/GPD)^2 - Clayton model.

the Bayes factors

$$B_{ij}(\mathbf{z}) = \frac{\int_{\Psi_i} f_i\left(\mathbf{z} \mid \boldsymbol{\psi}_i\right) \pi_i\left(\boldsymbol{\psi}_i\right) d\boldsymbol{\psi}_i}{\int_{\Psi_j} f_j\left(\mathbf{z} \mid \boldsymbol{\psi}_j\right) \pi_j\left(\boldsymbol{\psi}_j\right) d\boldsymbol{\psi}_j}, \qquad i,j \in \{1,2,3\}\ ,\ i \neq j\,,$$

can be approximated by the strongly consistent estimates

$$\tilde{B}_{ij}(\mathbf{z}) = \frac{n_i^{-1} \sum_{k=1}^{n_i} f_i\left(\mathbf{z} \mid \boldsymbol{\psi}_{i,k}\right)}{n_j^{-1} \sum_{k=1}^{n_j} f_j\left(\mathbf{z} \mid \boldsymbol{\psi}_{j,k}\right)}, \qquad i,j \in \{1,2,3\}\ ,\ i \neq j\,.$$

Here $\boldsymbol{\psi}_{i,k}$, $k = 1, ..., n_i$, and $\boldsymbol{\psi}_{j,k}$, $k = 1, ..., n_j$, are two independent samples generated from the prior distributions π_i and π_j, respectively, and $\mathbf{z}$ represents the fire insurance data for the period January to December 2002.

The strength of evidence in favor of the (Burr/GPD)^2 - Clayton model versus the (Loggamma/GPD)^2 - Clayton or the (Burr)^2 - Clayton model, respectively, can be evaluated according to Jeffreys' Bayes factor scale, see e.g. Gelman et al. [7]. There Bayes factors B_{ij} greater than 100 are considered as decisive evidence for model i against model j. Following the results in Table 5, our (Burr/GPD)^2 - Clayton model is clearly the best among the competing three models.

(Burr/GPD)^2 vs. (Loggamma/GPD)^2	2.1×10^3
(Burr/GPD)^2 vs. (Burr)^2	1.4×10^8

Table 5: Approximated Bayes factors for the (Burr/GPD)^2 model, results for Γ_1 priors. 10,000 simulations are used for calculation.

7 Final remarks

In the simulation study we have seen that the choice of the prior distributions plays an important role. One must be aware, that decisions based on the posterior mean estimates and the posterior distributions are affected also by the priors. Therefore, the choice of a certain informative prior must be well-founded, in particular when the data sets used for the analysis are small. Here the impact of the prior is usually higher than for large data sets.

In practice, it might also be useful to transform the data before one starts with the analysis, e.g. by taking logarithms. Such transformations may change the dependence structure, and outliers may get closer to the rest of the data (or even farer away). Note that Esmaeili and Klüppelberg [6] have analyzed the log-transformed fire insurance data set using maximum likelihood estimation. This is, of course, also a very reasonable approach which leads to similar informative findings. Moreover, when we fitted several models to the fire insurance data and selected the best among these models using Bayes factors, this was intended to illustrate the procedure of model fitting and model selection. Of course, there may be other models which also describe the data very well.

References

[1] Basel Committee on Banking Supervision (2004) International convergence of capital measurement and capital standards. Basel.

[2] Böcker, K. and Klüppelberg, C. (2005) Operational VaR: a closed-form solution. *RISK Magazine*, December, 90-93.

[3] Böcker, K. and Klüppelberg, C. (2008) Modelling and measuring multivariate operational risk with Lévy copulas. *J. Operational Risk* **3**(2), 3-27.

[4] Cont, R. and Tankov, P. (2004) *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, Boca Raton.

[5] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[6] Esmaeili, H. and Klüppelberg, C. (2010) Parameter Estimation of a Bivariate Compound Poisson process. *Insurance: Mathematics and Economics*, to appear.

[7] Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (1995) *Bayesian Data Analysis.* Chapman & Hall, London.