

Article

Bayesian Modelling, Monte Carlo Sampling and Capital Allocation of Insurance Risks

Gareth W. Peters ^{1,2,3,*}, Rodrigo S. Targino ⁴ and Mario V. Wüthrich ⁵¹ Department of Statistical Science, University College London, London WC1E 6BT, UK² Oxford-Man Institute, Oxford University, Oxford OX1 2JD, UK³ System Risk Center, London School of Economics, London WC2A 2AE, UK⁴ Fundação Getulio Vargas, Escola de Matemática Aplicada, Botafogo, RJ 22250-040, Brazil; rodrigo.targino@fgv.br⁵ RiskLab, Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland; mario.wuethrich@math.ethz.ch

* Correspondence: garethpeters78@gmail.com; Tel.: +44-20-7679-1238

Received: 3 May 2017; Accepted: 31 August 2017; Published: 22 September 2017

Abstract: The main objective of this work is to develop a detailed step-by-step guide to the development and application of a new class of efficient Monte Carlo methods to solve practically important problems faced by insurers under the new solvency regulations. In particular, a novel Monte Carlo method to calculate capital allocations for a general insurance company is developed, with a focus on coherent capital allocation that is compliant with the Swiss Solvency Test. The data used is based on the balance sheet of a representative stylized company. For each line of business in that company, allocations are calculated for the one-year risk with dependencies based on correlations given by the Swiss Solvency Test. Two different approaches for dealing with parameter uncertainty are discussed and simulation algorithms based on (pseudo-marginal) Sequential Monte Carlo algorithms are described and their efficiency is analysed.

Keywords: capital allocation; premium and reserve risk; Solvency Capital Requirement (SCR); Sequential Monte Carlo (SMC); Swiss Solvency Test (SST)

1. Introduction

Due to the new risk based solvency regulations (such as the Swiss Solvency Test [FINMA \(2007\)](#) and Solvency II [European Commission \(2009\)](#)), insurance companies must perform two core calculations. The first one involves computing and setting aside the risk capital to ensure the company's solvency and financial stability, and the second one is related to the *capital allocation* exercise. This exercise is a process of splitting the (economic or regulatory) capital amongst its various constituents, which could be different lines of business (LoBs), types of exposures, territories or even individual products in a portfolio of insurance policies. One of the reasons for performing such an exercise is to utilize the results for a risk-reward management tool to analyse profitability. The amount of capital (or risk) allocated to each LoB, for example, may assist the central management's decision to further invest in or discontinue a business line.

In contrast to the quantitative risk assessment, where there is an unanimous view shared by regulators world-wide that it should be performed through the use of *risk measures*, such as the Value at Risk (VaR) or Expected Shortfall (ES), there is no consensus on how to perform capital allocation to sub-units. In this work we follow the Euler allocation principle (see, e.g., [Tasche \(1999\)](#) and [McNeil et al. 2010](#), *sct. 6.3*), which is briefly revised in the next section. For other allocation principles we refer the reader to [Dhaene et al. \(2012\)](#).

Under the Euler principle the allocation for each one of the portfolio's constituents can be calculated through an expectation conditional on a rare event. Even though, in general, these expectations are not available in closed form, some exceptions exist, such as the multivariate Gaussian model, first discussed in this context in [Panjer \(2001\)](#) and extended to the case of multivariate elliptical distributions in [Landsman and Valdez \(2003\)](#) and [Dhaene et al. \(2008\)](#); the multivariate gamma model of [Furman and Landsman \(2005\)](#); the combination of the Farlie-Gumbel-Morgenstern (FGM) copula and (mixtures of) exponential marginals from [Bargès et al. \(2009\)](#) or (mixtures of) Erlang marginals [Cossette et al. \(2013\)](#); and the multivariate Pareto-II from [Asimit et al. \(2013\)](#).

In this work we develop algorithms to calculate the marginal allocations for a generic model, which, invariably, leads to numerical approximations. Although simple Monte Carlo schemes (such as rejection sampling or importance sampling) are flexible enough to be used for a generic model, they can be shown to be computationally highly inefficient, as the majority of the samples do not satisfy the necessary conditioning event (which is a rare event). We build upon ideas developed in [Targino et al. \(2015\)](#) and propose an algorithm based on methods from Bayesian Statistics, namely a combination of Markov Chain Monte Carlo (for parameter estimation) and (pseudo-marginal) Sequential Monte Carlo (SMC) for the capital allocation.

As a side result of the allocation algorithm, we are able to efficiently compute both the company's overall Value at Risk (VaR) and also its Expected Shortfall (ES), (partially) addressing one of the main concerns of [Embrechts et al. \(2014\)](#): For High Confidence Levels, e.g., 95% and beyond, the “*statistical quantity*” VaR Can only Be Estimated with Considerable Statistical, as well as Model Uncertainty. Even though the issue of model uncertainty is not resolved, our algorithm can, at least, help to reduce the “*statistical uncertainty*”, measured by a variance reduction factor taking as basis a standard Monte Carlo simulation with comparable computational cost.

The proposed allocation procedure is described for a fictitious general insurance company with 9 LoBs (see Table 1 and Section 8). Further, within each LoB we also allocate the capital to the one-year reserve risk (due to claims from previous years) and the one-year premium risk.

Table 1. Initial synthetic balance sheet.

LoB	Reserves	Premium
1 MTPL	2391.64	503.14
2 Motor Hull	99.08	573.26
3 Property	449.26	748.76
4 Liability	870.27	299.73
5 Workers Compensation (UVG)	1104.66	338.63
6 Commercial Health	271.54	254.21
7 Private Health	7.32	7.20
8 Credit and Surety	49.50	34.64
9 Others	67.64	46.28
Total	5310.92	2805.87

In order to study the premium risk we follow the framework prescribed by the Swiss Solvency Test (SST) in ([FINMA 2007](#), sct. 4.4). In this technical document, given company-specific quantities, the distribution of the premium risk is deterministically defined and no parameter uncertainty is involved. For the reserve risk, we use a fully Bayesian version of the gamma-gamma chain ladder model, analysed in [Peters et al. \(2017\)](#). As this model is described via a set of unknown parameters, two different approaches to capital allocation are proposed: a *marginalized* one, where the unknown parameters are marginalized prior to the allocation process and a *conditional* one, which is performed conditional on the unknown parameters and the parameter is integrated out numerically ex-poste.

The remainder of this paper is organized as follows. Section 2 formally describes marginal risk contributions (allocations) under the marginalized and conditional models. Section 3 reviews concepts

of SMC algorithms and how they can be used to compute the quantities described in Section 2. We set the notation used for claims reserving in Section 4, before formally defining the models for the reserve risk (Section 5) and the premium risk (Section 6); these are merged together through a copula in Section 7. Sections 8 and 9 provide details of the synthetic data used, the inferential procedure for the unknown parameters and the implementation of the SMC algorithms. Results and conclusions are presented, respectively, in Sections 10 and 11.

2. Risk Allocation for the Swiss Solvency Test

In this section we follow the Euler allocation principle (see, e.g., Tasche (1999) and (McNeil et al. 2010, sct. 6.3)) and discuss how the risk capital that is held by an insurance company can be split into different risk triggers. As stochastic models for these risks involve a set of unknown parameters, we present an allocation procedure for a marginalized model (which arises when the parameter uncertainty is resolved beforehand) and a conditional model (which is still dependent on unknown parameters).

Although we postpone the construction of the specific claims payments model to Section 5 we now assume its behaviour is given by a Bayesian model depending on a generic parameter vector θ , for which a prior distribution is assigned. Probabilistic statements, such as the calculation of the risks allocated to each trigger, have to be made based on the available data, described by the filtration $\{\mathcal{F}(t)\}_t$ and formally defined in Section 4. This requirement implies that the uncertainty on the parameter values needs to be integrated out, in a process that must typically be performed numerically.

Therefore, to calculate the risk allocations we approximate the stochastic behaviour of functions of future observations, with the functions defined in Section 4. For the moment, let us denote by \bar{Z} a multivariate function of $\mathcal{F}(t+1)$, the future data, and θ the vector of model parameters. On the one hand, in the *conditional model*, we approximate the distribution of the components of the vector $\bar{Z} | \theta, \mathcal{F}(t)$. On the other hand, in the *marginalized model*, the approximation is performed after the parameter uncertainty has been integrated out (i.e., marginalized). In this later framework, we approximate the distribution of the components of $Z | \mathcal{F}(t)$, where the random vector Z is defined as $Z = \mathbb{E}[\bar{Z} | \mathcal{F}(t)]$, with expectation taken with respect to $\theta | \mathcal{F}(t)$. Note that, given $\mathcal{F}(t)$, Z is a random variable, as it depends on future information, i.e., $\mathcal{F}(t+1)$. Both in the conditional and in the marginalized models we use moment matching and log-normal distributions for the approximations and couple the distributions via a Gaussian copula.

Suppressing the dependence on the available information, $\mathcal{F}(t)$, these two models (marginalized and conditional) are defined through their probability density functions (p.d.f.'s), $f_Z(z)$ and $f_{\bar{Z}|\theta}(\bar{z} | \theta)$, respectively, which are both assumed to be combinations of log-normal distributions and a gaussian copula. For the conditional model, as we work in a Bayesian framework, the unknown parameter vector θ has a (posterior) distribution with p.d.f. $f_\theta(\theta)$. This is, then, combined with the likelihood $f_{\bar{Z}|\theta}(\bar{z} | \theta)$ to construct $f_{\bar{Z}}(\bar{z})$, the density used for inference under the conditional model.

For the methodology discussed in this work, the important features of these two models are that $f_Z(z)$ is known in closed form, whilst $f_{\bar{Z}}(\bar{z})$ is not.

In summary, the two models presented in Sections 4 to 7 are defined as

$$\text{Marginalized model: } Z \sim f_Z(z); \quad (1)$$

$$\text{Conditional model: } \bar{Z} \sim f_{\bar{Z}}(\bar{z}) = \int f_{\bar{Z}|\theta}(\bar{z} | \theta) f_\theta(\theta) d\theta. \quad (2)$$

Remark 1. As the “original” model for claims payments is a Bayesian model, we use the Bayesian nomenclature for both the marginalized and the conditional model. For the former, the Bayesian structure of prior and likelihood is hidden in Equation (1), as the parameter θ has already been marginalized (with respect to its posterior distribution). For the later, we explicitly make use of the posterior distribution of θ in Equation (2). Another strategy, followed in Wüthrich (2015), is to use an “empirical Bayes” approach, fixing the value of the unknown parameter vector θ , for example at its maximum likelihood estimator (MLE).

Under the marginalized model we define $S = \sum_{i=1}^d Z_i$ as the company's overall risk. The SST requires the total capital to be calculated as the 99% ES of S , given by

$$\rho(S) = \mathbb{E}[S | S \geq \text{VaR}_{99\%}(S)]. \quad (3)$$

In turn, the Euler allocation principle states that the contribution of each component Z_i to the total capital in Equation (3) is given by

$$\rho_i = \mathbb{E}[Z_i | S \geq \text{VaR}_{99\%}(S)], \quad \forall i = 1, \dots, d. \quad (4)$$

The allocations for the conditional model follows the same structure, with Z_i and S replaced, respectively, by \bar{Z}_i and \bar{S} in Equation (4) and reads as

$$\bar{\rho}_i = \mathbb{E}[\bar{Z}_i | \bar{S} \geq \text{VaR}_{99\%}(\bar{S})], \quad \forall i = 1, \dots, d, \quad (5)$$

with $\bar{S} = \sum_{i=1}^d \bar{Z}_i$. For the models discussed below the density of $f_{\bar{Z}}(\bar{z})$ is not known in closed form, adding one more layer of complexity to the proposed method.

Remark 2. Observe that the log-normal approximations are done at different stages in the marginalized and the conditional models. Therefore, we expect that the results will differ.

Although computing ρ_i and $\bar{\rho}_i$ is a static problem, for the sake of transforming the Monte Carlo estimation into an efficient computational framework, we embed the calculation of these quantities into a sequential procedure, where at each step we solve a simpler problem, through a relaxation of the rare-event conditioning constraint to a sequence of less extreme rare-events. In the next section we discuss the methodological Monte Carlo approach used to perform this task. The reader familiar with the concepts of Sequential Monte Carlo methods may skip Section 3.1.

3. SMC Samplers and Capital Allocation

For the marginalized and conditional models presented in Sections 4 to 7 the marginal contributions in Equations (4) and (5) cannot be calculated in analytic form for a generic model, so a simulation technique needs to be employed. In the sequel we provide a brief overview of a class of Monte Carlo methods, named Sequential Monte Carlo (SMC). For a recent survey in the topic, with focus on economics, finance and insurance applications the reader is referred to Creal (2012) and Del Moral et al. (2013). For a generic introductory review we refer the reader to Doucet and Johansen (2009).

3.1. A Brief Introduction to SMC Methods

The class of Sequential Monte Carlo (SMC) algorithms, also called Particle Filters, has its roots in the fields of engineering, probability and statistics where it was primarily used for sampling from a sequence of distributions (see, e.g., Gordon et al. (1993) and Del Moral (1996)). In the context of state-space models, SMC methods can be used to sequentially approximate the filtering distributions of non-linear and non-Gaussian state space models, solving the same problem as the Kalman filter—a technique with a long-standing tradition in actuarial mathematics (see, e.g., De Jong and Zehnwrith (1983) and Verrall (1989)).

The general context of a standard SMC method is that one wants to approximate a (often naturally occurring) sequence of p.d.f.'s $\{\tilde{\pi}_t\}_{t \geq 1}$ with the support of each function in this sequence is given by $\text{supp}(\tilde{\pi}_t) = \mathbb{R}^d \times \dots \times \mathbb{R}^d = \mathbb{R}^{d \times t}$, for $t \geq 1$, where t can be any artificial ordering of the sequence that is problem specific. We assume that $\tilde{\pi}_t$ is (only) known up to a normalizing constant, and we write

$$\tilde{\pi}_t(\mathbf{z}_{1:t}) = \mathcal{Z}_t^{-1} \tilde{\gamma}_t(\mathbf{z}_{1:t}),$$

where $\mathbf{z}_{1:t} = (z_1, \dots, z_t) \in \mathbb{R}^{d \times t}$.

3.1.1. SMC Algorithm

Procedurally, we initialize the algorithm sampling a set of N independent particles (as the samples are denoted in the literature) from the distribution $\tilde{\pi}_1$ and set normalized weights to $W_1^{(j)} = 1/N$, for all $j = 1, \dots, N$. If it is not possible to sample directly from $\tilde{\pi}_1$, one should sample from an importance distribution \tilde{q}_1 and calculate its weights accordingly. Then the particles are sequentially propagated through each distribution $\tilde{\pi}_t$ in the sequence via three main processes: mutation, correction (incremental importance weighting) and resampling. In the first step (mutation) we propagate particles from time $t - 1$ to time t , in the second one (correction) we calculate the new importance weights of the particles.

Without resampling, this method can be seen as a sequence of importance sampling (IS) steps, where the target distribution at each step t is $\tilde{\gamma}_t$ (the unnormalized version of $\tilde{\pi}_t$) and the importance distribution is given by

$$\tilde{q}_t(\mathbf{z}_{1:t}) = \tilde{q}_1(\mathbf{z}_1) \prod_{j=2}^t K_j(\mathbf{z}_{j-1}, \mathbf{z}_j), \quad (6)$$

where $K_j(\mathbf{z}_{j-1}, \cdot)$ is the mechanism used to propagate particles from time $t - 1$ to t , known as the mutation kernel. Therefore, after the mutation step each particle $j = 1, \dots, N$ has (unnormalized) importance weight given by

$$w_t^{(j)} = \frac{\tilde{\gamma}_t(\mathbf{z}_{1:t}^{(j)})}{\tilde{q}_t(\mathbf{z}_{1:t}^{(j)})} = w_{t-1}^{(j)} \underbrace{\frac{\tilde{\gamma}_t(\mathbf{z}_{1:t}^{(j)})}{\tilde{\gamma}_{t-1}(\mathbf{z}_{1:t-1}^{(j)}) K_t(\mathbf{z}_{t-1}^{(j)}, \mathbf{z}_t^{(j)})}}_{\text{incremental weight: } \alpha_t^{(j)}} = w_{t-1}^{(j)} \alpha_t^{(j)}. \quad (7)$$

These importance weights can be normalized to create a set of (normalized) weighted particles $\{\mathbf{z}_{1:t}^{(j)}, W_t^{(j)}\}_{j=1}^N$, with normalized weights $W_t^{(j)} = \frac{w_t^{(j)}}{\sum_{k=1}^N w_t^{(k)}}$. In this case, from the Law of Large Numbers,

$$\sum_{j=1}^N W_t^{(j)} \varphi(\mathbf{z}_{1:t}^{(j)}) \longrightarrow \mathbb{E}_{\tilde{\pi}_t}[\varphi(\mathbf{Z}_{1:t})] = \int_{\mathbb{R}^{d \times t}} \varphi(\mathbf{z}_{1:t}) \tilde{\pi}_t(\mathbf{z}_{1:t}) d\mathbf{z}_{1:t}, \quad (8)$$

$\tilde{\pi}_t$ -almost surely as $N \rightarrow \infty$, for any test function φ such that the expectation of φ under $\tilde{\pi}_t$ exists (see Geweke (1989)).

Remark 3. The reader should note that the knowledge of $\tilde{\pi}_t$ up to a normalizing constant is sufficient for the implementation of a generic SMC algorithm, since the normalized weights $W_t^{(j)}$ are the same for both $\tilde{\pi}_t$ and $\tilde{\gamma}_t$.

In simple implementations of the SMC algorithm (such as the one discussed above), when the algorithmic time t increases, the estimates in Equation (8) become, eventually, effectively a function of one sample point $\{\mathbf{z}_t^{(j)}, W_t^{(j)}\}$; what is observed, in practice, is that for some particle j , $W_t^{(j)} \approx 1$ and for all the others the normalized weights are negligible. This degeneracy is measured using the Effective Sample Size (ESS) defined in Liu and Chen (1995) and Liu and Chen (1998) as

$$\text{ESS}_t = \left[\sum_{j=1}^N (W_t^{(j)})^2 \right]^{-1} \in [1, N].$$

This quantity has the interpretation that ESS_t is maximized when $\{W_t^{(j)}, \mathbf{z}_t^{(j)}\}_{j=1}^N$ forms a uniform distribution on $\{\mathbf{z}_t^{(j)}\}_{j=1}^N$ and minimized when $W_t^{(j)} = 1$ for some j . One may also use the Gini index or the entropy as a degeneracy measure, as discussed, for example, in Martino et al. (2017).

One way to tackle this degeneracy problem is to unbiasedly resample the whole set of weighted particles, for example, choosing (with replacement) N samples from the system where each $\mathbf{z}_t^{(j)}$ is selected with probability weight $W_t^{(j)}$. In our algorithms we propose to resample the whole set of weighted samples whenever it is “too degenerate” and our degeneracy threshold is $\text{ESS}_t < N/2$. Many

different resampling schemes have been suggested in the literature and for a comparison between them we refer the reader to [Douc and Cappé \(2005\)](#) and [Gandy and Lau \(2015\)](#).

Although the resample step alleviates the degeneracy problem, its successive reapplication at each stage of the sampler produces the so-called sample impoverishment, where the number of distinct particles is extremely small. In [Gilks and Berzuini \(2001\)](#) it was proposed to add a “move” step with any kernel such that the target distribution is invariant with respect to it in order to rejuvenate the system. This kernel may be, for example, as a Markov Chain Monte Carlo (MCMC) kernel, which would begin with equally chosen weighted samples from the target distribution and then perturb them under a single step of a Metropolis Hastings acceptance-rejection mechanism. Note that in this case the samples start exactly in the target distribution’s stationary regime. Therefore, a single step of the Metropolis-Hastings accept-reject mutation is strictly valid and no burn-in is required.

More precisely, we can apply any kernel $M(\hat{\mathbf{z}}_{1:t}^{(j)}, \mathbf{z}_{1:t}^{(j)})$ that leaves $\tilde{\pi}_t$ invariant to move the sample $\hat{\mathbf{z}}_{1:t}^{(j)}$ to $\mathbf{z}_{1:t}^{(j)}$ (the hat denotes a sample *after* the resample step but *before* the “move” step), i.e.,

$$\tilde{\pi}_t(\mathbf{z}_{1:t}) = \int M(\hat{\mathbf{z}}_{1:t}, \mathbf{z}_{1:t}) \tilde{\pi}_t(\hat{\mathbf{z}}_{1:t}) d\hat{\mathbf{z}}_{1:t}.$$

3.1.2. SMC Samplers

Although very general, the SMC algorithm presented above, in principle, requires the sequence of p.d.f.’s to have an increasing support. However it has been shown in [Peters \(2005\)](#) and [Del Moral et al. \(2006\)](#) that these algorithms can be applied to sequences of p.d.f.’s defined on the same support, leading to the so-called SMC sampler algorithm discussed below. This development is central for the insurance applications explored in this paper.

Given the sequence of densities $\{\pi_t\}_{t \geq 1}$ (and its unnormalized version, $\{\gamma_t\}_{t \geq 1}$), where each element is defined over the same support, say \mathbb{R}^d , we create another sequence, defined on $\mathbb{R}^{d \times t}$, the *path space*

$$\tilde{\pi}_t(\mathbf{z}_{1:t}) \propto \tilde{\gamma}_t(\mathbf{z}_{1:t}) = \gamma_t(\mathbf{z}_t) \prod_{s=1}^{t-1} L_s(\mathbf{z}_{s+1}, \mathbf{z}_s), \quad (9)$$

which, for any Markov kernel $L_s(\mathbf{z}_{s+1}, \mathbf{z}_s)$ is a density with $\pi_t(\mathbf{z}_t)$ as marginal (which can be seen by integrating out $\mathbf{z}_{1:t-1}$). Note that, in Equation (9) time runs backwards, from t to 1. For completeness we define $\tilde{\pi}_1(\mathbf{z}_{1:1}) = \pi_1(\mathbf{z}_1)$ and $\tilde{\gamma}_1(\mathbf{z}_{1:1}) = \gamma_1(\mathbf{z}_1)$.

If $q_1 \equiv \tilde{q}_1$ is an IS density targeting $\pi_1 \equiv \tilde{\pi}_1$ then, see Equation (6),

$$\tilde{q}_t(\mathbf{z}_{1:t}) = q_1(\mathbf{z}_1) \prod_{j=2}^t K_j(\mathbf{z}_{j-1}, \mathbf{z}_j),$$

is defined, for Markov kernels $K_j(\mathbf{z}_{j-1}, \mathbf{z}_j)$, as an IS density targeting $\tilde{\pi}_t(\mathbf{z}_{1:t})$. For $t = 1$ we define $K_1 \equiv 1$.

As in the SMC algorithm, to generate a set of weighted samples $\{\mathbf{z}_t^{(j)}, W_t^{(j)}\}_{j=1}^M$ from $\pi_t(\mathbf{z}_t)$ one can use a sequence of IS steps on the path space where the unnormalized importance weights are, at each time step $t \geq 1$, given by, see Equation (7),

$$w_t^{(j)} = \frac{\tilde{\gamma}_t(\mathbf{z}_{1:t}^{(j)})}{\tilde{q}_t(\mathbf{z}_{1:t}^{(j)})} = w_{t-1}^{(j)} \frac{\gamma_t(\mathbf{z}_t^{(j)}) L_{t-1}(\mathbf{z}_t^{(j)}, \mathbf{z}_{t-1}^{(j)})}{\gamma_{t-1}(\mathbf{z}_{t-1}^{(j)}) K_t(\mathbf{z}_{t-1}^{(j)}, \mathbf{z}_t^{(j)})} = w_{t-1}^{(j)} \alpha_t^{(j)}, \quad (10)$$

where $w_0^{(j)} \equiv 1$, for all $j = 1, \dots, N$. The normalized weights are then computed as

$$W_t^{(j)} = \frac{w_t^{(j)}}{\sum_{k=1}^N w_t^{(k)}}.$$

The pseudo-code of the SMC sampler procedure just described is found in Algorithm 1.

Algorithm 1: SMC sampler algorithm.

Inputs: (Forward) mutation kernels $\{K_t(z_{t-1}, z_t)\}_{t=2}^T$, (artificial) backward kernels $\{L_{t-1}(z_t, z_{t-1})\}_{t=2}^T$, MCMC move kernel $\{M_t(\hat{z}_t, z_t)\}_{t=1}^T$;

for $j = 1, \dots, N$ **do**

Sample i.i.d. $z_1^{(j)} \sim \pi_1(\cdot)$;

Set $w_1^{(j)} = 1$;

Set $W_1^{(j)} = \frac{1}{N}$;

end

for $t = 2, \dots, T$ **do**

for $j = 1, \dots, N$ **do**

Sample $z_t^{(j)} \sim K_t(z_{t-1}^{(j)}, \cdot)$;

Calculate the weights $w_t^{(j)} = w_{t-1}^{(j)} \frac{\gamma_t(z_t^{(j)}) L_{t-1}(z_t^{(j)}, z_{t-1}^{(j)})}{\gamma_{t-1}(z_{t-1}^{(j)}) K_t(z_{t-1}^{(j)}, z_t^{(j)})} = w_{t-1}^{(j)} \alpha_t^{(j)}$;

end

Calculate the normalized weights $W_t^{(j)} = \frac{w_t^{(j)}}{\sum_{j=1}^N w_t^{(j)}}$;

if $ESS_t < N/2$ **then**

for $j = 1, \dots, N$ **do**

Resample $\hat{z}_t^{(j)}$, i.i.d from $\{z_t^{(k)}, W_t^{(k)}\}_{k=1}^N$;

Sample $z_t^{(j)} \sim M_t(\hat{z}_t^{(j)}, \cdot)$;

Set $W_t^{(j)} = 1/N$;

end

end

end

Result: Weighted random samples $\{z_t^{(j)}, W_t^{(j)}\}_{j=1}^N$ approximating π_t , for all $t = 1, \dots, T$.

The introduction of the sequence of kernels $\{L_{t-1}\}_{t=2}^T$ creates a new degree of freedom in the design of SMC samplers compared with the usual SMC algorithms, where only the forward mutation kernels $\{K_t\}_{t=2}^T$ should be designed. As discussed in [Peters \(2005\)](#) and [Del Moral et al. \(2006\)](#) if one wants to minimize the variance of the importance weights one strategy is to use the following approximation to the optimal backward kernel that minimizes the variance of the incremental weights (which cannot be computed in practice and must be approximated)

$$L_t(z_{t+1}, z_t) = \frac{\gamma_t(z_t) K_{t+1}(z_t, z_{t+1})}{\frac{1}{N} \sum_{j=1}^N w_t^{(j)} K_{t+1}(z_t^{(j)}, z_{t+1})}, \quad (11)$$

which leads to incremental weights

$$\alpha_t^{(j)} = \frac{\gamma_t(z_t^{(j)})}{\frac{1}{N} \sum_{k=1}^N w_{t-1}^{(k)} K_t(z_{t-1}^{(k)}, z_t^{(j)})}. \quad (12)$$

With the methodological tools provided by the SMC samplers we now proceed on how to adapt these methods to the allocation of risks under our generic marginalized and conditional models.

3.2. Allocations for the Marginalized Model

For a generic random vector $\mathbf{Z} = (Z_1, \dots, Z_d)$ with known marginal densities and distribution functions, respectively $f_{Z_i}(z_i)$ and $F_{Z_i}(z_i)$, and copula density $c(u_1, \dots, u_d)$ on $[0, 1]^d$, due to Sklar's theorem (see Sklar (1959) and (McNeil et al. 2010, chp. 5)) the joint density of \mathbf{Z} can be written as

$$f_{\mathbf{Z}}(\mathbf{z}) = c(\mathbf{u}) \prod_{i=1}^d f_{Z_i}(z_i),$$

where $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$ and $u_i = F_{Z_i}(z_i)$. In order to approximate the marginal risk contributions ρ_i from Equation (4) we can use samples from the distribution

$$\pi(\mathbf{z}) = f_{\mathbf{Z}}(\mathbf{z} | \mathbf{z} \in \mathcal{G}_{\mathbf{Z}}) = \frac{f_{\mathbf{Z}}(\mathbf{z}) \mathbb{1}_{\mathcal{G}_{\mathbf{Z}}}(\mathbf{z})}{\mathbb{P}[\mathbf{Z} \in \mathcal{G}_{\mathbf{Z}}]}, \quad (13)$$

where the set $\mathcal{G}_{\mathbf{Z}} = \mathcal{G}_{\mathbf{Z}}(B)$ is defined, for $B = \text{VaR}_{99\%}(S)$, as

$$\mathcal{G}_{\mathbf{Z}} = \left\{ \mathbf{z} \in \mathbb{R}^d : \sum_{i=1}^d z_i \geq B \right\}, \quad (14)$$

and the indicator function $\mathbb{1}_{\mathcal{G}_{\mathbf{Z}}}(\mathbf{z})$ is one when $\mathbf{z} \in \mathcal{G}_{\mathbf{Z}}$ and zero otherwise. It should be noted that since the boundary B in Equation (14) is given by $\text{VaR}_{99\%}(S)$ with $S = \sum_{i=1}^d Z_i$ we have $\mathbb{P}[\mathbf{Z} \in \mathcal{G}_{\mathbf{Z}}] = 0.01$, see discussion on this point in Targino et al. (2015).

3.2.1. Reaching a Rare Event Using Intermediate Steps

Instead of directly targeting the conditional distribution $(Z_1, \dots, Z_d) | \{S \geq \text{VaR}_{99\%}(S)\}$ the idea of the SMC sampler of Algorithm 1 is to sequentially sample from intermediate distributions with conditioning events that become rarer until the point we reach the distribution of interest (see Equation (14)). The benefit of such an approach is that the samples (particles) from a previous step (with a less rare conditioning event) are “guided” to the next algorithmic step (when targeting a rarer conditioning set) and, if carefully designed, no samples are wasted on the way to the target distribution, in the sense that no samples are incrementally weighted with a strictly zero weight. This “herds” the samples into the target sampling region of interest.

In order to sample from the target distribution defined in Equation (13) we use a sequence of intermediate distributions $\{\pi_t\}_{t=1}^T$, such that $\pi_T \equiv \pi$ and

$$\pi_t(\mathbf{z}) = f_{\mathbf{Z}}(\mathbf{z} | \mathbf{z} \in \mathcal{G}_{\mathbf{Z}_t}), \quad (15)$$

with $\mathcal{G}_{\mathbf{Z}_t} = \mathcal{G}_{\mathbf{Z}_t}(B_t)$ given by

$$\mathcal{G}_{\mathbf{Z}_t} = \left\{ \mathbf{z} \in \mathbb{R}^d : \sum_{i=1}^d z_i \geq B_t \right\}.$$

Remark 4. Differently from Targino et al. (2015), in order to make the algorithm more easily comparable with the one used for the conditional model, we do not transform the original random variable \mathbf{Z} through its marginal distribution functions. Therefore, instead of sampling from the conditional copula we sample from the conditional joint distribution of \mathbf{Z} .

The thresholds B_1, \dots, B_{T-1} are chosen in order to have increasingly rarer conditioning events as a function of t , starting from the unconditional joint density. In other words, $\{B_t\}_{t=1}^T$ needs to satisfy $0 = B_1 < \dots < B_{T-1} < B_T = B = \text{VaR}_{99\%}(S)$. Note that the choice $B_1 = 0$ assumes $S > 0$, \mathbb{P} -a.s., otherwise $B_1 = -\infty$. Depending on the choice of the thresholds $\{B_t\}_{t=1}^{T-1}$ it may be the case that the densities defined in Equation (15) are only known up to a normalizing constant so, from now on, we work with γ_t , the unnormalized version of π_t :

$$\pi_t(\mathbf{z}) \propto \gamma_t(\mathbf{z}) = f_{\mathbf{Z}}(\mathbf{z}) \mathbb{1}_{\mathcal{G}_{Z_t}}(\mathbf{z}). \quad (16)$$

If, at algorithmic time t , we have a set of N weighted samples $\{W_t^{(j)}, \mathbf{z}_t^{(j)}\}_{j=1}^N$ from π_t , with $\mathbf{z}_t^{(j)} = (z_{1,t}^{(j)}, \dots, z_{d,t}^{(j)})$ then we construct the following empirical approximation:

$$\mathbb{E}[Z_i | S \geq B_t] \approx \sum_{j=1}^N W_t^{(j)} z_{i,t}^{(j)}. \quad (17)$$

It should be noticed, though, that in our application the final threshold $B_T = B = \text{VaR}_{99\%}(S)$ is not previously known. In these cases, an adaptive strategy, similar to the one studied in Cérou et al. (2012) can be implemented, where neither B_1, \dots, B_{T-1} nor B_T needs to be previously known. More details on this aspect of the algorithm are provided in Section 9.1.

3.3. Allocations for The Conditional Model

From the discussion in Section 2 we see that the main difference between the marginalized and conditional models is the fact that the former density is analytically known (in fact, it is approximated by an analytically known density) whilst the latter is defined through an integral of a known density, see Equations (1) and (2). In this section we discuss how to adapt the algorithm presented in Section 3.2 for situations where the target density cannot be analytically computed but a positive and unbiased estimator for it can be calculated.

Following the recent developments on *pseudo-marginal* methods (see Andrieu and Roberts (2009) and Finke (2015) for a survey in the topic) we substitute the unknown density $f_{\bar{\mathbf{Z}}}$ in Equation (2) by a positive and unbiased estimate $\hat{f}_{\bar{\mathbf{Z}}}$ and show the SMC procedure still targets the correct distribution—a strategy similar to the ones proposed in Everitt et al. (2016) and McGree et al. (2015). In the context of rare event simulations a similar idea has been independently developed in Vergé et al. (2016) where the authors study the impact of the parameter uncertainty in the probability of the rare event, whilst we analyse the impact in expectations conditional to the rare event (as in Equation (5)).

The idea of replacing an unknown density by a positive and unbiased estimate is in the core of many recently proposed algorithms, such as the Particle Markov Chain Monte Carlo (PMCMC) of Andrieu et al. (2010), the Sequential Monte Carlo Squared (SMC²) of Chopin et al. (2013) and Fulop and Li (2013) (see also the island particle filter of Vergé et al. (2015)) and the Importance Sampling Squared (IS²) of Tran et al. (2014). In the context of Sequential Monte Carlo algorithms this argument first appeared as a brief note in Rousset and Doucet's comments of Beskos et al. (2006), where it reads that "(...) a straightforward argument shows that it is not necessary to know $w_k(X_{t_0:t_k}^{(i)})$ [the weights] exactly. Only an unbiased positive estimate $\hat{w}_k(X_{t_0:t_k}^{(i)})$ of $w_k(X_{t_0:t_k}^{(i)})$ is necessary to obtain asymptotically consistent SMC estimates under weak assumptions".

To introduce the concept we first estimate $f_{\bar{\mathbf{Z}}}$ by $f_{\bar{\mathbf{Z}}}(\cdot | \theta)$, which can be seen as a "one sample" approximation to the integral in Equation (2); then we show how to use an estimator based on $M \geq 1$ samples from f_{θ} . These two approaches have been named in the literature (see Everitt et al. (2016) and references therein) as, respectively, the single auxiliary variable (SAV) and the multiple auxiliary variable (MAV) methods.

3.3.1. Single Auxiliary Variable Method

To avoid direct use of $f_{\bar{\mathbf{Z}}}$ on the SMC sampler algorithm we provide a procedure on the joint space of $\bar{\mathbf{Z}}$ and the parameter θ , defined as $\mathcal{Y} = \mathbb{R}^d \times \Theta$. The reader is referred to Finke (2015) for an extensive list of known algorithms which can also be interpreted in a *extended space* way. The target distribution on this new space is defined as the joint distribution of $\bar{\mathbf{Z}}$ and θ and its marginal with respect to $\bar{\mathbf{Z}}$ is precisely the density of the conditional model.

Formally, for $\mathbf{y} = (\bar{\mathbf{z}}, \theta)$, $\mathcal{G}_{\bar{\mathbf{Z}}}(\bar{B}) = \mathcal{G}_{\bar{\mathbf{Z}}} = \left\{ \bar{\mathbf{z}} \in \mathbb{R}^d : \sum_{i=1}^d \bar{z}_i \geq \bar{B} \right\}$ and $\bar{B} = \text{VaR}_{99\%}(\bar{S})$ we define

$$\pi^y(y) \propto \gamma^y(y) = f_{\bar{Z}}(\bar{z} | \theta) f_{\theta}(\theta) \mathbb{1}_{\mathcal{G}_{\bar{Z}}}(\bar{z}),$$

which has the desired marginal target distribution of interest:

$$\bar{\pi}(\bar{z}) \propto \bar{\gamma}(\bar{z}) = \int_{\Theta} f_{\bar{Z}}(\bar{z} | \theta) f_{\theta}(\theta) d\theta \mathbb{1}_{\mathcal{G}_{\bar{Z}}}(\bar{z}). \quad (18)$$

Similarly to the densities defined in Equations (9) and (16) we define a sequence of target distributions both in \mathcal{Y} and \mathcal{Y}^t , respectively, as

$$\pi_t^y(y_t) \propto \gamma_t^y(y_t) = f_{\bar{Z}}(\bar{z}_t | \theta_t) f_{\theta}(\theta_t) \mathbb{1}_{\mathcal{G}_{\bar{Z}_t}}(\bar{z}_t),$$

and

$$\begin{aligned} \tilde{\pi}_t^y(y_{1:t}) &\propto \tilde{\gamma}_t^y(y_{1:t}) = \gamma_t^y(y_t) \prod_{s=1}^{t-1} L_s^y(y_{s+1}, y_s) \\ &= f_{\bar{Z}}(\bar{z}_t | \theta_t) f_{\theta}(\theta_t) \mathbb{1}_{\mathcal{G}_{\bar{Z}_t}}(\bar{z}_t) \prod_{s=1}^{t-1} \bar{L}_s(\bar{z}_{s+1}, \bar{z}_s | \theta_s) f_{\theta}(\theta_s), \end{aligned}$$

where the second identity specifies the choices of L_s^y , in terms of \bar{L}_s and f_{θ} .

Assuming we can perfectly sample from the distribution of θ (in our application this distribution is a posterior, from which samples are generated via simulation algorithms), to move y samples backwards from time $s+1$ to s we split this process into sampling θ_s from f_{θ} (ignoring θ_{s+1}) and then, conditional on θ_s , moving \bar{z}_{s+1} to \bar{z}_s . In other words, to sample

$$y_s = (\bar{z}_s, \theta_s) \mid \{y_{s+1} = (\bar{z}_{s+1}, \theta_{s+1})\} \sim L_s^y(y_{s+1}, y_s),$$

we split the process in two stages,

1. $\theta_s \sim f_{\theta}(\theta_s)$;
2. $\bar{z}_s \mid \bar{z}_{s+1} \sim \bar{L}_s(\bar{z}_{s+1}, \bar{z}_s \mid \theta_s)$.

The importance distribution on the path space of y can be expressed as

$$\tilde{q}_t^y(y_{1:t}) = q_1^y(y_1) \prod_{s=2}^t K_s^y(y_{s-1}, y_s) = \bar{q}_1(\bar{z}_1) f_{\theta}(\theta_1) \prod_{s=2}^t \bar{K}_s(\bar{z}_{s-1}, \bar{z}_s \mid \theta_s) f_{\theta}(\theta_s),$$

and, once again, the second identity provides the choices of q_1^y and K_s^y , i.e.,

$$q_1^y(y_1) = \bar{q}_1(\bar{z}_1) f_{\theta}(\theta_1) \quad \text{and} \quad K_s^y(y_{s-1}, y_s) = \bar{K}_s(\bar{z}_{s-1}, \bar{z}_s \mid \theta_s) f_{\theta}(\theta_s).$$

Therefore, a SMC procedure targeting the sequence $\{\pi_t^y(y_t)\}_{t=1}^T$ produces unnormalized weights

$$\begin{aligned} w_t^y &= \frac{\tilde{\gamma}_t^y(y_{1:t})}{\tilde{q}_t^y(y_{1:t})} \\ &= w_{t-1}^y \frac{\gamma_t^y(y_t) L_{t-1}^y(y_t, y_{t-1})}{\gamma_{t-1}^y(y_{t-1}) K_t^y(y_{t-1}, y_t)} \\ &= w_{t-1}^y \frac{f_{\bar{Z}}(\bar{z}_t | \theta_t) f_{\theta}(\theta_t) \mathbb{1}_{\mathcal{G}_{\bar{Z}_t}}(\bar{z}_t) \bar{L}_{t-1}(\bar{z}_t, \bar{z}_{t-1} | \theta_{t-1}) f_{\theta}(\theta_{t-1})}{f_{\bar{Z}}(\bar{z}_{t-1} | \theta_{t-1}) f_{\theta}(\theta_{t-1}) \mathbb{1}_{\mathcal{G}_{\bar{Z}_{t-1}}}(\bar{z}_{t-1}) \bar{K}_t(\bar{z}_{t-1}, \bar{z}_t | \theta_t) f_{\theta}(\theta_t)} \\ &= w_{t-1}^y \frac{f_{\bar{Z}}(\bar{z}_t | \theta_t) \mathbb{1}_{\mathcal{G}_{\bar{Z}_t}}(\bar{z}_t) \bar{L}_{t-1}(\bar{z}_t, \bar{z}_{t-1} | \theta_{t-1})}{f_{\bar{Z}}(\bar{z}_{t-1} | \theta_{t-1}) \mathbb{1}_{\mathcal{G}_{\bar{Z}_{t-1}}}(\bar{z}_{t-1}) \bar{K}_t(\bar{z}_{t-1}, \bar{z}_t | \theta_t)}, \end{aligned}$$

that can be used to create weighted samples from $\bar{\pi}_t(\bar{z}_t)$, which is the desired marginal of $\pi_t^y(y_t)$, the density required for the capital allocation.

Remark 5. From the structure of the mutation kernels K_t^y it should be noticed that at each iteration t a new value of θ_t needs to be generated and used to sample $\bar{z}_t | \theta_t$. In other words, for each particle $j = 1, \dots, N$ a different $\theta_t^{(j)}$ is to be used for each $\bar{z}_t^{(j)} | \theta_t^{(j)}$.

3.3.2. Multiple Auxiliary Variable

In the previous algorithm we, indirectly, estimate the density $f_{\bar{Z}}(\bar{z})$ by $f_{\bar{Z}}(\bar{z} | \theta)$. In this section we discuss how to use a different and more robust estimator, using $M \geq 1$ samples from θ . In the context of pseudo-marginal Monte Carlo Markov Chain (MCMC) [Andrieu and Vihola \(2015\)](#) show that reducing the variance of the estimate of the unknown density $f_{\bar{Z}}(\bar{z})$ leads to reduced asymptotic variance of estimators from the MCMC. For SMC algorithms this strategy has been used, for example, in [McGree et al. \(2015\)](#) and [Everitt et al. \(2016\)](#).

Before proceeding, we note that even in the case that $M = 1$ the algorithm still produces asymptotic and unbiased estimators (when the number of particles $N \rightarrow \infty$). However, the rate of variance reduction in the asymptotic estimates is directly affected by the choice of M (in a non-trivial manner). Furthermore, the asymptotic variance of Central Limit Theorem (CLT) estimators under the class of such pseudo-marginal Monte Carlo approaches is strictly ordered in M , with M increasing reducing the the asymptotic variance.

For any $M \geq 1$, a positive and unbiased estimate for $f_{\bar{Z}}(\bar{z})$ can be constructed as

$$\hat{f}_{\bar{Z}}(\bar{z}; \boldsymbol{\theta}) = \frac{1}{M} \sum_{i=1}^M f_{\bar{Z}}(\bar{z} | \theta^{(i)}), \quad (19)$$

where $\boldsymbol{\theta} = (\theta^{(1)}, \dots, \theta^{(M)}) \in \Theta^M$ and each $\theta^{(m)}$ is sampled independently from $f_{\theta}(\theta)$. Note that when only one sample of θ is used to estimate $f_{\bar{Z}}(\bar{z})$ the estimator is reduced to $\hat{f}_{\bar{Z}}(\bar{z}; \boldsymbol{\theta}) = f_{\bar{Z}}(\bar{z} | \theta)$. Also, note that $\hat{f}_{\bar{Z}}(\bar{z}; \boldsymbol{\theta}) \rightarrow f_{\bar{Z}}(\bar{z})$ point-wise when $M \rightarrow \infty$, by the law of large numbers. Indeed, since the random variable $\boldsymbol{\theta}$ has density $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \prod_{i=1}^M f_{\theta}(\theta^{(i)})$ we obtain

$$\begin{aligned} \int_{\Theta^M} \hat{f}_{\bar{Z}}(\bar{z}; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta} &= \int_{\Theta^M} \frac{1}{M} \sum_{i=1}^M f_{\bar{Z}}(\bar{z} | \theta^{(i)}) \prod_{i=1}^M f_{\theta}(\theta^{(i)}) d\theta^{(1)} \dots d\theta^{(M)} \\ &= \int_{\Theta} f_{\bar{Z}}(\bar{z} | \theta) f_{\theta}(\theta) d\theta = f_{\bar{Z}}(\bar{z}). \end{aligned}$$

Therefore the density $\bar{\pi}(\bar{z})$ constructed in Equation (18) is the marginal of the new target density defined on $\mathcal{Y}_M = \mathbb{R}^d \times \Theta^M$

$$\pi^y(y; \boldsymbol{\theta}) \propto \gamma^y(y; \boldsymbol{\theta}) = \hat{f}_{\bar{Z}}(\bar{z}; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \mathbb{1}_{\mathcal{G}_{\bar{Z}}}(\bar{z}).$$

Apart from the cumbersome notation, the same argument from the previous section can be used to show that a SMC procedure with estimated density $\hat{f}_{\bar{Z}}(\bar{z}; \boldsymbol{\theta})$ replacing $f_{\bar{Z}}(\bar{z})$ has unnormalized weights given by

$$w_t^y = w_{t-1}^y \frac{\hat{f}_{\bar{Z}}(\bar{z}_t; \boldsymbol{\theta}_t) \mathbb{1}_{\mathcal{G}_{\bar{Z}_t}}(\bar{z}_t) \bar{L}_{t-1}(\bar{z}_t, \bar{z}_{t-1} | \boldsymbol{\theta}_{t-1})}{\hat{f}_{\bar{Z}}(\bar{z}_{t-1}; \boldsymbol{\theta}_{t-1}) \mathbb{1}_{\mathcal{G}_{\bar{Z}_{t-1}}}(\bar{z}_{t-1}) \bar{K}_t(\bar{z}_{t-1}, \bar{z}_t | \boldsymbol{\theta}_t)},$$

when targeting a sequence $\{\bar{\pi}_t(\bar{z}_t)\}_{t=1}^T$ with $\bar{\pi}_T(\bar{z}_T) = \bar{\pi}(\bar{z})$.

The algorithms described in this section contain several degrees of freedom, whose choices are discussed in detail in Section 9. In the next section we formally define the elements necessary for constructing the statistical models underlying the risk drivers \bar{Z} and Z . We also present the formulas for the Solvency Capital Requirements (SCRs) under both the conditional and marginalized models.

After this brief introduction to SMC algorithms, in the following section we introduce the random variables used in the risk allocation process. In particular, we formally define the random vectors Z and \bar{Z} discussed in Section 2, and identify its components with the one-year reserve risk and the one-year premium risk.

4. Swiss Solvency Test and Claims Development

For the rest of this work we assume all random variables are defined in the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}(t)\}_{t \geq 0})$. We denote cumulative payments for accident year $i = 1, \dots, t$ until development year $j = 0, \dots, J$ (with $t > J$) on the $\ell = 1, \dots, L$ LoB by $C_{i,j}^{(\ell)}$. Moreover, in the ℓ -th LoB incremental payments for claims with accident year i and development year j are denoted by $X_{i,j}^{(\ell)} = C_{i,j}^{(\ell)} - C_{i,j-1}^{(\ell)}$. Remark that these payments are made in accounting year $i + j$.

The information (regarding claims payments) available at time $t = 0, \dots, I + J$ for the ℓ -th LoB is assumed to be given by

$$\mathcal{D}^{(\ell)}(t) = \{X_{i,j}^{(\ell)} : 1 \leq i \leq t, 0 \leq j \leq J, 1 \leq i + j \leq t\},$$

and, similarly, the total information (regarding claims payments) available at time t is denoted as

$$\mathcal{D}(t) = \bigcup_{1 \leq \ell \leq L} \mathcal{D}^{(\ell)}(t). \quad (20)$$

Remark 6. By a slight abuse of notation we also use $\mathcal{D}^{(\ell)}(t)$ and $\mathcal{D}(t)$ for the sigma-field generated by the corresponding sets. Note that $\mathcal{D}(t) \subset \mathcal{F}(t)$ for all $t \geq 0$, as we assume that $\mathcal{F}(t)$ contains not only information about claims payments, but also about premium and administrative costs.

The general aim now is to predict the future cumulative payments $C_{i,j}^{(\ell)}$ for $i + j > t$ at time t , given the information $\mathcal{F}(t)$, in particular, the so-called ultimate claim $C_{i,J}^{(\ell)}$. For more information we refer to [Wüthrich \(2015\)](#).

4.1. Conditional Predictive Model

As noted previously, we generically denote parameters in the Bayesian model for the ℓ LoB by $\theta^{(\ell)}$. For the ease of exposition, whenever a quantity is defined conditional on $\theta^{(\ell)}$ it is going to be denoted with a bar on top of it.

At time $t \geq I$, LoB ℓ and accident year $i > t - J$ predictors for the ultimate claim $C_{i,J}^{(\ell)}$ and the corresponding claims reserves are defined, respectively, as

$$\bar{\mathcal{C}}_{i,J}^{(\ell)}(t) = \mathbb{E}[C_{i,J}^{(\ell)} | \theta^{(\ell)}, \mathcal{F}(t)] \quad \text{and} \quad \bar{\mathcal{R}}_i^{(\ell)}(t) = \bar{\mathcal{C}}_{i,J}^{(\ell)}(t) - C_{i,t-i}^{(\ell)}. \quad (21)$$

Under modern solvency regulations, such as Solvency II [European Commission \(2009\)](#) and the Swiss Solvency Test [FINMA \(2007\)](#) an important variable to be analysed is the claims development result (CDR). For accident year $i = 1, \dots, I$, accounting year $t + 1 > I$ and LoB ℓ , the CDR is defined as

$$\begin{aligned} \overline{\text{CDR}}_i^{(\ell)}(t + 1) &= \bar{\mathcal{R}}_i^{(\ell)}(t) - (X_{i,t-i+1}^{(\ell)} + \bar{\mathcal{R}}_i^{(\ell)}(t + 1)) \\ &= \bar{\mathcal{C}}_{i,J}^{(\ell)}(t) - \bar{\mathcal{C}}_{i,J}^{(\ell)}(t + 1), \end{aligned} \quad (22)$$

and an application of the tower property of the expectation shows that (subject to integrability)

$$\mathbb{E}[\overline{\text{CDR}}_i^{(\ell)}(t + 1) | \theta^{(\ell)}, \mathcal{F}(t)] = 0. \quad (23)$$

Thus, the prediction process in Equation (21) is a martingale in t and we aim to study the volatility of these martingale innovations.

Equation (23) justifies the prediction of the CDR by zero and the uncertainty of this prediction can be assessed by the conditional mean squared error of prediction (mse):

$$\text{mse}_{\overline{\text{CDR}}_i^{(\ell)}(t+1) | \theta^{(\ell)}, \mathcal{F}(t)}(0) = \mathbb{E}[(\overline{\text{CDR}}_i^{(\ell)}(t+1) - 0)^2 | \theta^{(\ell)}, \mathcal{F}(t)] \quad (24)$$

$$\begin{aligned} &= \text{Var}(\overline{\text{CDR}}_i^{(\ell)}(t+1) | \theta^{(\ell)}, \mathcal{F}(t)) \\ &= \text{Var}(\overline{\mathcal{C}}_{i,J}^{(\ell)}(t+1) | \theta^{(\ell)}, \mathcal{F}(t)). \end{aligned} \quad (25)$$

Moreover, we denote the aggregated (over all accident years) CDR and the reserves, conditional on the knowledge of the parameter $\theta^{(\ell)}$, respectively, by

$$\overline{\text{CDR}}^{(\ell)}(t+1) = \sum_{i=t-J+1}^t \overline{\text{CDR}}_i^{(\ell)}(t+1) \quad \text{and} \quad \overline{\mathcal{R}}^{(\ell)}(t) = \sum_{i=t-J+1}^t \overline{\mathcal{R}}_i^{(\ell)}(t). \quad (26)$$

Using this notation we also define the total prediction uncertainty incurred when predicting $\overline{\text{CDR}}^{(\ell)}(t+1)$ by zero as

$$\text{mse}_{\overline{\text{CDR}}^{(\ell)}(t+1) | \theta^{(\ell)}, \mathcal{F}(t)}(0) = \text{Var} \left(\sum_{i=t-J+1}^t \overline{\mathcal{C}}_{i,J}^{(\ell)}(t+1) \mid \theta^{(\ell)}, \mathcal{F}(t) \right).$$

Remark 7. It should be remarked that, in general, as the parameter vector $\theta^{(\ell)}$ is unknown none of the quantities presented in this section can be directly calculated unless an explicit estimate for the parameter is used.

4.2. Marginalized Predictive Model

Even though cumulative claims models are defined conditional on unobserved parameter values, any quantity that has to be calculated based on these models should only depend on observable variables. Under the Bayesian paradigm, unknown quantities are modelled using a prior probability distribution reflecting prior beliefs about these parameters.

Analogously to Section 4.1 we define the marginalized (Bayesian) ultimate claim predictor and its reserves, respectively, as

$$\mathcal{C}_{i,J}^{(\ell)}(t) = \mathbb{E}[C_{i,J}^{(\ell)} | \mathcal{F}(t)] = \mathbb{E}[\overline{\mathcal{C}}_{i,J}^{(\ell)}(t) | \mathcal{F}(t)] \quad \text{and} \quad \mathcal{R}_i^{(\ell)}(t) = \mathcal{C}_{i,J}^{(\ell)}(t) - C_{i,t-i}^{(\ell)}. \quad (27)$$

We also define the marginalized CDR and notice, again using the tower property, that its mean is equal to zero

$$\text{CDR}_i^{(\ell)}(t+1) = \mathcal{C}_{i,J}^{(\ell)}(t) - \mathcal{C}_{i,J}^{(\ell)}(t+1) \quad \text{with} \quad \mathbb{E}[\text{CDR}_i^{(\ell)}(t+1) | \mathcal{F}(t)] = 0.$$

Furthermore, summing over all accident years i we follow Equation (26) and denote by $\mathcal{R}^{(\ell)}(t)$ and $\text{CDR}^{(\ell)}(t+1)$ the aggregated version of the marginalized reserves and CDR, where the uncertainty in the later is measured via

$$\text{mse}_{\text{CDR}^{(\ell)}(t+1) | \mathcal{F}(t)}(0) = \text{Var} \left(\sum_{i=t-J+1}^t \mathcal{C}_{i,J}^{(\ell)}(t+1) \mid \mathcal{F}(t) \right). \quad (28)$$

4.3. Solvency Capital Requirement (SCR)

In this section we discuss how two important concepts in actuarial risk management, namely the technical result (TR) and the solvency capital requirement (SCR), can be defined for both the conditional and the marginalized models.

In this context the TR is calculated netting all income and expenses arising from the LoBs, while the SCR denotes the minimum capital required by the regulatory authorities in order to cover the company's business risks. More precisely, the SCR for accounting year $t+1$ quantifies the risk of having a substantially distressed result at time $t+1$, evaluated in light of the available information at time t .

As an important shorthand notation, we introduce three sets of random variables, representing the total claim amounts of the current year (CY) claims and of prior year (PY) claims, the later for both the conditional and marginalized models. These random variables are defined, respectively, as

$$Z_{CY}^{(\ell)} = \mathcal{C}t + 1, J^{(\ell)}(t + 1), \quad \bar{Z}_{PY}^{(\ell)} = \sum_{i=t-J+1}^t \left(\bar{\mathcal{C}}_{i,J}^{(\ell)}(t + 1) - C_{i,t-i}^{(\ell)} \right) \quad \text{and}$$

$$Z_{PY}^{(\ell)} = \sum_{i=t-J+1}^t \left(\mathcal{C}i, J^{(\ell)}(t + 1) - C_{i,t-i}^{(\ell)} \right). \quad (29)$$

In the standard SST model, CY claims do not depend on any unknown parameters and are split into small claims $Z_{CY,s}^{(\ell)}$ for the LoBs $\ell = 1, \dots, L$ and into large events $Z_{CY,l}^{(p)}$ for the perils $p = 1, \dots, P$. Small claims are also called attritional claims and large claims can be individual large claims or catastrophic events, like earthquakes. In this context the company can choose thresholds $\beta^{(\ell)}$ such that claims larger than these amounts are classified as large claims in its respective LoBs.

To further simplify the notation we also group all the random variables related to the conditional and the marginalized models in two random vectors, defined as follows

$$\bar{\mathbf{Z}} = (\bar{Z}_1, \dots, \bar{Z}_{2L+P}) = (\bar{Z}_{PY}^{(1)}, \dots, \bar{Z}_{PY}^{(L)}, Z_{CY,s}^{(1)}, \dots, Z_{CY,s}^{(L)}, Z_{CY,l}^{(1)}, \dots, Z_{CY,l}^{(P)}), \quad (30)$$

$$\mathbf{Z} = (Z_1, \dots, Z_{2L+P}) = (Z_{PY}^{(1)}, \dots, Z_{PY}^{(L)}, Z_{CY,s}^{(1)}, \dots, Z_{CY,s}^{(L)}, Z_{CY,l}^{(1)}, \dots, Z_{CY,l}^{(P)}). \quad (31)$$

Next we give more details on how the TR and the SCR are calculated in the generic structure of the conditional and the marginalized models.

4.3.1. SCR for the Conditional Model

At time $t + 1$ the technical result (TR) of the ℓ -th LoB in accounting year $(t, t + 1]$ based on the conditional model is defined as the following $\mathcal{F}(t + 1)$ -measurable random variable:

$$\overline{\text{TR}}^{(\ell)}(t + 1) = \Pi^{(\ell)}(t + 1) - K^{(\ell)}(t + 1) - \mathcal{C}_{t+1,J}^{(\ell)}(t + 1) + \overline{\text{CDR}}^{(\ell)}(t + 1),$$

where $\Pi^{(\ell)}(t + 1)$ and $K^{(\ell)}(t + 1)$ are, respectively, the earned premium and the administrative costs of accounting year $(t, t + 1]$. For simplicity, we assume that these two quantities are known at time t , i.e., the premium and administrative costs of accounting year $(t, t + 1]$ are assumed to be previsible and, hence, $\mathcal{F}(t)$ -measurable. Moreover, it should be noticed that in this context $\mathcal{F}(t)$ not only includes the claims payment information defined in Equation (20). The general sigma-field $\mathcal{F}(t)$ should be seen as a sigma-field generated by the inclusion in $\mathcal{D}(t)$ of the information about $\Pi^{(\ell)}(t + 1)$ and $K^{(\ell)}(t + 1)$, for $\ell = 1, \dots, L$.

Given the technical result for all the LoBs, the company's overall TR based on the conditional model, and aggregated cost and premium are denoted, respectively, by

$$\overline{\text{TR}}(t + 1) = \sum_{\ell=1}^L \overline{\text{TR}}^{(\ell)}(t + 1), \quad \Pi(t + 1) = \sum_{\ell=1}^L \Pi^{(\ell)}(t + 1) \quad \text{and} \quad K(t + 1) = \sum_{\ell=1}^L K^{(\ell)}(t + 1).$$

In order to cover the company's risks over an horizon of one year, the Swiss Solvency Test is concerned with the 99% ES (in light of all the data up to time t):

$$\overline{\text{SCR}}(t + 1) = \text{ES}_{99\%}[-\overline{\text{TR}}(t + 1) \mid \mathcal{F}(t)],$$

where $\overline{\text{SCR}}$ denotes the solvency capital requirement.

It is important to notice that even though the ES operator is being applied to a "conditional random variable", namely $\overline{\text{TR}}$, the operator is *not* being taken conditional on the knowledge of $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(L)})$, otherwise this quantity would not be computable (as discussed in Remark 7). Instead, the SCR is calculated based on the marginalized version of the conditional model, where

the parameter uncertainty is integrated out. More precisely, the expected shortfall is based on the following (usually intractable) distribution

$$f_{\bar{Z}}(\bar{z} | \mathcal{F}(t)) = \int f_{\bar{Z}}(\bar{z} | \theta, \mathcal{F}(t)) \pi(\theta | \mathcal{F}(t)) d\theta.$$

In order to compute the SCR based on the conditional model we first discuss the measurability of the terms in the conditional TR, which can be rewritten as

$$\overline{\text{TR}}(t+1) = -K(t+1) + \Pi(t+1) + \sum_{\ell=1}^L \sum_{i=t-J+1}^t \left(\overline{\mathcal{C}}_{i,J}^{(\ell)}(t) - C_{i,t-i}^{(\ell)} \right) - \sum_{\ell=1}^L \left(\bar{Z}_{PY}^{(\ell)} + Z_{CY}^{(\ell)} \right).$$

From the above equation we see the first two terms are, by assumption, $\mathcal{F}(t)$ measurable and so are all the terms of the form $C_{i,t-i}^{(\ell)}$ (payments already completed by time t), while the last summation is $\mathcal{F}(t+1)$ measurable and, therefore, a random variable at time t . Due to the dependence on the unknown parameter θ the conditional ultimate claim predictor $\overline{\mathcal{C}}_{i,J}^{(\ell)}(t)$ is usually *not* $\mathcal{F}(t)$ measurable. However, under the special models introduced in Section 5 we have that $\overline{\mathcal{C}}_{i,J}^{(\ell)}(t)$ depends only on the claims data up to time t and not on the unknown parameter vector, making it $\mathcal{F}(t)$ measurable. In this case one has

$$\overline{\text{SCR}}(t+1) = K(t+1) - \Pi(t+1) - \sum_{\ell=1}^L \overline{\mathcal{R}}^{(\ell)}(t) + \text{ES}_{99\%} \left[\sum_{\ell=1}^L \bar{Z}_{PY}^{(\ell)} + Z_{CY}^{(\ell)} \mid \mathcal{F}(t) \right], \quad (32)$$

where, by assumption, $\sum_{\ell=1}^L \overline{\mathcal{R}}^{(\ell)}(t) = \sum_{\ell=1}^L \sum_{i=t-J+1}^t \left(\overline{\mathcal{C}}_{i,J}^{(\ell)}(t) - C_{i,t-i}^{(\ell)} \right)$ is $\mathcal{F}(t)$ -measurable.

4.3.2. SCR for the Marginalized Model

As the parameter uncertainty is dealt with in a previous step, the calculation of the SCR for the marginalized model is simpler than its conditional counterpart.

Similarly to the conditional case, we define the TR for the marginalized model as

$$\text{TR}^{(\ell)}(t+1) = \Pi^{(\ell)}(t+1) - K^{(\ell)}(t+1) - \mathcal{C}_{t+1,J}^{(\ell)}(t+1) + \text{CDR}^{(\ell)}(t+1),$$

and its aggregated version as

$$\text{TR}(t+1) = \sum_{\ell=1}^L \text{TR}^{(\ell)}(t+1).$$

Furthermore, the SCR for the marginalized model is given by

$$\text{SCR}(t+1) = \text{ES}_{99\%}[-\text{TR}(t+1) \mid \mathcal{F}(t)] \quad (33)$$

$$= K(t+1) - \Pi(t+1) - \sum_{\ell=1}^L \overline{\mathcal{R}}^{(\ell)}(t) + \text{ES}_{99\%} \left[\sum_{\ell=1}^L Z_{PY}^{(\ell)} + Z_{CY}^{(\ell)} \mid \mathcal{F}(t) \right], \quad (34)$$

where in this case the expected shortfall is calculated with respect to the density $f_{\mathbf{Z}}(\mathbf{z} | \mathcal{F}(t))$.

Remark 8. For the models discussed in Section 5, as $\overline{\mathcal{C}}_{i,J}^{(\ell)}(t)$ does not depend on the parameter vector θ and we also have that $\overline{\mathcal{R}}^{(\ell)}(t) = \mathcal{R}^{(\ell)}(t)$.

Remark 9. As we assume the cost of claims processing and assessment $K(t+1)$ and premium $\Pi(t+1)$ are known at time t they do not differ from the conditional to the marginalized model.

5. Modelling of Individual LoBs PY Claims

For the modelling of the PY claims reserving risk we need to model \bar{Z}_{PY} or Z_{PY} as given in Equation (29). The uncertainty in these random variables will be assessed by the conditional and marginalized mean square error of prediction (mse_p), introduced in Equations (25) and (28). In order to calculate the mse_p we must first expand our analysis to the study of the claims reserving uncertainty. To do so, in this section we present a fully Bayesian version of the gamma-gamma chain-ladder (CL) model, which has been studied in Peters et al. (2017).

Since in this section we present the model for individual LoBs, for notational simplicity we omit the upper index (ℓ) from all random variables and parameters.

Model Assumptions 1. [Gamma-gamma Bayesian chain ladder model] We make the following assumptions:

- (a) Conditionally, given $\boldsymbol{\phi} = (\phi_0, \dots, \phi_{J-1})$ and $\boldsymbol{\sigma} = (\sigma_0, \dots, \sigma_{J-1})$, cumulative claims $(C_{i,j})_{j=0,\dots,J}$ are independent (in accident year i) Markov processes (in development year j) with

$$C_{i,j+1} \mid \{\mathcal{F}(i+j), \boldsymbol{\phi}, \boldsymbol{\sigma}\} \sim \Gamma(C_{i,j}\sigma_j^{-2}, \phi_j\sigma_j^{-2}),$$

for all $1 \leq i \leq t$ and $0 \leq j \leq J-1$.

- (b) The parameter vectors $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}$ are independent.

- (c) For given hyper-parameters $f_j > 0$ the components of $\boldsymbol{\phi}$ are independent such that

$$\phi_j \sim \lim_{\gamma_j \rightarrow 1} \Gamma(\gamma_j, f_j(\gamma_j - 1)),$$

for $0 \leq j \leq J-1$, where the limit infers that they are eventually distributed from an improper uninformative prior.

- (d) The components σ_j of $\boldsymbol{\sigma}$ are independent and F_{σ_j} -distributed, having support in $(0, d_j)$ for given constants $0 < d_j < \infty$ for all $0 \leq j \leq J-1$.

- (e) $\boldsymbol{\phi}$, $\boldsymbol{\sigma}$ and $C_{1,0}, \dots, C_{t,0}$ are independent and $\mathbb{P}[C_{i,0} > 0] = 1$, for all $1 \leq i \leq t$.

In Model Assumptions 1 (c) the (improper) prior distribution for $\boldsymbol{\phi}$ should be seen as a non-informative limit when $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_{J-1}) \rightarrow \mathbf{1} = (1, \dots, 1)$ of the (proper) prior assumption

$$\phi_j \sim \Gamma(\gamma_j, f_j(\gamma_j - 1)).$$

The limit in (c) does not lead to a proper probabilistic model for the prior distribution, however, based on “reasonable” observations $\{C_{i,j}\}_{i,j}$ the posterior model can be shown to be well defined (see Equation (38)), a result that has been proved using the dominated convergence theorem in Peters et al. (2017).

From Model Assumptions 1 (a), conditional on a specific value of the parameter vectors $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}$, we have that

$$\begin{aligned} \mathbb{E}[C_{i,j+1} \mid \mathcal{F}(i+j), \boldsymbol{\phi}, \boldsymbol{\sigma}] &= \phi_j^{-1} C_{i,j}, \\ \text{Var}(C_{i,j+1} \mid \mathcal{F}(i+j), \boldsymbol{\phi}, \boldsymbol{\sigma}) &= \phi_j^{-2} \sigma_j^2 C_{i,j}, \end{aligned} \quad (35)$$

which provides a stochastic formulation of the classical CL model of Mack (1993).

Even though the prior is assumed improper and does not integrate to one, the conditional posterior for $\phi_j \mid \sigma_j, \mathcal{F}(t)$ is proper and, in addition, also gamma distributed (see Appendix A and (Merz and Wüthrich 2015, Lemma 3.2)). More precisely, we have that

$$\phi_j \mid \boldsymbol{\sigma}, \mathcal{F}(t) \sim \Gamma(a_j, b_j), \quad (36)$$

with the following parameters

$$a_j = 1 + \sum_{i=1}^{t-j-1} C_{i,j} \sigma_j^{-2} \quad \text{and} \quad b_j = \sum_{i=1}^{t-j-1} C_{i,j+1} \sigma_j^{-2}. \quad (37)$$

Therefore, given σ this model belongs to the family of Bayesian models with conjugate priors that allows for closed form (conditional) posteriors – for details see [Wüthrich \(2015\)](#).

The marginal posterior distribution of the elements of the vector σ is given by

$$\pi(\sigma_j | \mathcal{F}(t)) \propto h_j(\sigma_j | \mathcal{F}(t)) = \Gamma(a_j) b_j^{-a_j} f_{\sigma_j}(\sigma_j) \prod_{i=1}^{t-j-1} \frac{(C_{i,j+1} \sigma_j^{-2})^{C_{i,j} \sigma_j^{-2}}}{\Gamma(C_{i,j} \sigma_j^{-2})}, \quad (38)$$

with a_j and b_j defined in Equation (37). We note that as long as Model Assumptions 1 (d) and the conditions in Lemma A1 are satisfied, then one can ensure the posterior distribution of σ is proper.

Therefore, under Model Assumptions 1 inference for all the unknown parameters can be performed. It should be noticed, though, that differently from the (conditional) posteriors for ϕ_j Equation (36), the posterior for σ_j Equation (38) is not recognized as a known distribution. Thus, whenever expectations with respect to the distribution of $\sigma_j | \mathcal{F}(t)$ need to be calculated one needs to make use of numerical procedures, such as numerical integration or Markov Chain Monte Carlo (MCMC) methods.

5.1. MSEP Results Conditional on σ

Following Model Assumptions 1 we now discuss how to explicitly calculate the quantities introduced in Section 4. We start with the equivalent of the classic CL factor. From the model structure in Equation (35) we define the posterior Bayesian CL factors, given σ , as

$$\hat{f}_j(t) = \mathbb{E}[\phi_j^{-1} | \sigma_j, \mathcal{F}(t)], \quad (39)$$

which, using the gamma distribution from Equation (36), takes the form

$$\hat{f}_j(t) = \frac{\sum_{k=1}^{t-j-1} C_{k,j+1}}{\sum_{k=1}^{t-j-1} C_{k,j}},$$

i.e., $\hat{f}_j(t)$ is identical to the classic CL factor estimate.

Following Equation (21) we define the conditional ultimate claim predictor

$$\bar{\mathcal{C}}_{i,J}(t) = \mathbb{E}[C_{i,J} | \sigma, \mathcal{F}(t)] = \mathbb{E}_{\phi} \left[\mathbb{E}[C_{i,J} | \phi, \sigma, \mathcal{F}(t)] \mid \sigma, \mathcal{F}(t) \right],$$

which can be shown (see ([Wüthrich 2015](#), Theorem 9.5)) to be equal to

$$\bar{\mathcal{C}}_{i,J}(t) = C_{i,t-i} \prod_{j=t-i}^{J-1} \hat{f}_j(t), \quad (40)$$

where this is exactly the classic chain ladder predictor of [Mack \(1993\)](#). For this reason we may take Model Assumptions 1 as a distributional model for the classical CL method. Additionally, the conditional reserves defined in Equation (21) and Equation (26) are also the same as the classic CL ones, that is,

$$\bar{\mathcal{R}}(t) = \sum_{i=1}^t \bar{\mathcal{C}}_{i,J}(t) - C_{i,t-i}. \quad (41)$$

The importance of Equation (40) relies on the fact that it does not depend on the parameter vector σ . In other words, the ultimate claim predictor based on the Bayesian model from Model Assumptions 1 conditional on σ – which is, in general, a random variable – is a real number (independent of σ). This justifies the argument used on the calculation of Equation (32).

Remark 10. Using the notation from the previous sections the parameter vector σ plays the role of θ as the only unknown, since, due to conjugacy properties, ϕ can be marginalized analytically.

For the Bayesian model from Model Assumptions 1 the mse conditional on σ has been derived in (Wüthrich 2015, Theorem 9.16) as follows, for $i + J > t$

$$\text{mse}_{\overline{\text{CDR}}_i(t+1) | \sigma, \mathcal{F}(t)}(0) = (\overline{\mathcal{C}}_{i,J}(t))^2 \left[\left(1 + \frac{\overline{\Psi}_{t-i}(t)}{\overline{\beta}_{t-i}(t)} \right) \prod_{j=t-i+1}^{J-1} \left(1 + \overline{\beta}_j(t) \overline{\Psi}_j(t) \right) - 1 \right], \quad (42)$$

where

$$\overline{\beta}_j(t) = \frac{C_{t-j,j}}{\sum_{i=1}^{t-j} C_{i,j}} \quad \text{and} \quad \overline{\Psi}_j(t) = \frac{\sigma_j^2}{\sum_{k=1}^{t-j-1} C_{k,j} - \sigma_j^2}. \quad (43)$$

Moreover, the conditional mse has been shown to be finite if, and only if, $\sigma_j^2 < \sum_{k=1}^{t-j-1} C_{k,j}$. We also refer to Remark 12, below.

The aggregated conditional mse for $\overline{\text{CDR}}(t+1) = \sum_{i=1}^t \overline{\text{CDR}}_i(t+1)$ is also derived in (Wüthrich 2015, Theorem 9.16), and given by

$$\begin{aligned} \text{mse}_{\overline{\text{CDR}}(t+1) | \sigma, \mathcal{F}(t)}(0) &= \sum_{i=t-J+1}^t \text{mse}_{\overline{\text{CDR}}_i(t+1) | \sigma, \mathcal{F}(t)}(0) \\ &+ 2 \sum_{t-J+1 \leq i < k \leq t} \overline{\mathcal{C}}_{i,J}(t) \overline{\mathcal{C}}_{k,J}(t) \left[\left(1 + \overline{\Psi}_{t-i}(t) \right) \prod_{j=t-i+1}^{J-1} \left(1 + \overline{\beta}_j(t) \overline{\Psi}_j(t) \right) - 1 \right]. \end{aligned} \quad (44)$$

Remark 11. The assumption that $\sigma_j^2 < \sum_{k=1}^{t-j-1} C_{k,j}$ is made in order to guarantee the conditional mse is finite and we enforce this assumption to hold for all the examples presented in this work. See also Remark 12, below.

5.2. Marginalized MSE Results

The results in the previous section are based on derivations presented in Merz and Wüthrich (2015) and Wüthrich (2015) where the parameter vector σ is assumed to be known. In this section we study the impact of the uncertainty in σ over the mean and variance of $\mathcal{C}_{i,J}(t+1) | \mathcal{F}(t)$ in light of Model Assumptions 1, which can be seen as a fully Bayesian version of the models previously mentioned.

In order to have well defined posterior distributions for σ , through this section we follow Lemma A1 and assume that, for all development years $0 \leq j \leq J-1$ and $t \geq I$, we have $(t-j-1) \wedge I = 1$ or at least one accident year $1 \leq i \leq (t-j-1) \wedge I$ is such that $\frac{C_{i,j+1}}{C_{i,j}} \neq \hat{f}_j(t)$. For all the numerical results presented this assumption is satisfied.

Lemma 1. The ultimate claim estimator under the marginalized model is equal to the classic chain ladder predictor, i.e., $\mathcal{C}_{i,J}(t) = \mathbb{E}[C_{i,J} | \mathcal{F}(t)] = \overline{\mathcal{C}}_{i,J}(t)$.

Proof. Due to the posterior independence of the elements of ϕ (also used in Equations (39) and (40)) and the fact that $\mathcal{C}_{i,J}(t)$ does not depend on σ we have

$$\begin{aligned}\mathcal{C}_{i,J}(t) &= \mathbb{E}[C_{i,J} | \mathcal{F}(t)] \\ &= \mathbb{E}[\mathbb{E}[C_{i,J} | \phi, \sigma, \mathcal{F}(t)] | \mathcal{F}(t)] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{E}[C_{i,J} | \phi, \sigma, \mathcal{F}(t)] | \sigma, \mathcal{F}(t)] | \mathcal{F}(t)] \\ &= \mathbb{E}[\mathbb{E}[C_{i,t-i} \prod_{j=t-i}^{J-1} \phi_j^{-1} | \sigma, \mathcal{F}(t)] | \mathcal{F}(t)] \\ &= \mathbb{E}[\overline{\mathcal{C}}_{i,J}(t) | \mathcal{F}(t)] = \overline{\mathcal{C}}_{i,J}(t).\end{aligned}$$

□

Proposition 1. The mse in the marginalized model is equal to the posterior expectation of the mse in the conditional model, i.e.,

$$\begin{aligned}mse_{CDR(t+1)|\mathcal{F}(t)}(0) &= Var\left(\sum_{i=1}^I \mathcal{C}_{i,J}(t+1) | \mathcal{F}(t)\right) \\ &= \mathbb{E}[mse_{\overline{CDR}(t+1)|\sigma, \mathcal{F}(t)}(0) | \mathcal{F}(t)].\end{aligned}\quad (45)$$

Proof. From the law of total variance we have that

$$\begin{aligned}Var\left(\sum_{i=1}^I \mathcal{C}_{i,J}(t+1) | \mathcal{F}(t)\right) &= Var\left(\mathbb{E}\left[\sum_{i=1}^I \mathcal{C}_{i,J}(t+1) | \mathcal{F}(t), \sigma\right] | \mathcal{F}(t)\right) \\ &\quad + \mathbb{E}\left[Var\left(\sum_{i=1}^I \mathcal{C}_{i,J}(t+1) | \mathcal{F}(t), \sigma\right) | \mathcal{F}(t)\right] \\ &= \mathbb{E}\left[Var\left(\sum_{i=1}^I \mathcal{C}_{i,J}(t+1) | \mathcal{F}(t), \sigma\right) | \mathcal{F}(t)\right],\end{aligned}$$

and the last equality follows from Lemma 1 and the fact that $\mathbb{E}[\overline{\mathcal{C}}_{i,J}(t+1) | \mathcal{F}(t), \sigma] = \overline{\mathcal{C}}_{i,J}(t)$ is independent of σ . □

Remark 12. Following the conditions required for finiteness of the conditional mse, in the unconditional case, one can see that $mse_{CDR(t+1)|\mathcal{F}(t)}(0) < \infty$ whenever $\sum_{k=1}^{t-j-1} C_{k,j} > d_j^2$. Furthermore, we note that this condition can be controlled during the model specification, i.e., the range of the σ_j^2 is chosen such that all posteriors are well-defined.

5.3. Statistical Model of PY Risk in the SST

Note that the distributional models derived in Sections 5.1 and 5.2 are rather complex. To maintain some degree of tractability, the overall PY uncertainty distribution is usually approximated by a log-normal distribution via a moment matching procedure.

5.3.1. Conditional PY Model

As discussed in Section 4.3, when modelling the risk of PY claims we work with the random variables \overline{Z}_{PY} , defined in Equation (29). Due to their relationship with the conditional CDR, see Equations (22) and (23) and the results discussed in Section 5.1, we can use the derived properties of these random variables to construct the model being used for \overline{Z}_{PY} .

The conditional mean (see Equations (22), (23) and (41)) and variance (see Equations (25) and (44)) of the random variable \bar{Z}_{PY} are as follows

$$\mathbb{E}[\bar{Z}_{PY} | \sigma, \mathcal{F}(t)] = \bar{\mathcal{R}}(t), \quad (46)$$

$$\text{Var}(\bar{Z}_{PY} | \sigma, \mathcal{F}(t)) = \text{mse}_{\bar{\text{CDR}}(t+1) | \sigma, \mathcal{F}(t)}(0). \quad (47)$$

Given mean and variance, we make the following approximation, also proposed in the Swiss Solvency Test (see (FINMA 2007, sct. 4.4.10)).

Model Assumptions 2 (Conditional log-normal approximation). *We assume that*

$$\bar{Z}_{PY} | \sigma, \mathcal{F}(t) \sim \text{LN}(\bar{\mu}_{PY}, \bar{\sigma}_{PY}^2),$$

$$\text{with } \bar{\sigma}_{PY}^2 = \log \left(\frac{\text{mse}_{\bar{\text{CDR}}(t+1) | \sigma, \mathcal{F}(t)}(0)}{\bar{\mathcal{R}}(t)^2} + 1 \right) \text{ and } \bar{\mu}_{PY} = \log(\bar{\mathcal{R}}(t)) - \frac{\bar{\sigma}_{PY}^2}{2}.$$

Although the distribution of $\bar{Z}_{PY} | \sigma, \mathcal{F}(t)$ under Model Assumptions 1 can not be described analytically it is simple to simulate from it. To test the approximation of Model Assumptions 2 we simulate its distribution under the gamma-gamma Bayesian CL model (with fixed σ) and compare it against the log-normal approximation proposed. For the hyper-parameters presented in Table 2 (and calculated in Section 8) the quantile-quantile plot of the approximation is presented in Figure 1. For all the LoBs we see that the log-normal distribution is a sensible approximation to the original model assumptions. Note that although the parameters used for the comparison are based on the marginalized model Figures 5 and 6 show that they are “representative” values for the distributions of $\bar{\mu}_{PY}$ and $\bar{\sigma}_{PY}$.

5.3.2. Marginalized PY Model

As an alternative to the conditional Model Assumptions 2 we use the moments of $Z_{PY} | \mathcal{F}(t)$ calculated in Lemma 1 and Proposition 1 and then approximate its distribution. Note that due to the intractability of the distribution of $\sigma | \mathcal{F}(t)$ the variance term defined in Equation (45) can only be calculated numerically, for example, via MCMC.

Model Assumptions 3 (Marginalized log-normal approximation). *We assume that*

$$Z_{PY} | \mathcal{F}(t) \sim \text{LN}(\mu_{PY}, \sigma_{PY}^2)$$

$$\text{with } \sigma_{PY}^2 = \log \left(\frac{\text{mse}_{\text{CDR}(t+1) | \mathcal{F}(t)}(0)}{\bar{\mathcal{R}}(t)^2} + 1 \right) \text{ and } \mu_{PY} = \log(\bar{\mathcal{R}}(t)) - \frac{\sigma_{PY}^2}{2}.$$

The same comparison based on the quantile-quantile plot of Figure 1 can be performed for the marginalized model and the results are presented in Figure 2. Once again, the log-normal model presents a viable alternative to the originally postulated gamma-gamma Bayesian CL model, even though for Motor Hull, Property and Others the right tail of the log-normal distribution is slightly heavier.

Table 2. Parameters and capital calculations for the marginalized and conditional models.

LoB	Reserve/ Premium	σ	μ	CoVa	Expectation	Standalone		Marginalized			Conditional		
						ES _{99%}	SCR	ES _{99%}	SCR	Div. Benefit	ES _{99%}	SCR	Div. Benefit
1	2365.44	0.0287	7.7659	2.87%	2365.44	2546.31	180.87	2489.85	124.41	31.22%	2492.05	126.61	30.00%
2	99.37	0.2164	4.5755	21.90%	99.37	173.23	73.86	131.73	32.36	56.19%	132.59	33.21	55.03%
3	405.99	0.1142	5.9998	11.46%	405.99	547.25	141.26	479.11	73.12	48.24%	485.27	79.28	43.88%
4	870.19	0.0315	6.7682	3.15%	870.19	946.06	75.87	905.48	35.29	53.49%	905.29	35.10	53.73%
5	1105.95	0.0193	7.0083	1.93%	1105.95	1,164.04	58.09	1137.06	31.11	46.44%	1136.88	30.93	46.76%
6	274.91	0.0410	5.6156	4.10%	274.91	306.43	31.52	287.33	12.42	60.59%	286.97	12.06	61.74%
7	7.150	0.0547	1.9657	5.48%	7.15	8.26	1.11	7.45	0.30	73.27%	7.43	0.28	74.50%
8	48.18	0.0493	3.8738	4.93%	48.18	54.89	6.71	50.51	2.32	65.36%	50.43	2.25	66.44%
9	72.20	0.1332	4.2706	13.38%	72.2	102.16	29.96	85.32	13.12	56.21%	85.15	12.95	56.77%
Total PY	5249.38				5249.38	5848.63	599.25	5573.84	324.45	45.86%	5582.06	332.67	44.49%
1	503.14	0.0685	6.0958	6.86%	448.94	533.07	84.13	499.16	50.21	40.32%	498.37	49.43	41.25%
2	573.26	0.0702	6.0356	7.03%	402.87	504.20	101.33	472.25	69.38	31.53%	471.66	68.79	32.11%
3	748.76	0.0683	6.3013	6.84%	547.23	654.38	107.15	603.36	56.13	47.62%	602.61	55.38	48.31%
4	299.73	0.0923	5.3596	9.25%	216.70	272.05	55.35	239.69	22.99	58.47%	239.57	22.87	58.69%
5	338.63	0.0648	5.6841	6.49%	303.77	349.69	45.92	319.17	15.40	66.47%	318.71	14.94	67.45%
6	254.21	0.0804	5.4296	8.05%	228.79	282.62	53.83	249.63	20.85	61.28%	249.31	20.52	61.88%
7	7.20	0.1047	1.8628	10.5%	6.48	8.52	2.04	7.01	0.53	73.84%	7.01	0.53	74.06%
8	34.64	0.0981	3.3172	9.84%	27.72	35.84	8.13	30.32	2.60	67.95%	30.28	2.57	68.44%
9	46.28	0.1004	3.6066	10.06%	37.03	48.16	11.14	41.83	4.81	56.83%	41.79	4.77	57.19%
Total CY,s	2805.85				2219.53	2688.53	469.02	2462.42	242.9	48.21%	2459.31	239.8	48.87%

Peril	$\beta^{(5)}$	γ	α	CoVa	Expectation	Standalone		Marginalized			Conditional		
						ES _{99%}	SCR	ES _{99%}	SCR	Div. benefit	ES _{99%}	SCR	Div. benefit
1	2.50		2.80		3.89	20.14	16.25	4.03	0.15	99.1%	4.01	0.12	99.27%
2	13.35	300	1.85		27.08	191.21	164.13	39.96	12.88	92.15%	39.61	12.53	92.36%
3	6.28	100	1.50		14.34	84.31	69.97	16.5	2.16	96.91%	16.45	2.11	96.98%
4	3.88	100	1.80		8.10	61.34	53.24	8.94	0.84	98.42%	8.91	0.81	98.48%
5	0.50		2.00		1.00	10.00	9.00	1.07	0.07	99.19%	1.12	0.12	98.69%
Total CY,l					54.41	367	312.59	70.5	16.1	94.85%	70.1	15.69	94.98%
Total	8055.26				7523.32	8904.18	1380.86	8106.77	583.45	57.75%	8111.5	588.18	57.40%

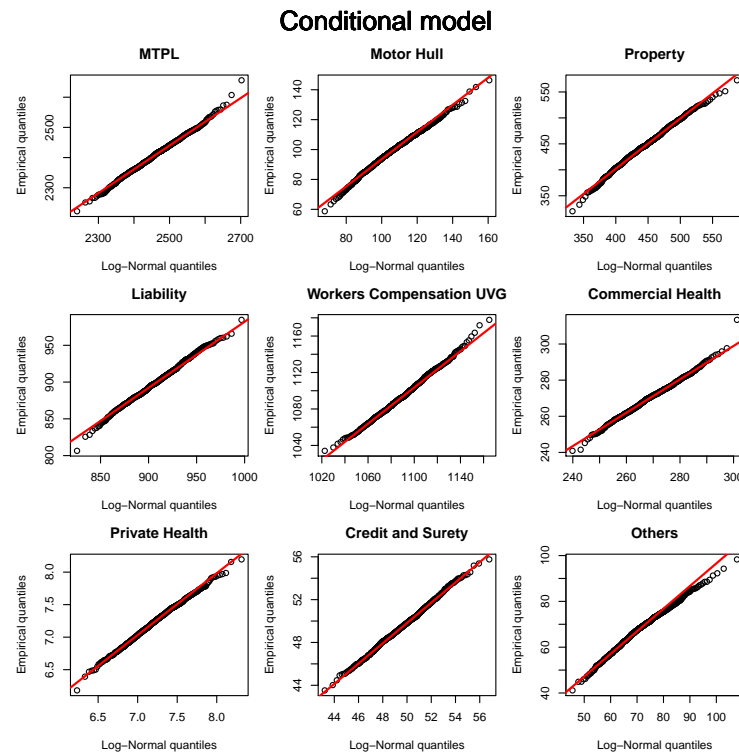


Figure 1. Quantile-Quantile plots for the different lines of business (LoBs) comparing (vertical axis) the empirical distribution of $\bar{Z}_{PY} | \sigma, \mathcal{F}(t)$ based on Model Assumptions 1 and (horizontal axis) the log-normal approximation from Model Assumptions 2. Based on 1000 samples.

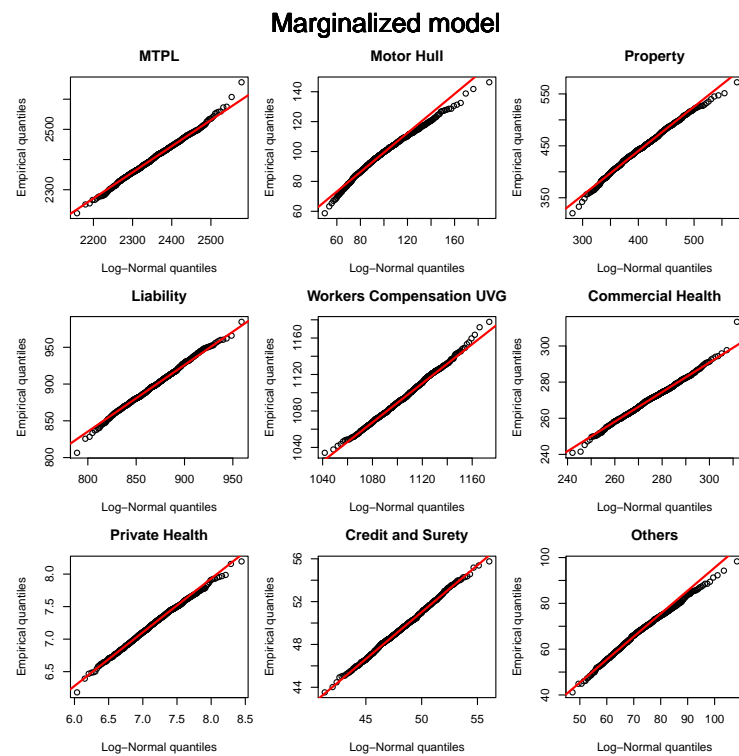


Figure 2. Quantile-Quantile plots for the different LoBs comparing (vertical axis) the empirical distribution of $Z_{PY} | \mathcal{F}(t)$ based on Model Assumptions 1 and (horizontal axis) the log-normal approximation from Model Assumptions 3 and using posterior samples as in Figures 5 and 6. Based on 1000 samples.

6. Modelling of Individual LoBs CY Claims

Model Assumptions 1 do not assume any specific distribution for $\mathbb{E}[C_{t+1,l} | \mathcal{F}(t+1)]$, the CY claims. These claims are treated differently in the Swiss Solvency Test from PY claims and the models used for these claims are explained in Sections 6.1 and 6.2, below. Throughout this section, we denote by $\lambda_{CY} = \lambda_{CY,s} + \lambda_{CY,l}$ the expected number of CY claims over the next year, which is the sum of the expected CY small claims $\lambda_{CY,s}$ and the expected CY large claims $\lambda_{CY,l}$.

6.1. Modelling of Small CY Claims

As mentioned in the SST Technical Document (FINMA 2007, sct. 4.4.7), the SST does not make any explicit assumption about the distribution of individual claims; instead, the annual claims expenses are only represented with their expected value and variance. More precisely, in (FINMA 2007, sct. 8.4.5.2) the distribution of the premium risk, $Z_{CY,s}$ is assumed to be such that

$$\text{CoVa}^2(Z_{CY,s} | \mathcal{F}(t)) = a_1 + \frac{a_2 + 1}{\lambda_{CY,s}}, \quad (48)$$

where the constants a_1 and a_2 are provided by the regulatory authority (under the names of *parameter uncertainty* and *random fluctuation*, respectively). Their values for the 2015 solvency test are found in FINMA (2016). In order to fully specify the model for CY small claims one also needs to decide on the mean of the variable $Z_{CY,s} | \mathcal{F}(t)$, but we postpone a detailed discussion on this point until Section 8.2, where we also present the value of $\lambda_{CY,s}$.

Model Assumptions 4 (Distribution of CY small claims). For known constants $v, r_s > 0$ and $\mathbb{E}[Z_{CY,s} | \mathcal{F}(t)]$ we set

$$Z_{CY,s} | \mathcal{F}(t) \sim LN(\mu_{CY,s}, \sigma_{CY,s}^2),$$

$$\text{with } \sigma_{CY,s}^2 = \log\left(a_1 + \frac{a_2 + 1}{\lambda_{CY,s}} + 1\right) \text{ and } \mu_{CY,s} = \log(\mathbb{E}[Z_{CY,s} | \mathcal{F}(t)]) - \frac{\sigma_{CY,s}^2}{2}.$$

6.2. Modelling of Large CY Claims

In the SST (see (FINMA 2007, sct. 4.4.8)), large CY claims are split into two groups. The first group of large claims are those triggered by the same market-wide event (a hailstorm, for example) and with many simultaneous (small) claims. These types of claims are likely to affect all market participants and are called “cumulated claims”. The second group encompasses individual claims with a large claim amount, which includes, as exemplified in (FINMA 2007, sct. 4.4.8), fire in a factory building.

For each risk trigger, CY large claims are required to be modelled as a compound Poisson random variable with i.i.d. Pareto severities, i.e.,

$$Z_{CY,l} = \sum_{k=1}^N Y_k, \quad (49)$$

where $N \sim \text{Pois}(\lambda)$ is the number of large claims in LoB under consideration and $Y_k \stackrel{\text{i.i.d.}}{\sim} \text{Pareto}(\beta, \alpha_\beta)$ model the intensity of large claims. Here we denote by $X \sim \text{Pareto}(\beta, \alpha_\beta)$ a random variable with density $f(x) = \frac{\alpha_\beta \beta^\alpha}{x^{\alpha+1}}$, for $x \geq \beta$. It is assumed in the SST that large claims are i.i.d. within the same risk trigger and also between different risk triggers, and independent of all Z_{PY} and $Z_{CY,s}$.

As a notational remark, if Z follows a Compound Poisson – Pareto model as a shorthand notation we write $Z \sim \text{CP-P}(\lambda, \beta, \alpha)$, with the same parameter interpretation as in Equation (49).

6.2.1. SST Model for Cumulated Claims

In this section we discuss the modelling of cumulated claims (those triggered by a market-wide event) which are modelled as an event that impacts the whole market and then scales down to

an individual insurance company through its market share. In particular, we present the modelling approach used in (1) Motor Hull LoB due to hail events and (2) Workers Compensation (UVG) LoB due to a market-wide large accident.

In both cases market-wide parameters for a compound Poisson model with Pareto intensities have been determined by the regulator, (based on a large claims data set). The aggregated market-wide loss is given by

$$Z_{mkt} = \sum_{k=1}^{N_{mkt}} Y_{k,mkt} \sim \text{CP-P}(\lambda_{mkt}, \beta_{mkt}, \alpha_{mkt}),$$

where $\text{CP-P}(\lambda_{mkt}, \beta_{mkt}, \alpha_{mkt})$ denotes a compound distribution with frequency given by $\text{Pois}(\lambda_{mkt})$ and severity given by $\text{Pareto}(\beta_{mkt}, \alpha_{mkt})$. The corresponding market-wide parameter values are found in [FINMA \(2016\)](#).

Denoting by β the company's threshold after which losses are classified as large and m its market share in the ℓ -th LoB, to be consistent with its assumption the company should model market-wide large events as events above the threshold of

$$\beta^* = \frac{\beta}{m}.$$

Then, the market-wide total loss (viewed from the specific company in consideration) is defined as

$$Z^* = \sum_{k=1}^{N^*} Y_k^* \sim \text{CP-P}(\lambda^*, \beta^*, \alpha_{mkt}),$$

from which it is easy to see that the only unknown parameter is λ^* , since in the SST the Pareto parameter α_{mkt} is kept the same. This frequency parameter is chosen such that the company's view of the market-wide events is equivalent to the suggested market-wide process. In other words, $\lambda_{mkt} = \mathbb{P}[Y_k^* > \beta_{mkt}] \lambda^*$ hence

$$\lambda^* = \lambda_{mkt} \left(\frac{\beta/m}{\beta_{mkt}} \right)^{-\alpha_{mkt}}. \quad (50)$$

Therefore, from the company's point of view, its own large claims are modelled as

$$Z_{comp} \sim \text{CP-P}(\lambda^*, \beta, \alpha_{mkt}).$$

Following the SST Technical Document [FINMA \(2007\)](#), an upper bound γ (provided by the regulator) is included in each Pareto random variable within the random sum. In other words, the final distribution of the company's large cumulated claims is given by

$$\tilde{Z} = \sum_{k=1}^{N^*} \tilde{Y}_k \sim \text{CP-P}(\lambda^*, \beta, \alpha_{mkt}, \gamma),$$

where $\tilde{Y}_k \sim \text{Pareto}(\beta, \alpha_{mkt}, \gamma)$, a Pareto distribution defined in $[\beta, \gamma]$ with tail index α_{mkt} .

For efficiency purposes, this distribution is approximated by a single Pareto, with the same mean. This leads us to the following model assumptions.

Model Assumptions 5 (Marginal distribution of cumulated claims). For α_{mkt} , β_{mkt} and γ provided by the regulator in [FINMA \(2016\)](#), $\beta \in \{1, 5\}$, $m \in (0, 1)$,

$$Z_{CY,l} \sim \text{Pareto} \left(\lambda^* \frac{\beta^{\alpha_{mkt}}}{1 - (\beta/\gamma)^{\alpha_{mkt}}} \left(\frac{1}{\beta^{\alpha_{mkt}-1}} - \frac{1}{\gamma^{\alpha_{mkt}-1}} \right), \alpha_{mkt} \right)$$

where λ^* is defined in Equation (50).

Remark 13. The reader should note that for large CY claims no parameter uncertainty is considered, since both λ_{mkt} , α_{mkt} and γ are given by the regulator, the market share, m can be perfectly calculated and β is chosen by the company.

6.2.2. SST Model for Individual Claims

For individual large events, the SST provides p_1 , the probability of observing losses larger than CHF 1 million and standard values for α_β , for $\beta = 1$ and $\beta = 5$ (see Table FINMA (2016)). Since the probability of large claims provided by the SST is based on a lower threshold of CHF 1 million, a thinning process of the CP-P has to be done if the company decides to use $\beta = 5$.

Following the same procedure presented in Section 6.2.1 we can see that the company's large individual claims are modelled as

$$Z_{comp} \sim \text{CP-P}(\lambda_\beta, \beta, \alpha_\beta),$$

with an expected number of claims larger than β equal to

$$\lambda_\beta = \lambda_{CY,l} = p_1 \lambda_{CY} \left(\frac{\beta}{1} \right)^{-\alpha_\beta}, \quad (51)$$

where λ_{CY} denotes the expected total number of CY claims in the ℓ -th LoB. Similarly, the regulator also requires an upper bound in the Pareto random variables, leading to the following distribution of large losses

$$\tilde{Z} \sim \text{CP-P}(\lambda_\beta, \beta, \alpha_{mkt}, \gamma).$$

As in Section 6.2.1, the distribution of $Z_{CY,l} | \mathcal{F}(t)$ is approximated by a single Pareto, with the same mean and Pareto index α_β .

Model Assumptions 6 (Marginal distribution of large individual claims). For α_β , p_1 and γ provided by the regulator in FINMA (2016), $\beta \in \{1, 5\}$ and $\lambda_{CY} > 0$,

$$Z_{CY,l} | \mathcal{F}(t) \sim \text{Pareto} \left(\lambda_\beta \frac{\beta^{\alpha_{mkt}}}{1 - (\beta/\gamma)^{\alpha_{mkt}}} \left(\frac{1}{\beta^{\alpha_{mkt}-1}} - \frac{1}{\gamma^{\alpha_{mkt}-1}} \right), \alpha_{mkt} \right),$$

with λ_β defined in Equation (51).

7. Joint Distribution of PY and CY Claims

Although the SST does not assume any parametric form for the joint distribution of $\mathbf{Z} | \mathcal{F}(t)$ or $\bar{\mathbf{Z}} | \mathcal{F}(t)$ (defined in Equations (30) and (31), respectively) it is required that a pre-specified correlation matrix Λ is used (see FINMA (2016)). In this section we discuss how to use the conditional and marginalized models to define a joint distribution satisfying this correlation assumption.

It is important to notice, though, that the SST correlation matrix may not be attainable for some joint distributions, as discussed in Appendix B in the case of log-normal marginals (in Devroye and Letac (2015) the authors discuss a similar problem). Let us denote by \mathcal{S}_n the set of all $n \times n$, symmetric, positive semi-definite matrices with diagonal terms equal to 1; and by $S(\mathbf{C}) = \text{Corr}(\mathbf{U})$ the correlation matrix of a random vector $\mathbf{U} \sim C$, with elements $U_i \sim [0, 1]$. The question asked in Devroye and Letac (2015) is: given $S \in \mathcal{S}_n$, does there exist a copula C such that $S(\mathbf{C}) = S$? The answer is yes, if $n \leq 9$ and the authors postulate that for $n \geq 10$ there exists $S \in \mathcal{S}_n$ such that there is no copula C such that $S(\mathbf{C}) = S$.

It should be noted that, since in the SST the CY large claims are assumed to be independent from all the other risks, the correlation matrix of $(\mathbf{Z}_{PY}, \mathbf{Z}_{CY,s}, \mathbf{Z}_{CY,l}) | \mathcal{F}(t)$ is essentially a correlation matrix between $(\mathbf{Z}_{PY}, \mathbf{Z}_{CY,s}) | \mathcal{F}(t)$ and the same is true also for the conditional model.

Regardless of assuming a conditional or a marginalized model, SST's correlation matrix Λ should be such that, for $i, j = 1, \dots, 2L + P$ (recall that L are the number of LoBs and P the number of perils),

$$\Lambda_{i,j} = \text{Corr}(Z_i, Z_j | \mathcal{F}(t)) = \text{Corr}(\bar{Z}_i, \bar{Z}_j | \mathcal{F}(t)).$$

Remark 14. In the conditional model we need to “integrate out” the parameter uncertainty, otherwise the (conditional) correlation would be dependent on an unknown parameter and could not be matched with the numbers provided by the SST.

7.1. Conditional Joint Model

Under Model Assumptions 2, 4, 5 and 6 our interest lies on modelling the joint behaviour of the vector $\bar{Z} | \sigma, \mathcal{F}(t)$. Under Model Assumptions 1 it can be shown that the required conditional independence between $Z_{CY,l}$ and $(Z_{PY}, Z_{CY,s})$ given $\mathcal{F}(t)$ is equivalent to the conditional independence between $Z_{CY,l}$ and $(Z_{PY}, Z_{CY,s})$ given $\mathcal{F}(t)$ and σ .

Moreover, since all the marginal conditional distributions of the prior year claims and small current year claims are assumed to be log-normal, following Equations (30) and (31), the notation can be further simplified to

$$\bar{Z}_i | \sigma, \mathcal{F}(t) \sim \text{LN}(\bar{m}_i(\sigma), \bar{V}_i(\sigma)), \text{ for } i = 1, \dots, 2L, \quad (52)$$

with $\bar{m}_i(\sigma)$, and $\bar{V}_i(\sigma)$ defined in Model Assumptions 2 and 4. For example, for $i = L + 1$, $\bar{m}_i(\sigma) = \mu_{CY,s}^{(1)}$, defined in Model Assumptions 4.

We are now ready to define the joint conditional model to be used.

Model Assumptions 7 (Conditional joint model). Based on Model Assumptions 2 and 4 we link the marginals of the conditional model through a Gaussian copula with correlation matrix $\bar{\Omega}$, with elements $(\bar{\Omega})_{i,j} = \bar{\omega}_{i,j}$. More formally, given $\mathcal{F}(t)$ and σ , the joint distribution of \bar{Z} is given by

$$F_{\bar{Z}}(\bar{z}_1, \dots, \bar{z}_{2L}; \bar{\Omega} | \mathcal{F}(t), \sigma) = C(F_{\bar{Z}_1}(\bar{z}_1 | \mathcal{F}(t), \sigma), \dots, F_{\bar{Z}_{2L}}(\bar{z}_{2L} | \mathcal{F}(t), \sigma); \bar{\Omega}),$$

where $F_{\bar{Z}_i}(\cdot | \mathcal{F}(t), \sigma)$ denotes the conditional distribution of $\bar{Z}_i | \mathcal{F}(t), \sigma$ defined in Equation (52) and $C(\cdot; \bar{\Omega})$ is the Gaussian copula with correlation matrix denoted by $\bar{\Omega}$.

Remark 15. In this section the parameter matrix $\bar{\Omega}$ should be understood as a deterministic variable, differently from σ and ϕ . For this reason we do not include it on the right hand side of the conditioning bar. Instead, whenever $\bar{\Omega}$ needs to be explicitly written, we include it on the left hand side of the bar, separated by the function (or functional, for expectations) arguments by a semicolon.

In order to match SST’s correlation matrix Λ , under Model Assumptions 1 and 4, the following equation needs to be solved with respect to $\bar{\Omega}$:

$$\Lambda_{i,j} = \text{Corr}(\bar{Z}_i, \bar{Z}_j; \bar{\Omega} | \mathcal{F}(t)). \quad (53)$$

To compute the right hand side of the equation above we first notice that

$$\text{Cov}(\bar{Z}_i, \bar{Z}_j; \bar{\Omega} | \mathcal{F}(t)) = \mathbb{E}[\bar{Z}_i \bar{Z}_j; \bar{\Omega} | \mathcal{F}(t)] - \mathbb{E}[\bar{Z}_i | \mathcal{F}(t)] \mathbb{E}[\bar{Z}_j | \mathcal{F}(t)],$$

where, from Equation (46) and the discussion in Section 6.1,

$$\begin{aligned} \mathbb{E}[\bar{Z}_i | \mathcal{F}(t)] &= \mathbb{E}[\mathbb{E}[\bar{Z}_i | \mathcal{F}(t), \sigma] | \mathcal{F}(t)] \\ &= \mathbb{E}[\bar{m}_i | \mathcal{F}(t)] = \begin{cases} \bar{\mathcal{R}}^{(i)}(t), & \text{if } 1 \leq i \leq L, \\ \mathbb{E}[Z_{CY,s}^{(i-L)} | \mathcal{F}(t)], & \text{if } L+1 \leq i \leq 2L, \end{cases} \end{aligned}$$

and from Equation (A3), Appendix B,

$$\begin{aligned} \mathbb{E}[\bar{Z}_i \bar{Z}_j; \bar{\Omega} | \mathcal{F}(t)] &= \mathbb{E}[\mathbb{E}[\bar{Z}_i \bar{Z}_j; \bar{\Omega} | \mathcal{F}(t), \sigma] | \mathcal{F}(t)] \\ &= \mathbb{E} \left[\exp \left\{ \bar{m}_i + \frac{\bar{V}_i^2 + 2\bar{V}_i \bar{\omega}_{i,j} \bar{V}_j + \bar{V}_j^2}{2} + \bar{m}_j \right\} \middle| \mathcal{F}(t) \right]. \end{aligned}$$

Therefore, to satisfy Equation (53) $\bar{\Omega}_{i,j}$ needs to be chosen such that the following implicit relationship (which can be solved through any univariate root search algorithm) holds:

$$\Lambda_{i,j} \sqrt{\text{Var}(\bar{Z}_i | \mathcal{F}(t)) \text{Var}(\bar{Z}_j | \mathcal{F}(t))} + \mathbb{E}[\bar{Z}_i | \mathcal{F}(t)] \mathbb{E}[\bar{Z}_j | \mathcal{F}(t)] - \mathbb{E}[\bar{Z}_i \bar{Z}_j; \bar{\Omega} | \mathcal{F}(t)] = 0.$$

7.2. Marginalized Joint Model

Similarly to Section 7.1, in this section we will fully characterize the joint distribution of $\mathbf{Z} | \mathcal{F}(t)$ under Model Assumptions 3, 4, 5 and 6.

From these assumptions we define the following notation:

$$Z_i | \mathcal{F}(t) \sim \text{LN}(m_i, V_i), \text{ for } i = 1, \dots, 2L. \quad (54)$$

Model Assumptions 8 (Marginalized joint model). *Based on Model Assumptions 3 and 4 we link the marginal distributions of the marginalized model through a Gaussian copula with correlation matrix Ω , with elements $(\Omega)_{i,j} = \omega_{i,j}$. More formally, given $\mathcal{F}(t)$, the joint distribution of \mathbf{Z} is given by*

$$F_{\mathbf{Z}}(z_1, \dots, z_{2L}; \Omega | \mathcal{F}(t)) = C\left(F_{Z_1}(z_1 | \mathcal{F}(t)), \dots, F_{Z_{2L}}(z_{2L} | \mathcal{F}(t)); \Omega\right),$$

where F_{Z_i} denotes the conditional distribution of $Z_i | \mathcal{F}(t)$ defined in (Equation 54) and $C(\cdot; \Omega)$ is the Gaussian copula with correlation matrix Ω .

In order to match SST's correlation matrix, in the joint marginalized model the Gaussian copula correlation Ω is chosen such that (see Equation (A4), Appendix B) it satisfies

$$\Lambda_{i,j} = \frac{\exp\{V_i \omega_{i,j} V_j\} - 1}{\left[(e^{V_i^2} - 1)(e^{V_j^2} - 1)\right]^{1/2}}.$$

8. Data Description and Parameter Estimation

In this section we discuss how we set up the parameters in the models discussed so far, starting from the balance sheet of a fictitious insurance company. Using this balance sheet and the information contained in the SST we generate realistic claims triangles (see Appendix C) and, based on them, we show how to perform Bayesian inference for the unknown parameters. Our starting point is the fictitious balance sheet shown in Table 1, which is intended to represent a large insurance company in Switzerland (for this reason all monetary units should be understood as millions of Swiss Francs (CHF)).

8.1. Hyperparameters for ϕ_j

Based on SST's standard runoff pattern (see Table 3) we first compute the implied CL factors $f_j^{(\ell)}$ as follows (once again we suppress the index ℓ of the LoB). If F_j is the deterministic cumulative claims payment pattern for development year j we define

$$f_j = \frac{F_{j+1}}{F_j}, \text{ for } j = 0, \dots, J-1.$$

These values can, then, be used as a hyperparameter in the prior for ϕ_j (see Model Assumptions 1, item (c)).

To generate data from the model (see Section C) we fix $\phi_j = 1/f_j$ and $\sigma_j = s_j/f_j$, where s_j is Mack's standard deviation estimate calculated from exogenous triangles. The values of s_j are presented in Table 4. That is, $\{F_j\}_j$ should be understood as a (deterministic) prior payment pattern.

Table 3. Swiss Solvency Test (SST)'s (2015) standard development patterns for claims provision (normalized to have at most 30 development years and rounded to 2 digits).

[illegible]

Table 4. Mack's standard deviation parameter estimates, s_j , based on exogenous triangles and for the development lengths given in Table 3.

[illegible]

8.2. Current Year Small and Large Claims

To calculate the expected number of CY claims, λ_{CY} , defined in Section 6, we first set our prior belief for the *claims ratio* for each LoB, i.e., how much of the premium in that LoB is used to cover incoming claims (all the rest covers business' costs). This information is available in Table 5, along with the *average claim amount*. Based on these values the expected number of claims is defined as

$$\lambda_{CY} = \frac{\text{Claims ratio} \times \text{Premium}}{\text{Average claim amount}}.$$

Table 5. Claims ratio, average claim amount (in millions of CHF) and market share.

LoB	Claims Ratio	Average Claim Amount	Market Share
1	90%	0.005	20%
2	75%	0.003	
3	75%	0.004	
4	75%	0.004	
5	90%	0.004	10%
6	90%	0.003	
7	90%	0.002	
8	80%	0.003	
9	80%	0.003	

Given the expected number of CY claims, λ_{CY} , this value is used to compute the expected number of individual large claims, $\lambda_{CY,l}$, as in Equation (51). Using the fact that $\lambda_{CY,s} = \lambda_{CY} - \lambda_{CY,l}$ we calculate the coefficient of variation for small CY claims as given in Equation (48).

The last ingredient in Model Assumptions 4 is $\mathbb{E}[Z_{CY,s} | \mathcal{F}(t)]$ which is given by

$$\mathbb{E}[Z_{CY,s} | \mathcal{F}(t)] = \text{Claims ratio} \times \text{Premium} - \mathbb{E}[Z_{CY,l} | \mathcal{F}(t)],$$

and the expectation on the right hand side is given either in Model Assumptions 5 or Model Assumptions 6, depending on the LoB.

For the large claims from Model Assumptions 5 and 6 we assume the threshold for large claims β to be equal to 5 (millions of CHF). For the large cumulated claims we use LoBs market share as given in Table 5. The resulting parameters can be found in Table 2. Note these parameters are the same both for the marginalized and conditional models.

8.3. Parameter Estimation

In this section we discuss how to compute the posterior distributions of the variance parameters σ_j in Model Assumptions 1, which are used to compute quantities such as the marginalized msef from Section 5.2.

In order to compute the posteriors of σ_j , we assume priors centred at Mack's Mack (1993) CL standard deviation estimator normalized by the CL factor f , both implied by the data. Formally,

$$\hat{\sigma}_j(t) = \frac{\sqrt{\hat{s}_j^2(t)}}{\hat{f}_j(t)}, \quad \forall 0 \leq j \leq J-1, \quad (55)$$

where $\hat{s}_{j-1}^2(t) = \min\{\hat{s}_{j-3}^2(t), \hat{s}_{j-2}^2(t), \hat{s}_{j-2}^4(t)/\hat{s}_{j-3}^2(t)\} = \min\{\hat{s}_{j-3}^2(t), \hat{s}_{j-2}^4(t)/\hat{s}_{j-3}^2(t)\}$.

To generate samples from the posteriors we use a Metropolis-Hastings algorithm, with proposals given by a truncated Normal centred at the current point and standard deviation equal to $10 \times d_j$. All the chains are started at the CL variance estimate and the upper limit for the prior, $d_j = k \times \hat{\sigma}_j(t)$ is set as $k = 5$ times the CL variance estimate. To be left with $N_{MCMC} = 1000$ samples from the posterior we

ran the Markov chains for 12,500 iterations, discarding the first 20% as a burn-in and keeping every 10th iteration of the remaining simulations.

Some of the results are presented in Figure 3 where one finds the unnormalized posteriors, the histogram of the MCMC outputs and a red dashed line indicating the CL variance estimate for three different LoBs: (a) MTPL, (b) Motor Hull and (c) Property. As expected, for unidimensional and unimodal densities the resulting estimates are highly accurate. It is also worth noticing that the larger the development year j the more diffuse the posterior is, due to the diminishing amount of data available. In the limit, when $j = J - 1$ the information available is not enough to estimate the variance parameter and, therefore, as can be seen from the posterior distribution derived in Equation (38), the posterior is the same as the prior.

Using the sample of size $N_{MCMC} = 1000$ mentioned above, the calculated parameters for the marginalized model are presented in Table 2. For the conditional model we use the same sample from the posterior and calculate the one value of σ_{PY} and μ_{PY} for each sampled value σ . The resulting (transformed) samples are presented as histograms in Figures 5 and 6 and, for comparison only, the relevant marginalized parameters are included as a red dashed line.

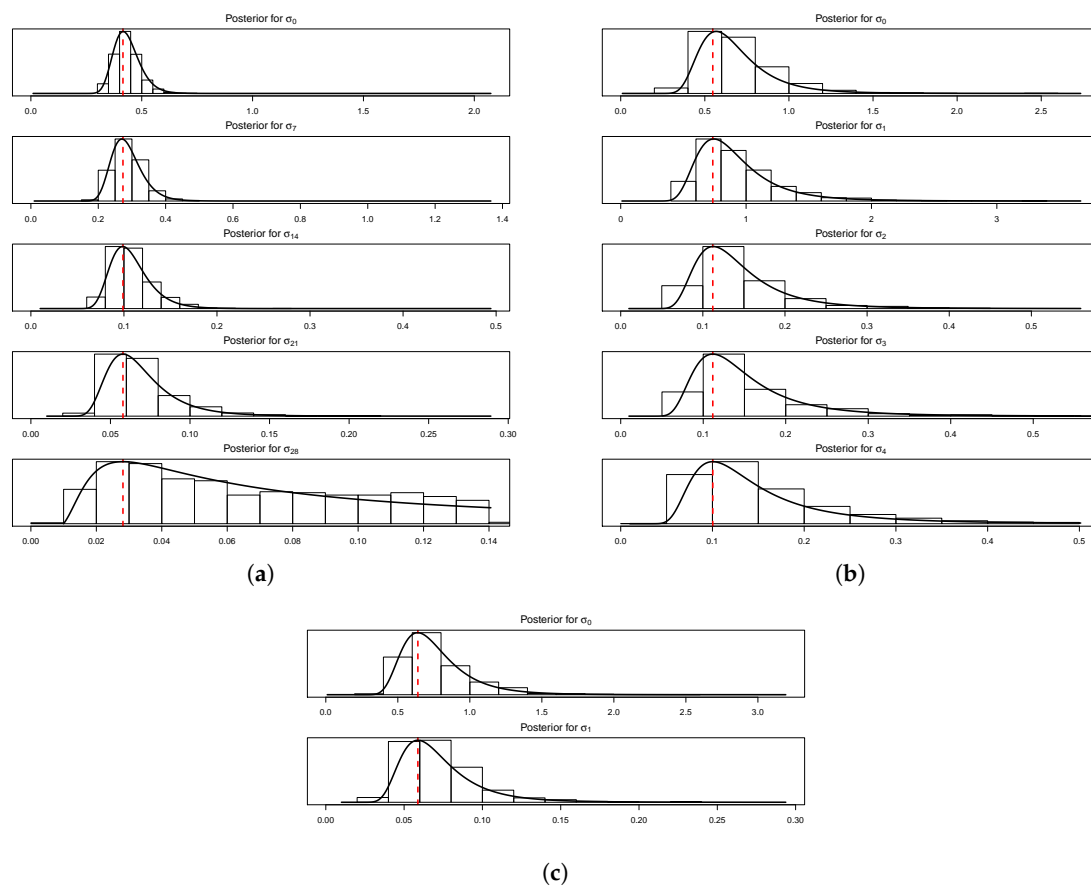


Figure 3. Posterior distributions for σ_j for the (a) Motor Third Part Liability (MTPL) (b) Property and (c) Motor Hull lines of business. One sees solid lines representing the unnormalized posteriors, the histogram of the Markov Chain Monte Carlo (MCMC) outputs and a red dashed line indicating the CL standard deviation estimate. Note that for LoB MTPL we only plot selected development periods: $j \in \{0, 7, 14, 21, 28\}$.

8.4. The Correlation Matrices

For the copula correlation matrices we follow the procedures outlined in Sections 7.1 and 7.2. The resulting matrix for the marginalized model is found in Table 6. From FINMA (2016) it can be seen

9. Details of the SMC Algorithm

9.1. Selection of Intermediate Sets

Recall that a key component of the proposed SMC Sampler solution is to create a relaxation of the rare-conditional events that constrain the target posterior into a sequence of increasingly difficult constraints. In this section we discuss how one can select the sequence of constraint relaxations in an adaptive manner.

For both the marginalized and the conditional models we use an adaptive strategy similar to Cérou et al. (2012) in order to select adaptively online (as the algorithm runs) the levels B_1, \dots, B_T , as well as the total number of intermediate sets T . When levels are being chosen adaptively one of the main advantages of the proposed SMC algorithm is the ability to estimate, in one run, the company-wide value at risk, the expected shortfall as well as the risk allocations.

Starting from $B_0 = 0$ (or $\bar{B}_0 = 0$ if the conditional model is being used) the idea consists of, at each algorithmic iteration $t - 1$, choosing the next level, B_t , such that a percentage $p_0 \in (0, 1)$ of the $(t - 1)$ -particles is above this set. More formally, we set B_t to be the $1 - p_0$ empirical quantile of the weighted sample $\{s_{t-1}^{(j)}, W_{t-1}^{(j)}\}_{j=1}^N$ or $\{\bar{s}_{t-1}^{(j)}, W_{t-1}^{(j)}\}_{j=1}^N$, where s_{t-1} and \bar{s}_{t-1} denote, respectively, the sum of the components of \mathbf{z}_t and $\bar{\mathbf{z}}_t$. Therefore, at algorithmic time t the level B_t corresponds to an estimate of the $(1 - p_0)$ -th quantile of the target distribution. In our examples we set $p_0 = 0.4, 0.5$ and 0.7 which induces intermediate quantiles seen in Table 10 for the algorithm. Note that, given a value of p_0 the number of levels in the algorithm is deterministic. For example, for $p_0 = 0.5$ there are 7 levels until the estimated quantile is above 99%.

Table 10. Intermediate quantiles for different values of p_0 .

p_0	1	2	3	4	5	6	7
0.4	0.6	0.84	0.936	0.9744	0.9898	0.9959	
0.5	0.5	0.75	0.875	0.9375	0.9688	0.9844	0.9922
0.7	0.3	0.51	0.657	0.7599	0.8319	0.8824	0.9176
p_0	8	9	10	11	12	13	
0.4							
0.5							
0.7	0.9424	0.9596	0.9718	0.9802	0.9862	0.9903	

An alternative approach to choosing the level sets is to use the classic normalizing constant estimator derived from the SMC sampler algorithm (see (Del Moral et al. 2006, sct. 3.2.1)). Using the notation from Section 3 we have that the normalizing constant $Z_t = \mathbb{P}[S > B_t]$ can be estimated as

$$\hat{Z}_t = \hat{Z}_{t-1} \sum_{j=1}^N W_{t-1}^{(j)} \tilde{\alpha}_t^{(j)}, \quad (56)$$

where W_{t-1} and $\tilde{\alpha}_t$ are, respectively the normalized and the incremental weights at time $t - 1$.

Similarly to our proposed estimate, in this alternative route one would choose B_t such that $p_0 \times 100\%$ of the time $t - 1$ particles are above this level. Using the estimator in Equation (56) one could stop the algorithm as soon as $\hat{Z}_t < \alpha$. The main disadvantage of this approach is that although \hat{Z}_t can be proven to be unbiased and asymptotically normally distributed when the number of particles $N \rightarrow \infty$ (see (Del Moral 2004, Propositions 7.4.1 and 9.4.1) and Pitt et al. (2012) for a proof in the special case of state-space models) one can not guarantee $\hat{Z}_t \in [0, 1]$. In our experiments the results based on this classic estimate were deemed unsatisfactory, as we observed estimates of the normalizing constant as large as 15, as finite sample realizations.

9.2. Marginalized Model

9.2.1. The Forward Kernel

Similarly to (Targino et al. 2015, sct. 6.1) we propose a mutation kernel $K_t(z_{t-1}, z_t)$ such that the condition $\sum_{i=1}^d z_{i,t} > B_t$ is always satisfied. Due to the independence assumption of the CY large claims (the P Pareto variables) we first independently mutate the Pareto coordinates, following their true (unconditional) marginal and then mutate the other $2L$ variables.

First we split the vector into its log-normal and Pareto components, $z_t = (z'_t, z''_t)$, where $z'_t = (z_{t,1}, \dots, z_{t,2L})$ and $z''_t = (z_{t,2L+1}, \dots, z_{t,2L+P})$. Using this notation and denoting $z_{t,-m}$ the vector z_t without its m -th component, we use

$$\begin{aligned} K_t(z_{t-1}, z_t) &= K'_t(z'_{t-1}, z'_t | z''_t) K''_t(z''_{t-1}, z''_t) \\ &= \left\{ \frac{1}{2L} \sum_{m=1}^{2L} \left[K_t^{(-m)}(z'_{t-1}, z'_{t,-m}) K_t^{(m)}(z'_{t-1}, z'_{t,m} | z'_{t,-m}, z''_t) \right] \right\} \\ &\quad \times \prod_{i=2L+1}^{2L+P} \text{Pareto}(z''_i; \alpha_i, \beta_i), \end{aligned}$$

where the kernel $K_t^{(-m)}(z'_{t-1}, \cdot)$, which mutates all but the m -th dimension of z'_{t-1} , consists of independent moves in each dimension, i.e.,

$$K_t^{(-m)}(z'_{t-1}, z'_{t,-m}) = \prod_{\substack{i=1 \\ i \neq m}}^{2L} K_t^{(-m,i)}(z'_{t-1,i}, z'_{t,i}).$$

Note that these moves are also independent of the P Pareto mutations.

Let us denote $\{z_{t-1}^{(j)}, W_{t-1}^{(j)}\}_{j=1}^N$ the weighted sample approximating

$$\pi_t(z_{t-1}) = f_Z(z_{t-1} | z_{t-1} \in \mathcal{G}_{Z_{t-1}}),$$

as defined in Equation (15). The components of the mutation kernel are then defined as

$$K_t^{(-m,i)}(z'_{t-1}, z'_{t,i}) = \text{LN}(z'_{t,i}; \hat{\mu}_i, \hat{\sigma}_i), \quad \text{for } i = 1, \dots, 2L, i \neq m, \quad (57)$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i$ are the empirical mean and variance of $\{z_{t-1}^{(j)}, W_{t-1}^{(j)}\}_{j=1}^N$ when $i = 1, \dots, 2L$

$$\begin{aligned} \hat{\mu}_{t-1,i} &= \sum_{j=1}^N W_{t-1}^{(j)} z_{t-1,i}^{(j)}, \\ \hat{\sigma}_{t-1,i}^2 &= \hat{\mu}_{t-1,i}^2 - \sum_{j=1}^N W_{t-1}^{(j)} \left(z_{t-1,i}^{(j)} \right)^2. \end{aligned}$$

For the mutation of the remaining dimension, m , to ensure all the samples satisfy the condition $\sum_{i=1}^d z_{i,t} > B_t$ we proceed as follows. First we define

$$B_t^z(m) = \max \left\{ 0, B_t - \sum_{\substack{i=1 \\ i \neq m}}^d z_{t,i} \right\},$$

and then sample the last component $z_{m,t} \in [B_t^z(m), +\infty)$ according to

$$K_t^{(m)}(z_{t-1}, z_{t,m} | z_{t,-m}) = \text{TN}(z_{t,m}; \hat{\mu}_m, \hat{\sigma}_m, B_t^z(m), +\infty), \quad \text{for } m = 1, \dots, 2L, \quad (58)$$

where $TN(\cdot; \mu, \sigma, a, b)$ denotes the density of a Normal distribution with mean μ and variance σ^2 truncated on support $[a, b]$.

9.2.2. The Backward Kernel

For the backward kernel we follow the discussion in Section 3.1.2 and use the (approximation to the) optimum kernel of Del Moral et al. (2006), given by equation Equation (11)

$$L_t(z_{t+1}, z_t) = \frac{\gamma_t(z_t) K_{t+1}(z_t, z_{t+1})}{\frac{1}{N} \sum_{j=1}^N w_t^{(j)} K_{t+1}(z_t^{(j)}, z_{t+1})},$$

where $w_t^{(j)}$ denotes the *unnormalized* weights at time t and the weighted sample $\{z_t^{(j)}, w_t^{(j)}\}_{j=1}^N$ targets the unnormalized density $\gamma_t(z_t)$. Proceeding in this way the unnormalized weights for the SMC sampler algorithm (see Algorithm 1) satisfy the following recursion

$$w_t^{(j)} = w_{t-1}^{(j)} \frac{\gamma_t(z_t)}{\frac{1}{N} \sum_{k=1}^N w_{t-1}^{(k)} K_t(z_{t-1}, z_t)}.$$

9.2.3. The MCMC Move Kernel

To improve particle diversity after a resampling step (which is performed whenever the effective sample size drops below $N/2$) the following MCMC move kernel is applied to the particles.

As in (Targino et al. 2015, sct. 6.2) we propose a Gibbs-type update combined with a slice sampler (see Neal (2003)). For notational simplicity we suppress the dependence in t in the vector z_t and denote $v^*(m) = (z_1^*, \dots, z_m^*, z_{m+1}, \dots, z_d)$ the vector where the first m components have already been updated in the Gibbs scan. The full conditional for the m -th component of z_t is given by

$$\pi_t(z_m^* | z_1^*, \dots, z_{m-1}^*, z_{m+1}, \dots, z_d) \propto \pi_t(v^*(m)) \propto f_Z(v^*(m)) \mathbb{1}_{\mathcal{G}_{Z_t}}(v^*(m)),$$

which can be sampled from using an unidimensional slice sampler (see Neal (2003)).

9.3. Conditional Model

Following the discussion in Section 3.3.2 we use equation Equation (19) as an approximation to the unknown density $f_{\bar{Z}}(\bar{z})$. For our simulations $M = 5$ samples of the unknown parameter θ are used, where

$$\theta = (\sigma^{(1)}, \dots, \sigma^{(L)}),$$

and each vector $\sigma^{(\ell)} = (\sigma_1^{(\ell)}, \dots, \sigma_J^{(\ell)})$ contains all the unknown variance parameters for the ℓ -th LoB. Therefore, $\theta = (\theta^{(1)}, \dots, \theta^{(M)})$ and it should be noticed the superscript have a different interpretation from those in $\sigma_j^{(\ell)}$.

As the parameter estimation step described in Section 8.3 is independent of the allocation process we assume N_{MCMC} samples for each unknown parameter vector σ have already been created. Therefore, to sample $\bar{z} \sim f_{\bar{Z}}(\bar{z})$ we first sample an index $n \sim U(\{1, \dots, N_{MCMC}\})$ and then $\bar{z} \sim f_{\bar{Z}}(\bar{z} | \theta^{(n)})$.

9.3.1. The Forward Kernel

The forward kernel used for the conditional model follows the same structure as the one used in the marginalized model and described in Section 9.2.1: first we sample the P independent Pareto variables (with the same distribution as in the marginalized case) and then the remaining $2L$ variables. More precisely,

$$\bar{K}_t^{(-m,i)}(\bar{z}'_{t-1}, \bar{z}'_{t,i} | \theta_t) = \bar{K}_t^{(-m,i)}(\bar{z}'_{t-1}, \bar{z}'_{t,i}) = K_t^{(-m,i)}(\bar{z}'_{t-1}, \bar{z}'_{t,i}),$$

where the last term is defined in equation Equation (57) and $\hat{\mu}_i$ and $\hat{\sigma}_i$ are now the empirical mean and variance of $\{\bar{z}_{t-1}^{(j)}, W_{t-1}^{(j)}\}_{j=1}^N$. Likewise,

$$\bar{K}_t^{(m)}(\bar{z}_{t-1}', \bar{z}_{t,m}' | \bar{z}_{t,-m}', \boldsymbol{\theta}_t) = \bar{K}_t^{(m)}(\bar{z}_{t-1}', \bar{z}_{t,m}' | \bar{z}_{t,-m}') = K_t^{(m)}(\bar{z}_{t-1}', \bar{z}_{t,m}' | \bar{z}_{t,-m}'),$$

with the last term defined in equation Equation (58). As samples from $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ have already been generated through MCMC then the mutation kernel in the extended space, $K_t^y(\mathbf{y}_{t-1}, \mathbf{y}_t)$, is completely characterized.

9.3.2. The Backward Kernel

As in Section 9.2.2 we use the optimum backward kernel in the extended space $\mathcal{Y} = \mathbb{R}^d \times \Theta^M$, which for the conditional model leads to the following incremental weights (see Equation (12))

$$\begin{aligned} \alpha_t &= \frac{\gamma_t^y(\mathbf{y}_t)}{\frac{1}{N} \sum_{j=1}^N w_{t-1}^{(j)} K_t^y(\mathbf{y}_{t-1}, \mathbf{y}_t)} \\ &= \frac{\hat{f}_{\bar{\mathbf{Z}}}(\bar{\mathbf{z}}_t; \boldsymbol{\theta}_t) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_t) \mathbb{1}_{\mathcal{G}_{\bar{\mathbf{Z}}}}(\bar{\mathbf{z}}_t)}{\frac{1}{N} \sum_{j=1}^N w_{t-1}^{(j)} K_t(\mathbf{z}_{t-1}, \mathbf{z}_t) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_t)} \\ &= \frac{\hat{f}_{\bar{\mathbf{Z}}}(\bar{\mathbf{z}}_t; \boldsymbol{\theta}_t) \mathbb{1}_{\mathcal{G}_{\bar{\mathbf{Z}}}}(\bar{\mathbf{z}}_t)}{\frac{1}{N} \sum_{j=1}^N w_{t-1}^{(j)} K_t(\mathbf{z}_{t-1}, \mathbf{z}_t)}. \end{aligned}$$

9.3.3. The MCMC Move Kernel

The MCMC move kernel used for the conditional model needs to keep the target distribution in the extended space, $\pi_t^y(\mathbf{y}_t)$, invariant. The strategy adopted is to first sample $\boldsymbol{\theta}^* \sim f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ and then $\mathbf{z}_t | \boldsymbol{\theta}^* \sim \hat{f}_{\bar{\mathbf{Z}}}(\bar{\mathbf{z}}_t; \boldsymbol{\theta}_t^*) \mathbb{1}_{\mathcal{G}_{\bar{\mathbf{Z}}}}(\bar{\mathbf{z}}_t)$.

For the second step above we use exactly the same Gibbs-sampler update as in Section 9.2.3, with $f_{\mathbf{Z}}(\cdot)$ replaced by $\hat{f}_{\bar{\mathbf{Z}}}(\cdot; \boldsymbol{\theta}_t)$.

10. Results

In this section we present the results of the SMC procedure when used to calculate the expected shortfall allocations from Equations (4) and (5) of the solvency capital requirement.

Before proceeding to the results calculated via the SMC algorithm, in order to understand the simulated data presented in Figure 4, in Table 2 we present some results based on a “brute force” Monte Carlo (rejection-sampling) simulation, which is taken as the base line for comparisons with the SMC algorithm. The table is divided in three blocks of rows, with PY claims, CY small (CY,s) claims and CY large (CY,l) claims.

First of all, it should be noticed that the reserves presented on the first block of Table 2 are the ones implied by the data, which we then assume to be the true ones (ignoring, from now on, the initial synthetic data from Table 1). That is, based on initial parameters we have generated synthetic claims development triangles, which naturally deviate slightly from their expected values. The parameters σ and μ for PY claims are related to the marginalized model (for the parameters of the conditional model see Figures 5 and 6). It is also important to note that only the PY parameters are different between the conditional and marginalized models.

For each LoB the standalone expected shortfall (ES) is calculated analytically and its value is, then, combined with the LoB’s expectation to calculate the solvency capital requirement (SCR). These values are added up, both within risk type (i.e., PY, CY,s and CY,l) and globally, in order to calculate the overall standalone capital. For the marginalized and conditional models the columns “ES” and “SCR” denote, respectively, the expected shortfall and capital allocations to each LoB. These values are

compared to their standalone counterparts to generate the diversification benefit, which is around 45% for PY and CY,s claims (regardless of the model used) and ranges between 30% and 70% within the PY and CY,s groups. Due to the independence assumptions the largest diversification benefit comes from the CY,l claims, where the capital is reduced by around 95%.

The data presented in Table 2 is calculated as follows. For the marginalized model (and conditional model in brackets), 5×10^9 (2.5×10^7) independent samples of the model are generated in order to calculate the overall $\text{VaR}_{99\%}$. Conditional on this value, for each LoB we then generate 5×10^7 (5×10^5) samples above the VaR and use the average of these samples as the true ES allocation (presented in Table 2). In order to assess the variance of the estimators, we divide these samples into $N_{rep} = 500$ groups of $N_{MC} = 10^5$ ($N_{MC} = 10^3$ for the conditional model) simulations. More formally, we approximate the ES allocations ρ_i , defined in Equation (4), by

$$\widehat{\mathbb{E}}[\widehat{\rho}_{i,MC}] = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \widehat{\rho}_{i,MC}^{(k)}, \quad (59)$$

where $\widehat{\rho}_{i,MC}^{(k)}$ stands for the estimate (using N_{MC} particles) from the k -th run (out of N_{rep}), which is defined according to

$$\widehat{\rho}_{i,MC} = \frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} Z_i^{(j)}.$$

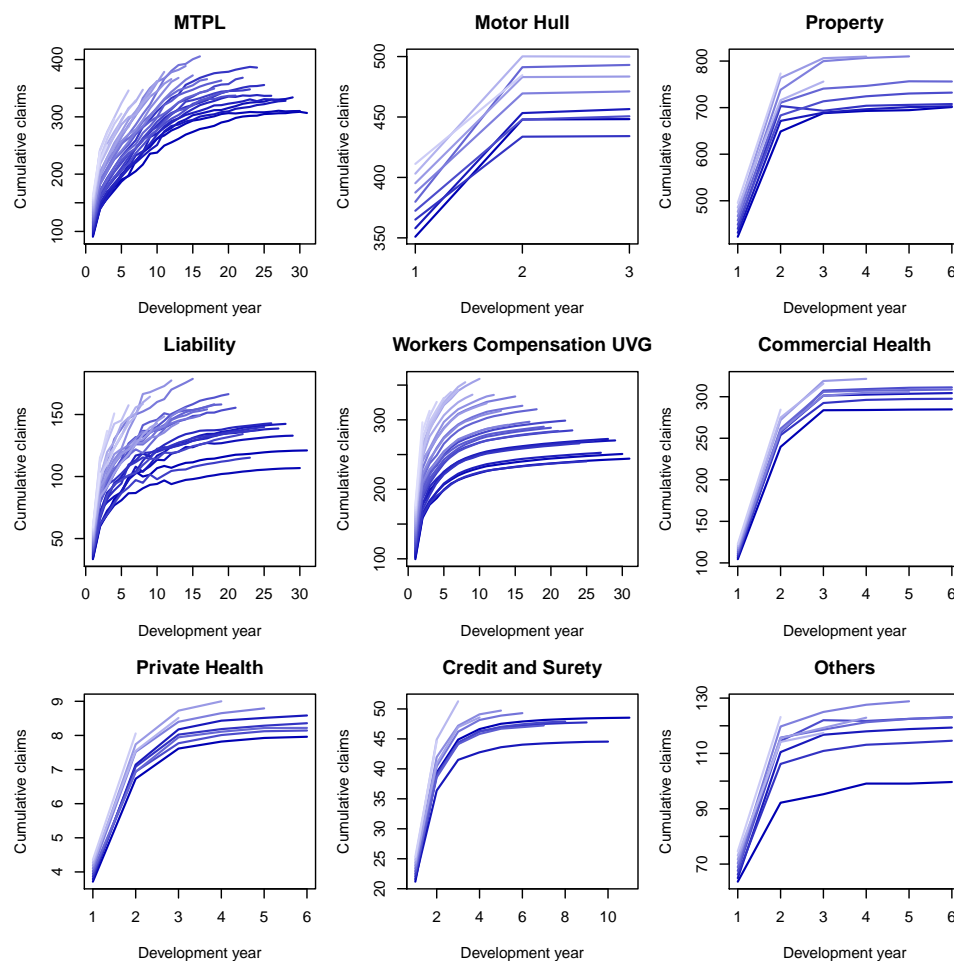


Figure 4. Cumulative claims payment (in millions of CHF). Lighter colours represent more recent accident years.

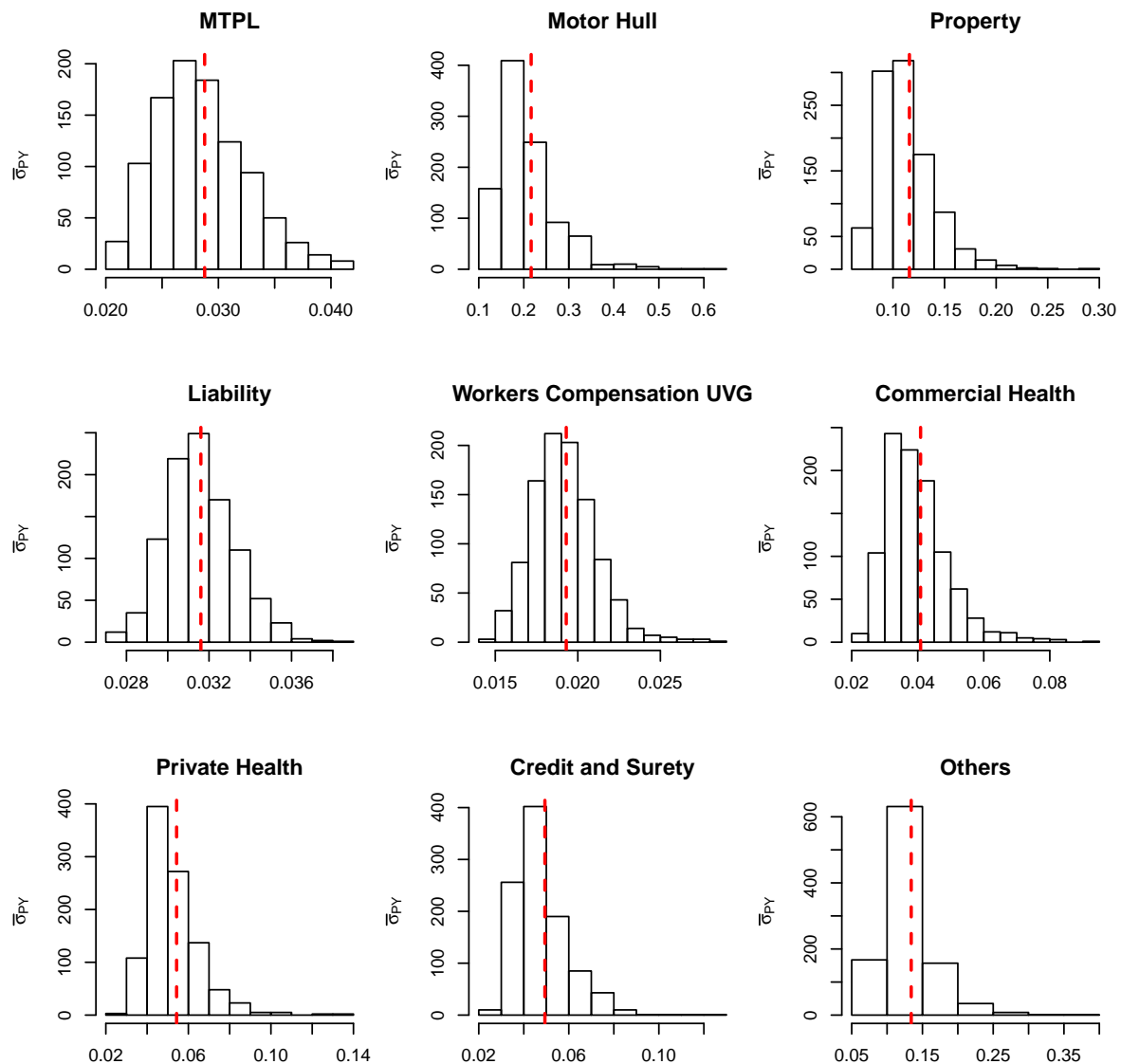


Figure 5. Histogram of the parameter $\bar{\sigma}_{PY}$ for the conditional model. Red dashed line: σ_{PY} .

Similarly to the analysis performed in [Peters et al. \(2017\)](#) the impact of the prior density can be assessed by comparing the sum of the SCR allocations with the SCR from the “empirical Bayes model”, i.e., the model where the prior for σ is set as a Dirac mass on $\hat{\sigma}_j(t)$, see Equation (55). In this case we have that the total capital is equal to SCR = 505.48 and the fully Bayesian model with prior defined with $k = 5$ (see Section 8.3) requires 15% more capital (both in the marginalized and conditional cases).

To check the accuracy of the SMC procedure we first analyse the estimate of the level sets (intermediate VaRs). For $p_0 = 0.5$, Figures 7 and 8 show, respectively, the histogram of the levels B_1, \dots, B_7 (as per Table 10) for the marginalized and conditional models. The red dashed bars represent the true value of the quantiles (based on the “brute force” MC simulations), which is very close to the mode of the empirical distribution of the SMC estimates. It should be noticed, though, that the SMC estimates seem to be negatively biased and the bias appears to become more pronounced for extreme quantiles. Apart from this negligible bias we assume the levels are being sensibly estimated and proceed, as in [Targino et al. \(2015\)](#), to calculate the relative bias and the variance reduction of the SMC method when compared to a MC procedure.

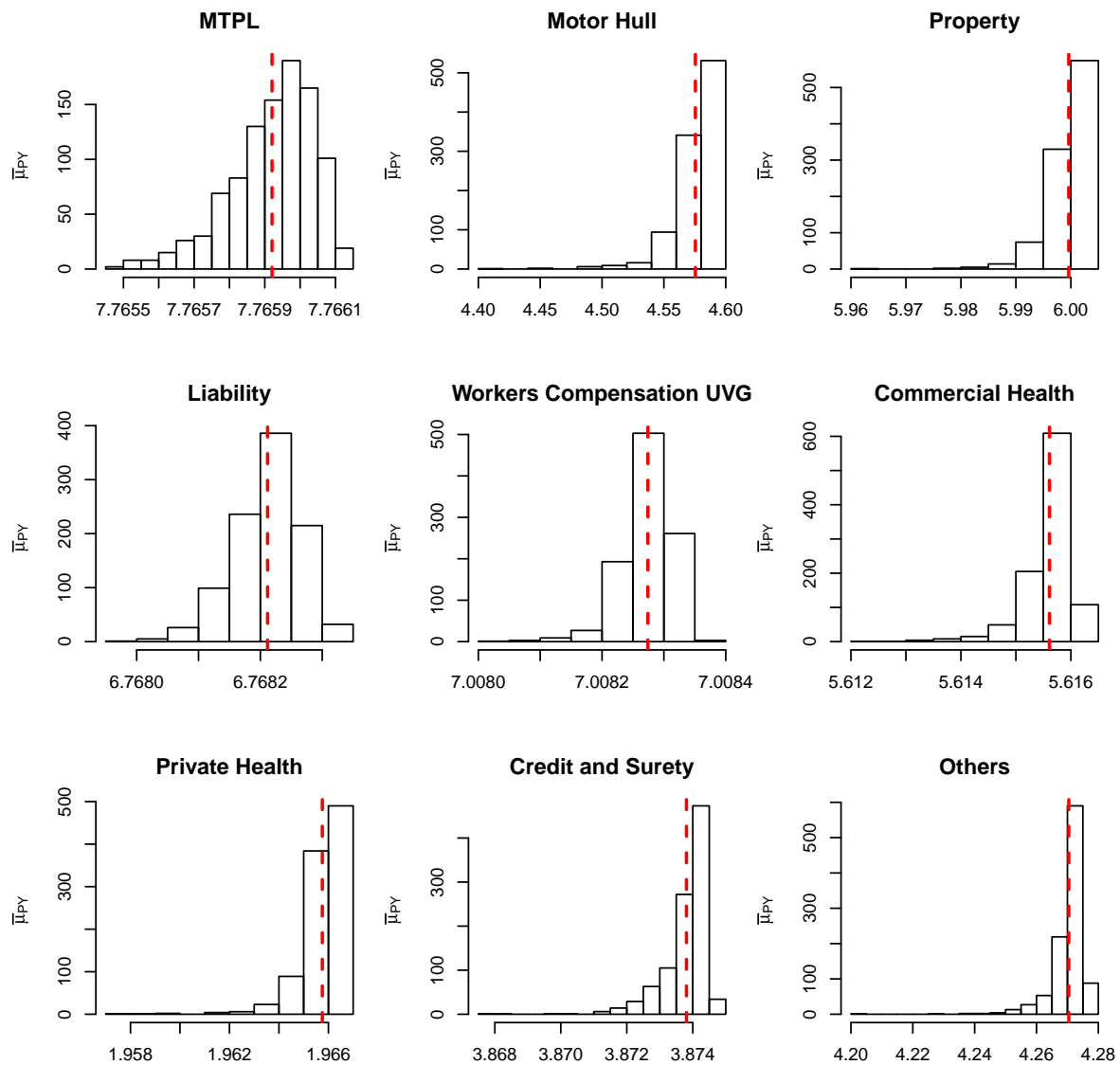


Figure 6. Histogram of the parameter $\bar{\mu}_{PY}$ for the conditional model. Red dashed line: μ_{PY} .

For each of the LoBs the plots on the Figures 9 and 10 show the relative bias, defined as

$$\text{Relative Bias} = \frac{\hat{\mathbb{E}}[\hat{\rho}_{i,SMC}] - \hat{\mathbb{E}}[\hat{\rho}_{i,MC}]}{\hat{\mathbb{E}}[\hat{\rho}_{i,MC}]},$$

where $\hat{\mathbb{E}}[\hat{\rho}_{i,SMC}]$ is computed analogously to the MC estimate but, instead, using the SMC method, with $N_{SMC} = 100$. The behaviour of the two models is very similar, and we observe that the bias in the PY and CYs allocations are negligible (less than 5%) while for some of the large CY risks a higher bias (of more than 10%) may be observed. Apart from the difficulty of performing the estimation based on Pareto distributions we stress the fact that although these errors may look large, as we can see from Table 2, their impact in the overall capital are almost imperceptible, due to the small capital charge due to these risks.

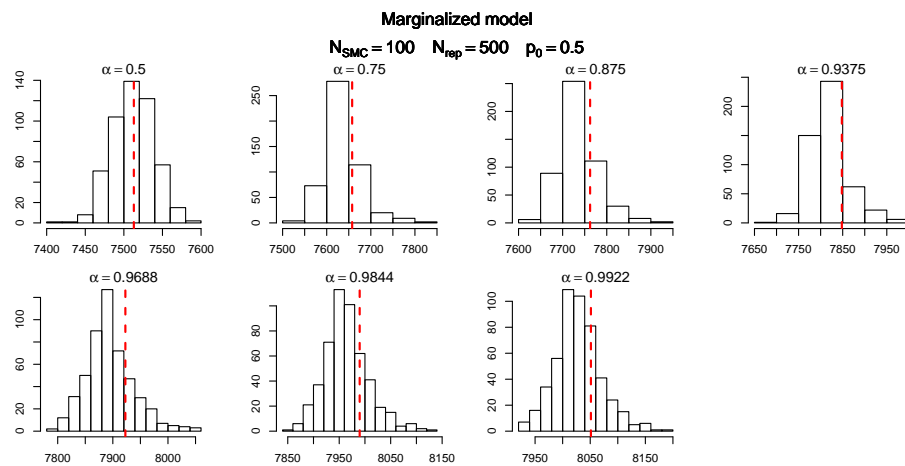


Figure 7. Histograms levels used in the SMC sampler algorithm with $p_0 = 0.5$ in the marginalized model. The red dashed bar represents the true value of the α quantile.

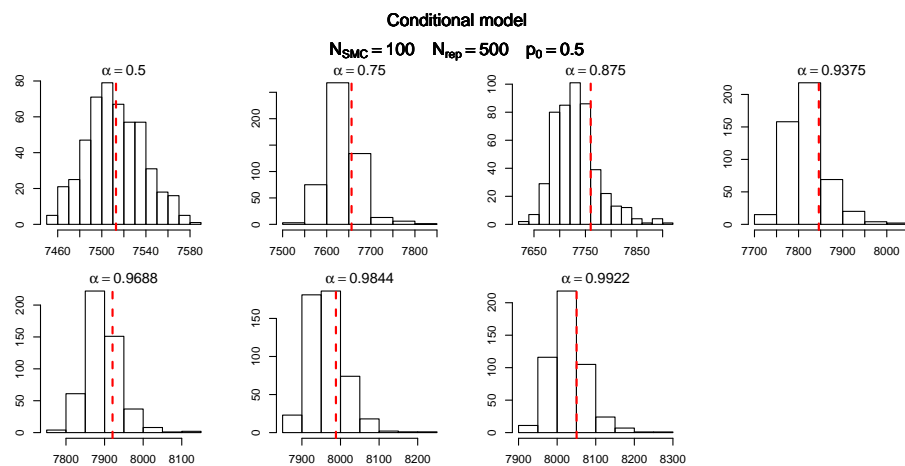


Figure 8. Histograms levels used in the Sequential Monte Carlo (SMC) sampler algorithm with $p_0 = 0.5$ in the conditional model. The red dashed bar represents the true value of the α quantile.

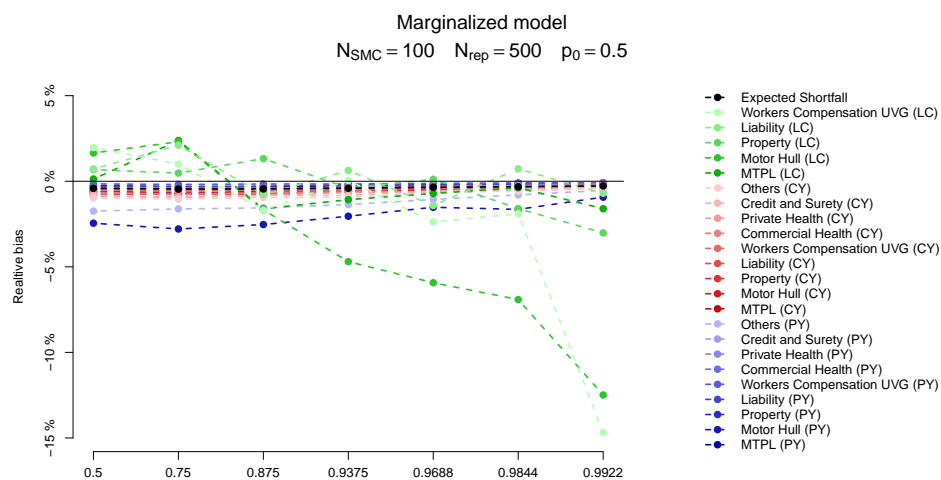


Figure 9. Bias for the marginalized model. Note that although the bias for some of the CY large claims is around 10% their allocated capital is rather small, as seen in Figure 11 (a).

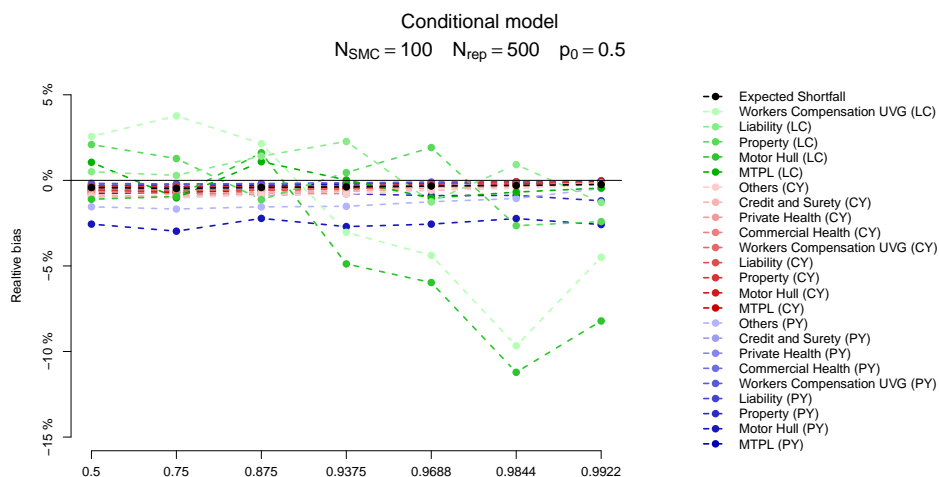


Figure 10. Bias for the conditional model. Note that although the bias for some of the current year (CY) large claims is around 10% their allocated capital is rather small, as seen in Figure 11 (b).

Another way to compare the SMC calculations is through the actual capital charges, as seen in Figure 11. In this figure we compare the 99% SCR calculated via the MC scheme discussed above with the SMC results for the quantile level right before 99% (which, for $p_0 = 0.5$ is 98.44%) and the one right after it (99.22%). From this figure we see the SMC calculation based on the 99.22% quantile is very precise, for both the marginalized and conditional models. Visually, the only perceivable difference comes from the CY,l claims, which accounts (in total) for less than 2% of the overall capital.

To calculate the improvement generated by the SMC algorithm compared to the MC procedure we need to analyse the variance of the estimates generated by both methods, under similar computational budgets.

We start by noticing that the expected number of samples in the Monte Carlo scheme in order to have N_{MC} samples satisfying the α condition is equal to $M_{MC} = N_{MC} / (1 - \alpha)$, which can be prohibitive if α is very close to 1. Then, similarly to Equation (59) we define the empirical variance of the MC and the SMC algorithms which are, then, compared as follows

$$\text{Variance Reduction} = M_{MC} \times \widehat{\text{Var}}(\hat{\rho}_{i,MC}) / T \times N_{SMC} \times \widehat{\text{Var}}(\hat{\rho}_{i,SMC}). \quad (60)$$

The variance reduction statistics defined in Equation (60) takes into account how many samples one needs to use in order to generate N_{MC} samples via rejection sampling or N_{SMC} using the SMC algorithm. The later also takes into account the fact that T levels are being used and in each one N_{SMC} samples need to be generated. For the conditional model we further multiply the denominator by the number of samples used to estimate the unknown density, which in our examples is set to $M = 5$.

The results follow on Figures 12 and 13. As in Targino et al. (2015) we observe that the variance of the SMC estimates become smaller (compared to the MC results) for larger quantiles. In particular, for the quantiles of interest the variance of the marginal ES allocation estimates are around $10^{0.5} \approx 3$ times smaller than its MC counterparts, while the overall ES estimate is slightly less variable for the MC scheme.

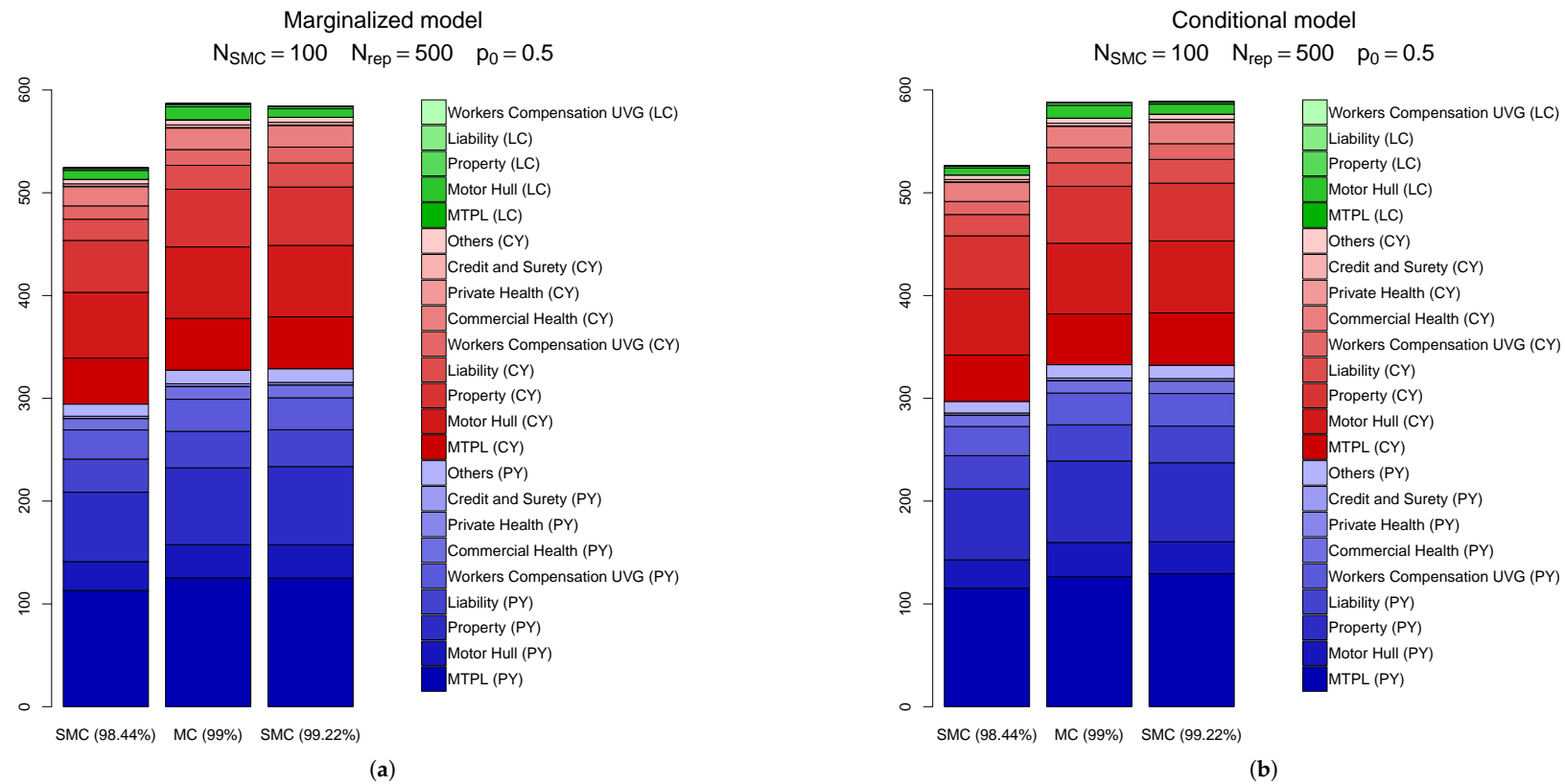


Figure 11. Comparison between the “true” allocations (calculated via a large Monte Carlo procedure) and the SMC sampler solution for the (a) marginalized and (b) conditional models.

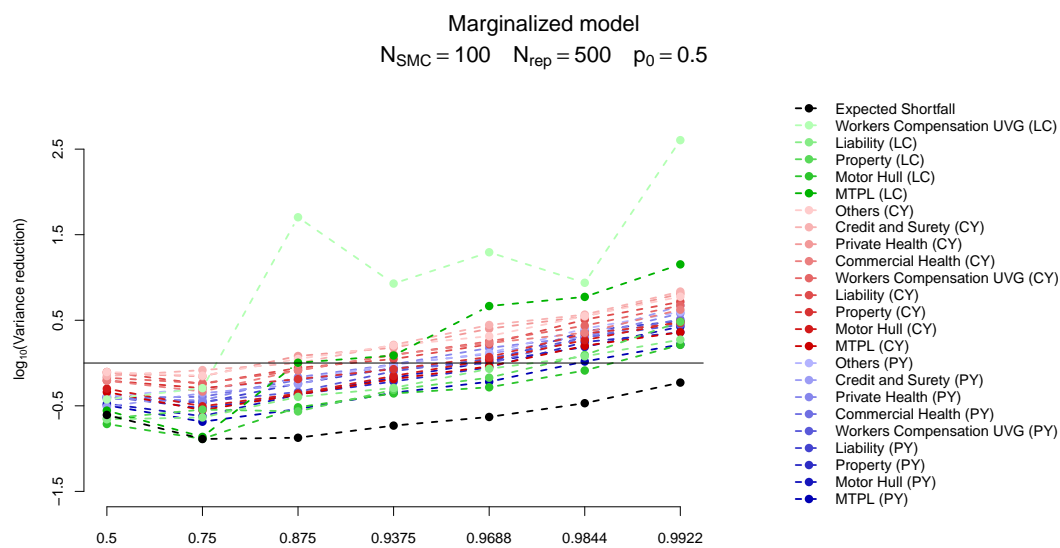


Figure 12. Variance reduction for the marginalized model.

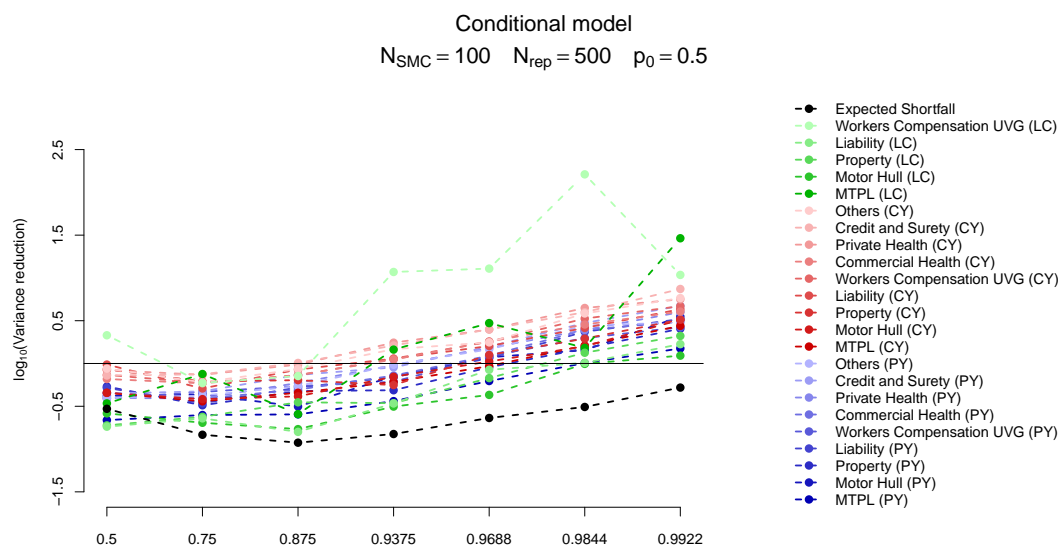


Figure 13. Variance reduction for the conditional model.

For the marginalized model we also present two plots in Figure 14, related, respectively, to the sensitivity to (a) the parameter p_0 and (b) the number of samples, N_{SMC} . In Figure 14a, for the same number of samples, $N_{SMC} = 100$ we analyse the bias relative to the 99% ES allocations of the first quantile larger than 99% (top plot) and the previous one (bottom plot) for $p_0 \in \{0.4, 0.5, 0.7\}$. The quantiles used in these different setups are presented in Table 10. Although the results may look slightly different, the main message is the same: the “higher” quantile is effectively unbiased for PY and CY,s risks but presents a negative bias of around 10% for some of the CY,l risks.

Regarding the sensitivity to the number of particles in the SMC algorithm, as expected, the absolute bias decreases when the number of samples increases, as seen in Figure 14b. Although the SMC algorithm is generically guaranteed to be unbiased when $N_{SMC} \rightarrow +\infty$ the trade-off between bias and the variance reduction in the allocation problem may lead us to accept a small bias in order to have a smaller variance.

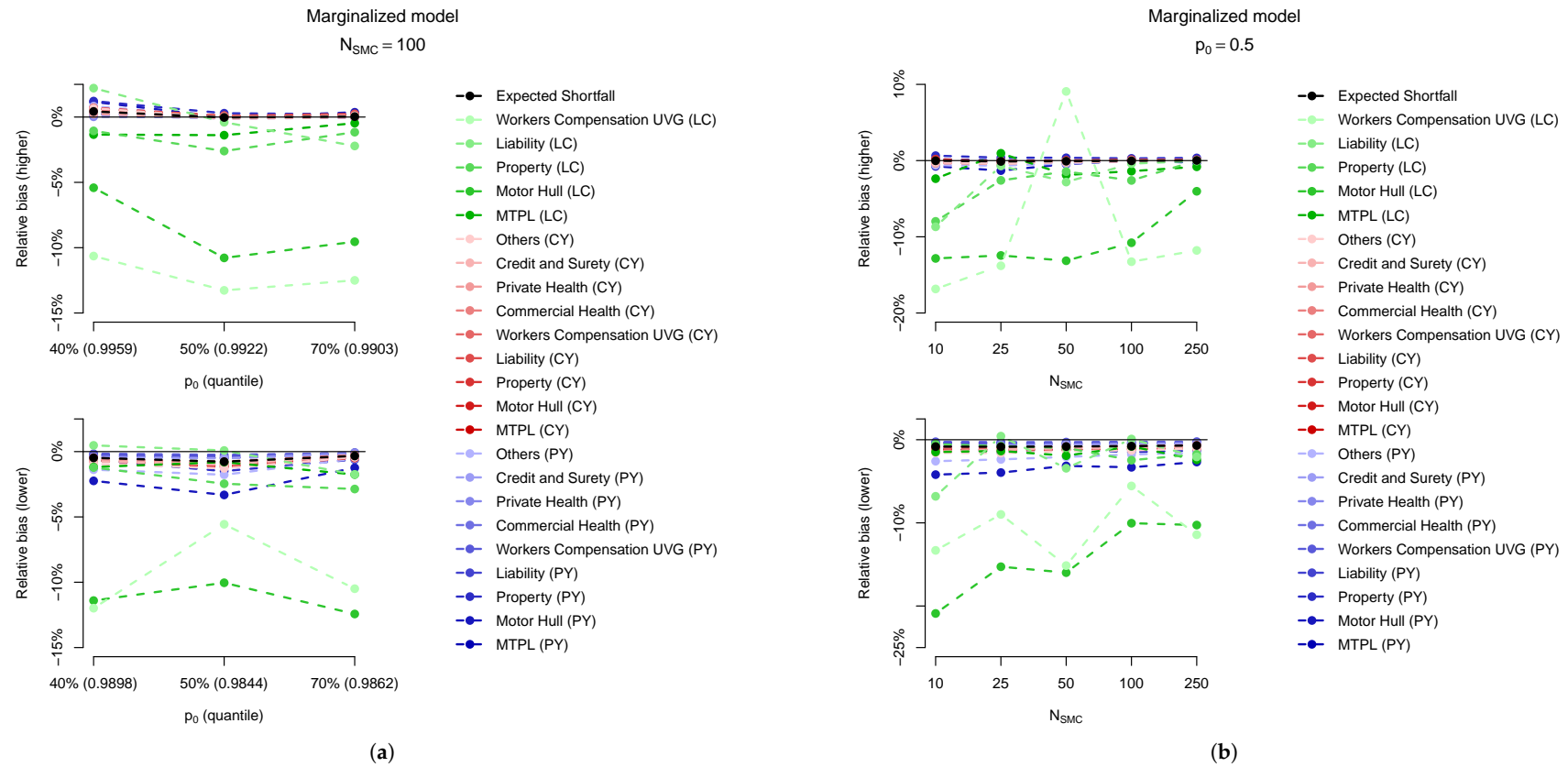


Figure 14. Relative bias in the marginalized model as a function of (a) the parameter p_0 and (b) the sample size in the SMC sampler, N_{SMC} .

11. Conclusions

In this paper we provide a complete and self-contained view of the capital allocation process for general insurance companies. As prescribed by the Swiss Solvency Test we break down the company's overall Solvency Capital Requirement (SCR) into the one-year reserve risk, due to claims from previous years (PY) and the one-year premium risk due to claims' payments in the current year (CY). The latter is further split into the risk of normal/small claims (CY,s) and large claims (CY,l). For the premium risk in each line of business we assume a log-normal distribution for CY,s risks with mean and variance as per the SST, which also describe a distribution for CY,l risks, in this case Pareto. For the reserve risk, as in Peters et al. (2017), we postulate a Bayesian gamma-gamma model which, for allocation purposes, is approximated by log-normal distributions leading to what we name the conditional (when the log-normal approximation is performed conditional on the unknown parameters) and the marginalized (when the log-normal approximation is performed after the parameter uncertainty has been integrated out) models.

As seen in Figures 1 and 2, when assuming a Bayesian gamma-gamma model these two approximations do not deviate considerably from the actual model assumptions. Regarding the allocations, Figure 11 shows the results for both models are, once again, very close to each other (and to the "true" allocations, calculated via a large Monte Carlo exercise). Therefore, the decision on which approximation to use should not interfere with the allocation or reserving results, and is left to the reader.

The allocation process is performed using state-of-the-art (pseudo-marginal) Sequential Monte Carlo (SMC) algorithms, which are presented in a self-contained and accessible format. Although the algorithms described form an extremely flexible class, we provide an off-the-shelf version, where minimal or no tuning is needed. The algorithms are also shown to be computationally efficient in a series of numerical experiments.

One of the advantages of our proposed methodology is that it is able to compute in one single loop (1) the value at risk (VaR) and (2) the Expected Shortfall (ES), both at the company level and (3) the capital allocations for the risk drivers. This procedure should be compared with routinely applied methodologies, where one simulation is performed to compute the VaR, which is used in a different simulation to compute the allocations, in a process that accumulate different errors.

Moreover, even ignoring the computational cost of calculating a precise estimate for the required VaR in a "brute force" Monte Carlo scheme, the proposed SMC algorithm is numerically shown to provide estimates that are less volatile than comparable "brute force" implementations.

Author Contributions: All authors contributed equally to the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Posterior Distributions

For ease of exposition we omit the LoB index ℓ . Under Model Assumptions 1 the posterior distribution of the parameter vectors ϕ and σ , for $t \geq I$, is given by

$$\begin{aligned}
\pi(\boldsymbol{\phi}, \boldsymbol{\sigma} | \mathcal{F}(t)) &\propto g(\mathcal{F}(t) | \boldsymbol{\phi}, \boldsymbol{\sigma}) f_{\boldsymbol{\phi}}(\boldsymbol{\phi}) f_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}) \\
&= \left[g(C_{1,0}, \dots, C_{t,0}) \prod_{j=0}^{J-1} \prod_{i=1}^{t-j-1} \frac{(\phi_j \sigma_j^{-2})^{C_{i,j} \sigma_j^{-2}}}{\Gamma(C_{i,j} \sigma_j^{-2})} C_{i,j+1}^{C_{i,j} \sigma_j^{-2}-1} \exp \left\{ -\phi_j \sigma_j^{-2} C_{i,j+1} \right\} \right] \\
&\times \left[\prod_{j=0}^{J-1} \lim_{\gamma_j \rightarrow 1} \frac{(f_j(\gamma_j - 1))^{\gamma_j}}{\Gamma(\gamma_j)} \phi_j^{\gamma_j-1} \exp \left\{ -\phi_j f_j(\gamma_j - 1) \right\} \right] \times \left[\prod_{j=0}^{J-1} f_{\sigma_j^2}(\sigma_j^2) \right] \\
&\propto \left[\prod_{j=0}^{J-1} \lim_{\gamma_j \rightarrow 1} \phi_j^{\gamma_j-1 + \sum_{i=1}^{t-j-1} C_{i,j} \sigma_j^{-2}} \exp \left\{ -\phi_j \left(f_j(\gamma_j - 1) + \sum_{i=1}^{t-j-1} C_{i,j+1} \sigma_j^{-2} \right) \right\} \right] \\
&\times \left[\prod_{j=0}^{J-1} f_{\sigma_j^2}(\sigma_j^2) \prod_{i=1}^{t-j-1} \frac{(C_{i,j+1} \sigma_j^{-2})^{C_{i,j} \sigma_j^{-2}}}{\Gamma(C_{i,j} \sigma_j^{-2})} \right] \\
&\propto \left[\prod_{j=0}^{J-1} \phi_j^{\sum_{i=1}^{t-j-1} C_{i,j} \sigma_j^{-2}} \exp \left\{ -\phi_j \left(\sum_{i=1}^{t-j-1} C_{i,j+1} \sigma_j^{-2} \right) \right\} \right] \\
&\times \left[\prod_{j=0}^{J-1} f_{\sigma_j^2}(\sigma_j^2) \prod_{i=1}^{t-j-1} \frac{(C_{i,j+1} \sigma_j^{-2})^{C_{i,j} \sigma_j^{-2}}}{\Gamma(C_{i,j} \sigma_j^{-2})} \right].
\end{aligned}$$

From the functional form of $\pi(\boldsymbol{\phi}, \boldsymbol{\sigma} | \mathcal{F}(t))$ it can be seen that the components ϕ_j of $\boldsymbol{\phi}$ and σ_j of $\boldsymbol{\sigma}$ are independent *a posteriori*, which is a direct consequence of the prior independence. Moreover, since $\pi(\boldsymbol{\phi} | \boldsymbol{\sigma}, \mathcal{F}(t)) \propto \pi(\boldsymbol{\phi}, \boldsymbol{\sigma} | \mathcal{F}(t))$, we have that

$$\phi_j | \boldsymbol{\sigma}, \mathcal{F}(t) \sim \Gamma(a_j, b_j), \quad (\text{A1})$$

with $a_j = 1 + \sum_{i=1}^{t-j-1} C_{i,j} \sigma_j^{-2}$ and $b_j = \sum_{i=1}^{t-j-1} C_{i,j+1} \sigma_j^{-2}$.

The marginal posterior $\pi(\boldsymbol{\sigma} | \mathcal{F}(t))$ and its unnormalized version $h(\boldsymbol{\sigma} | \mathcal{F}(t))$ are calculated as

$$\begin{aligned}
\pi(\boldsymbol{\sigma} | \mathcal{F}(t)) &= \int \pi(\boldsymbol{\phi}, \boldsymbol{\sigma} | \mathcal{F}(t)) d\boldsymbol{\phi} \\
&\propto \prod_{j=0}^{J-1} \frac{\Gamma(a_j)}{b_j^{a_j}} f_{\sigma_j^2}(\sigma_j^2) \prod_{i=1}^{t-j-1} \frac{(C_{i,j+1} \sigma_j^{-2})^{C_{i,j} \sigma_j^{-2}}}{\Gamma(C_{i,j} \sigma_j^{-2})} = h(\boldsymbol{\sigma} | \mathcal{F}(t)).
\end{aligned}$$

Lemma A1. (from [Peters et al. \(2017\)](#)) For $0 \leq j \leq J-1$ and $t \geq 1$ if either $t-j-1 = 1$ or at least one accident year $1 \leq i \leq t-j-1$ is such that $\frac{C_{i,j+1}}{C_{i,j}} \neq \hat{f}_j(t)$ then the marginal posterior $\pi(\boldsymbol{\sigma} | \mathcal{F}(t))$ is integrable, i.e.,

$$\int_0^{d_j} h_j(\sigma_j | \mathcal{F}(t)) d\sigma_j < \infty.$$

Appendix B. Correlation Bounds in the Log-Normal–Gaussian Copula Model

As mentioned in Section 7 and discussed, for example in ([Embrechts et al. 2002](#), Fallacy 2), for given marginal distributions not all linear correlations between -1 and 1 can be achieved. This can also be seen in the following Lemma (see ([Denuit and Dhaene 2003](#), sct. 2)).

Lemma A2 (Correlation bounds). Let (X_1, X_2) be a bivariate random variable with marginal distributions F_1 and F_2 . Then the correlation between X_1 and X_2 is bounded by

$$\frac{\text{Cov}(F_1^{-1}(U), F_2^{-1}(1-U))}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} \leq \text{Corr}(X_1, X_2) \leq \frac{\text{Cov}(F_1^{-1}(U), F_2^{-1}(U))}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}},$$

for U uniformly distributed in $[0, 1]$.

Although theoretically interesting, Lemma A2 may provide bounds that are too wide and, in some cases just state that the correlation lies between -1 and 1 . In the sequel we show that in the particular case of a random vector with log-normal marginals and Gaussian copula it is possible to calculate precisely the intended correlation and numerically check its limits.

Let us assume a random vector $\mathbf{X} = (X_1, \dots, X_{2L})$ is normally distributed with $\mathbf{X} \sim N(\mathbf{m}, \mathbf{V})$, where a general term of the covariance matrix \mathbf{V} is given by $(\mathbf{V})_{i,j} = V_{i,j}$ and $V_{i,i} = V_i^2$. Moreover, we denote by $\mathbf{\Omega} = \text{Corr}(\mathbf{X})$ the correlation matrix of the random vector \mathbf{X} , i.e.,

$$\mathbf{V} = \text{diag}(V_1, \dots, V_{2L}) \mathbf{\Omega} \text{diag}(V_1, \dots, V_{2L}),$$

with $(\mathbf{\Omega})_{i,j} = (\mathbf{\Omega})_{j,i} = \omega_{i,j}$.

If we define $Z_i = e^{X_i}$, for $i = 1, \dots, 2L$ then $Z_i \sim \text{LN}(m_i, V_i)$ with

$$\begin{aligned} \mathbb{E}[Z_i] &= \exp \left\{ m_i + \frac{V_i^2}{2} \right\}, \\ \text{Var}(Z_i) &= (\mathbb{E}[Z_i])^2 (e^{V_i^2} - 1). \end{aligned} \quad (\text{A2})$$

On the other hand, since $X_i + X_j \sim N(m_i + m_j, V_i^2 + V_j^2 + 2V_i\omega_{i,j}V_j)$ we have that

$$\mathbb{E}[Z_i Z_j] = \mathbb{E}[e^{X_i + X_j}] = \exp \left\{ m_i + m_j + \frac{V_i^2 + V_j^2 + 2V_i\omega_{i,j}V_j}{2} \right\}. \quad (\text{A3})$$

Therefore, using Equations (A2) and (A3) the correlation between Z_i and Z_j can be written as

$$\text{Corr}(Z_i, Z_j) = \frac{\exp\{V_i\omega_{i,j}V_j\} - 1}{\left[(e^{V_i^2} - 1)(e^{V_j^2} - 1)\right]^{1/2}}. \quad (\text{A4})$$

Since $\exp(\cdot)$ is a strictly increasing function and the marginal distributions of (X_1, \dots, X_{2L}) are continuous, from (McNeil et al. 2010, Proposition 5.6) we can conclude that (Z_1, \dots, Z_{2L}) has the same copula as (X_1, \dots, X_{2L}) : a Gaussian copula with correlation matrix $\mathbf{\Omega}$.

From equation Equation (A4) it is easy to see the correlation between Z_i and Z_j is a monotone function of $\omega_{i,j}$ which implies that $\text{Corr}(Z_i, Z_j)$ will be minimal when $\omega_{i,j} = -1$ and maximal when $\omega_{i,j} = 1$. Therefore, for a given pair of standard deviations it is possible to compute the interval of admissible correlations for the pair (Z_i, Z_j) . In Figure A1 the lower (left plot) and upper (right plot) present bounds for the correlations.

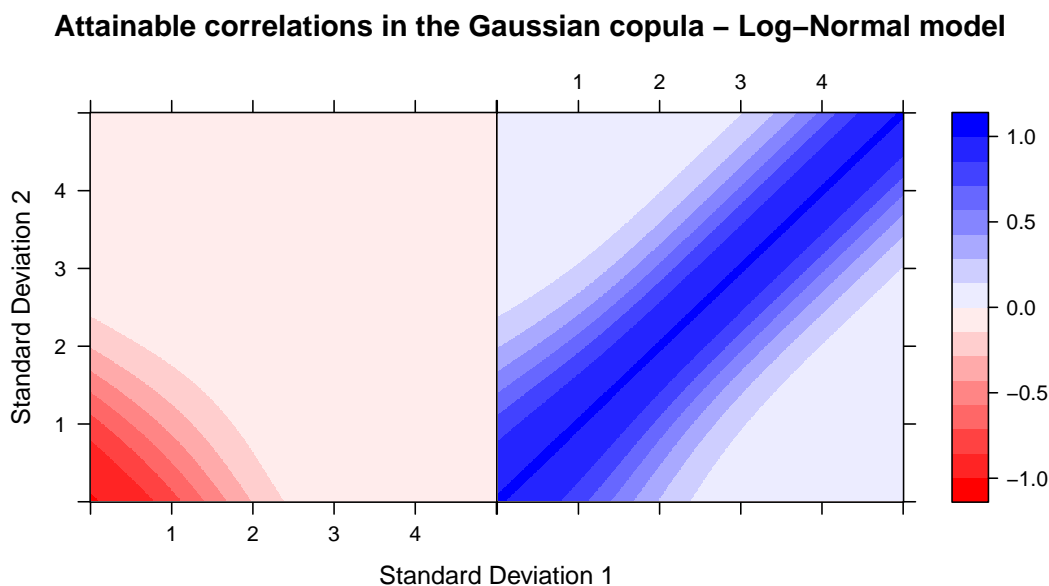


Figure A1. Lower (**left**) and upper (**right**) bound for correlations in a Gaussian-copula model with Log-Normal marginal distributions, as a function of the scale parameters σ_1 and σ_2 .

Figure A1 shows that even when the copula correlation is set to -1 if at least one of the standard deviation parameters is “large” then minimum possible correlation between the log-normal variables is close to zero. For example, if $\sigma_1 = \sigma_2 = 2$ then the lower bound for the correlations is approximately -2% . As actuarial risks are usually positively correlated this may not be a problem from the modelling point of view. The upper limit for the correlations have a different behaviour. If both standard deviations are the same then the range of attainable correlations is upper bounded by 1 , meaning that any positive correlation can be achieved. Problems arrive when the standard deviations are sufficiently different from each other. If $\sigma_1 = 1$ then the correlation is upper bounded by 66% if $\sigma_2 = 2$, 16% if $\sigma_2 = 3$ and about 1% if $\sigma_2 = 4$.

Appendix C. Data Generating Process

In this appendix we describe the process used to generate claims triangles using the balance sheet data from Table 1 in a way that the estimated reserves from the data match closely the reserves from Table 1.

First of all, for each LoB we set the maximum number of development years as the number of years it takes until $F_j = 1$, where F_j denotes the *cumulative* payment pattern for development year j (see Section 8.1). As claims in the “Motor Third Part Liability (MTPL)” and “Workers Compensation (UVG)” LoBs should take between 20 and 30 years to settle, we make a simplifying assumption that $I = \max(J + 1, 10)$.

For different accident years we calculate the present value of the runoff pattern, using a constant claim inflation $r = 2\%$ for all years and LoBs. More precisely, we have that

$$PV_i(F_j) = (1 + r)^{-i} F_j \text{ for } j = 1, \dots, J \text{ and } j + i > I.$$

For the most recent accident year, $i = I$, we define the expected ultimate claim by

$$C_{I,J}^* = R \times \frac{\sum_{j=1}^J P_{I,j}}{\sum_{j=1}^J F_j},$$

where R denotes the reserves from Table 1 and

$$P_{I,j} = \frac{PV_I(F_j)}{\sum_{i=1}^I \sum_{j=1}^J PV_i(F_j)}.$$

Note that $C_{I,J}^*$ is neither the ultimate claim predictor for the conditional model defined in Equation (21) nor the marginalized one from Equation (27). In this context $C_{I,J}^*$ is just an auxiliary variable being used in order to simulate triangles which have estimated reserves similar to the original ones in Table 1.

For the remaining accident years the expected ultimate claim is taken as the present value of $C_{I,J}^*$. In other words,

$$C_{i,J}^* = PV_{i-I}(C_{I,J}^*) = (1+r)^{I-i} C_{I,J}^*.$$

Given all the values of $C_{i,J}^*$, we compute $E_i^* = F_0 \times C_{i,J}^*$, the expected initial payment for each accident year. These values are, then, combined with the coefficients of variation for CY small claims and used to simulate the first column of our triangles as

$$C_{i,0} \sim LN(m_i^*, V_i^*),$$

with the auxiliary parameters $m_i^* = \log(E_i^*) - V_i^*/2$, $V_i^* = \log(1 + \text{CoVa}_{CY}^2)$ and CoVa_{CY} the coefficient of variation of CY small claims, based on Model Assumptions 4. For the remaining development years we follow Model Assumptions 1 (a) with $\phi_j = 1/f_j$ and $\sigma_j = s_j/f_j$, as discussed in Section 8.1.

Figure 4 presents the generated cumulative claims payments for all LoBs, where each line represents the cumulative claims payment. In each plot the lighter colours represent more recent accident years which are not yet fully developed. The reserves calculated based on this dataset are presented in Table 2 and given these values the original reserves from Table 1 are ignored.

References

- Andrieu, Christophe, Arnaud Doucet, and Roman Holenstein. 2010. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72: 269–342.
- Andrieu, Christophe, and Gareth O. Roberts. 2009. The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics* 37: 697–725.
- Andrieu, Christophe, and Matti Vihola. 2015. Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *The Annals of Applied Probability* 25: 1030–77.
- Asimit, Alexandru V., Raluca Vernic, and Ricardas Zitakis. 2013. Evaluating risk measures and capital allocations based on multi-losses driven by a heavy-tailed background risk: The multivariate Pareto-II model. *Risks* 1: 14–33.
- Bargès, Mathieu, Hélène Cossette, and Etienne Marceau. 2009. TVaR-based capital allocation with copulas. *Insurance: Mathematics and Economics* 45: 348–61.
- Beskos, Alexandros, Omiros Papaspiliopoulos, Gareth O. Roberts, and Paul Fearnhead. 2006. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68: 333–82.
- Cérou, Frédéric, Pierre Del Moral, Teddy Furon, and Arnaud Guyader. 2012. Sequential Monte Carlo for rare event estimation. *Statistics and Computing* 22: 795–808.
- Chopin, Nicolas, Pierre E Jacob, and Omiros Papaspiliopoulos. 2013. SMC²: An efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75: 397–426.
- European Commission. 2009. *Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)*. Technical Report. Available Online: <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009L0138> (accessed on 3 May 2017).

- Cossette, Hélène, Marie-Pier Côté, Etienne Marceau, and Khouzeima Moutanabbir. 2013. Multivariate distribution defined with Farlie–Gumbel–Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics and Economics* 52: 560–72.
- Creal, Drew. 2012. A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews* 31: 245–96.
- De Jong, Piet, and Ben Zehnwirth. 1983. Claims reserving, state-space models and the Kalman filter. *Journal of the Institute of Actuaries* 110: 157–81.
- Del Moral, Pierre. 1996. Non-linear filtering: interacting particle resolution. *Markov Processes and Related Fields* 2: 555–81.
- Del Moral, Pierre. 2004. *Feynman-Kac Formulae*. Berlin: Springer.
- Del Moral, Pierre, Arnaud Doucet, and Ajay Jasra. 2006. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68: 411–36.
- Del Moral, Pierre, Gareth W Peters, and Christelle Vergé. 2013. An introduction to stochastic particle integration methods: With applications to risk and insurance. In *Monte Carlo and Quasi-Monte Carlo Methods 2012*. Berlin: Springer, pp. 39–81.
- Denuit, Michel, and Jan Dhaene. 2003. Simple characterizations of comonotonicity and countermonotonicity by extremal correlations. *Belgian Actuarial Bulletin* 3: 22–27.
- Devroye, Luc, and Gérard Letac. 2015. Copulas with prescribed correlation matrix. In *In Memoriam Marc Yor-Séminaire de Probabilités XLVII*. Berlin: Springer, pp. 585–601.
- Dhaene, Jan, Luc Henrard, Zinovy Landsman, Antoine Vandendorpe, and Steven Vanduffel. 2008. Some results on the CTE-based capital allocation rule. *Insurance: Mathematics and Economics* 42: 855–63.
- Dhaene, Jan, Andreas Tsanakas, Emiliano A. Valdez, and Steven Vanduffel. 2012. Optimal capital allocation principles. *Journal of Risk and Insurance* 79: 1–28.
- Douc, Randal, and Olivier Cappé. 2005. Comparison of resampling schemes for particle filtering. Paper presented the 4th International Symposium on Image and Signal Processing and Analysis (ISPA 2005), Zagreb, Croatia, September 15–17, pp. 64–69.
- Doucet, Arnaud, and Adam M. Johansen. 2009. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering* 12: 656–704.
- Embrechts, Paul, Alexander McNeil, and Daniel Straumann. 2002. Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and Beyond*. Cambridge: Cambridge University Press, pp. 176–223.
- Embrechts, Paul, Giovanni Puccetti, Ludger Rüschendorf, Ruodu Wang, and Antonela Beleraj. 2014. An academic response to Basel 3.5. *Risks* 2: 25–48.
- Everitt, Richard G., Adam M. Johansen, Ellen Rowing, and Melina Evdemon-Hogan. 2016. Bayesian model comparison with un-normalised likelihoods. *Statistics and Computing* 27: 403–22.
- Finke, Axel. 2015. On Extended State-Space Constructions for Monte Carlo Methods. Ph.D. dissertation, University of Warwick, Coventry, UK.
- FINMA. 2007. *Technical Document on the Swiss Solvency Test*. Technical Report. Bern: FINMA.
- FINMA. 2016. Standardmodell Schadenversicherung. Available Online: <https://www.finma.ch/de/~media/finma/dokumente/dokumentencenter/myfinma/2ueberwachung/sst/standard-model-nonlife-2016.zip?la=de> (accessed on 13 July 2016).
- Fulop, Andras, and Junye Li. 2013. Efficient learning via simulation: A marginalized resample-move approach. *Journal of Econometrics* 176: 146–61.
- Furman, Edward, and Zinovy Landsman. 2005. Risk capital decomposition for a multivariate dependent gamma portfolio. *Insurance: Mathematics and Economics* 37: 635–49.
- Gandy, Axel, and F. Din-Houn Lau. 2015. The chopthn algorithm for resampling. *arXiv*, preprint arXiv:1502.07532.
- Geweke, John. 1989. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica: Journal of the Econometric Society*: 1317–39. doi:0012-9682(198911)57:6<1317:BIEMU>2.0.CO;2-5.
- Gilks, Walter R., and Carlo Berzuini. 2001. Following a moving target Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63: 127–46.
- Gordon, Neil J., David J. Salmond, and Adrian F.M. Smith. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings F-Radar and Signal Processing* 140: 107–13.

- Landsman, Zinoviy M., and Emiliano A. Valdez. 2003. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal* 7: 55–71.
- Liu, Jun S., and Rong Chen. 1995. Blind deconvolution via sequential imputations. *Journal of the American Statistical Association* 90: 567–76.
- Liu, Jun S., and Rong Chen. 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association* 93: 1032–44.
- Mack, Thomas. 1993. Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin* 23: 213–25.
- Martino, Luca, Víctor Elvira, and Francisco Louzada. 2017. Effective sample size for importance sampling based on discrepancy measures. *Signal Processing* 131: 386–401.
- McGree, James M., Christopher C. Drovandi, Gentry White, and Anthony N. Pettitt. 2015. A pseudo-marginal sequential Monte Carlo algorithm for random effects models in Bayesian sequential design. *Statistics and Computing* 26: 1–16.
- McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. 2010. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton: Princeton University Press.
- Merz, Michael, and Mario V. Wüthrich. 2015. Claims run-off uncertainty: The full picture. Available at SSRN 2524352, version of 3/Jul/2015. Available Online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2524352 (accessed on 3 May 2017).
- Neal, Radford M. 2003. Slice sampling. *Annals of Statistics* 31: 705–41.
- Panjer, Harry H. 2001. *Measurement of Risk, Solvency Requirements and Allocation of Capital within Financial Conglomerates*. Waterloo: University of Waterloo, Institute of Insurance and Pension Research.
- Peters, Gareth W. 2005. Topics in Sequential Monte Carlo Samplers. Master's thesis, University of Cambridge, Cambridge, UK.
- Peters, Gareth W., Rodrigo S. Targino, and Mario V. Wüthrich. 2017. Full bayesian analysis of claims reserving uncertainty. *Insurance: Mathematics and Economics* 73: 41–53.
- Pitt, Michael K., Ralph dos Santos Silva, Paolo Giordani, and Robert Kohn. 2012. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics* 171: 134–51.
- Sklar, Abe. 1959. Fonctions de répartition à n dimensions et leurs marges. *Fonctions de Répartition à n Dimensions et Leurs Marges* 8: 229–31.
- Targino, Rodrigo S., Gareth W. Peters, and Pavel V. Shevchenko. 2015. Sequential Monte Carlo samplers for capital allocation under copula-dependent risk models. *Insurance: Mathematics and Economics* 61: 206–26.
- Tasche, Dirk. 1999. Risk contributions and performance measurement. *Report of the Lehrstuhl für mathematische Statistik, TU München*. Available Online: <https://pdfs.semanticscholar.org/2659/60513755b26ada0b4fb688460e8334a409dd.pdf> (accessed on 3 May 2017).
- Tran, Minh-Ngoc, Marcel Scharth, Michael K. Pitt, and Robert Kohn. 2014. Importance sampling squared for bayesian inference in latent variable models. Available at SSRN 2386371. Available Online: <https://ssrn.com/abstract=2386371> (accessed on 3 May 2017).
- Vergé, Christelle, Cyrille Dufour, Pierre Del Moral, and Eric Moulines. 2015. On parallel implementation of sequential Monte Carlo methods: The island particle model. *Statistics and Computing* 25: 243–60.
- Vergé, Christelle, Jérôme Morio, and Pierre Del Moral. 2016. An island particle algorithm for rare event analysis. *Reliability Engineering & System Safety* 149: 63–75.
- Verrall, Richard J. 1989. A state space representation of the chain ladder linear model. *Journal of the Institute of Actuaries* 116: 589–609.
- Wüthrich, Mario V. 2015. Non-Life Insurance: Mathematics & Statistics. Available at SSRN 2319328, version of 29/Jun/2015. Available at SSRN 2386371. Available Online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2319328 (accessed on 3 May 2017).

