

Article

# One-Year Change Methodologies for Fixed-Sum Insurance Contracts

Michel Dacorogna <sup>1</sup>, Alessandro Ferriero <sup>2,3,\*</sup> and David Krief <sup>4</sup>

<sup>1</sup> Prime Re Solutions, 6340 Zug, Switzerland; michel@dacorogna.ch

<sup>2</sup> Department of Mathematics, UAM, Campus de Cantoblanco, 28049 Madrid, Spain

<sup>3</sup> SCOR, General Guisan-Quai 26, 8022 Zurich, Switzerland

<sup>4</sup> LPSM, Université Paris Diderot, 75013 Paris, France; krief@lpsm.paris

\* Correspondence: alessandro.ferriero@gmail.com

Received: 27 June 2018; Accepted: 26 July 2018; Published: 30 July 2018



**Abstract:** We study the dynamics of the one-year change in P&C insurance reserves estimation by analyzing the process that leads to the ultimate risk in the case of “fixed-sum” insurance contracts. The random variable ultimately is supposed to follow a binomial distribution. We compute explicitly various quantities of interest, in particular the Solvency Capital Requirement for one year change and the Risk Margin, using the characteristics of the underlying model. We then compare them with the same figures calculated with existing risk estimation methods. In particular, our study shows that standard methods (Merz–Wüthrich) can lead to materially incorrect results if the assumptions are not fulfilled. This is due to a multiplicative error assumption behind the standard methods, whereas our example has an additive error propagation as often happens in practice.

**Keywords:** one-year risk; Merz–Wüthrich; solvency II; solvency capital requirement; risk margin; fixed-sum insurance

## 1. Introduction

In the new solvency regulations, companies are required to estimate their risk over one year and to compute a Risk Margin (RM) for the rest of the time until the ultimate risk margin is reached. Actuaries are not accustomed to do so. Until recently, their task mostly consisted of estimating the ultimate claims and their risk. There are only very few actuarial methods that are designed to look at the risk over one year. Among the most popular ones is the approach proposed by [Merz and Wüthrich \(2008\)](#) as an extension of the Chain–Ladder following Mack’s assumptions ([Mack 1993](#)). They obtain an estimation of the mean square error of the one-year change based on the development of the reserve triangles using the Chain–Ladder method. An alternative way to model the one-year risk is developed by Ferriero: the Capital Over Time (COT) method ([Ferriero 2016](#)). The latter assumes a modified jump–diffusion Lévy process to the ultimate and gives a formula, based on this process, for determining the one-year risk as a portion of the ultimate risk. In a previous paper, Dacorogna et al. have presented a simple model with a binomial distribution to develop the principles of risk pricing ([Dacorogna and Hummel 2008](#)) in insurance. This has been formalized and extended to study the diversification effects in a subsequent work ([Busse et al. 2013](#)).

The goal of this paper is to study the time evolution of the risk that leads to the ultimate risk through a simple example that is easy to handle, and so calculate exactly the one year and ultimate risks. This is why we take here the example (as in [Busse et al. 2013](#)) of a binomial distribution to model the ultimate loss, and see it as the result of an evolution of different binomials over time. The company is exposed to the risk  $n$  times. Different exposures come in different steps, and each exposure has a

probability  $p$  to incur in a fixed loss, say 1 hundred/thousand/million euros, and, at the end of the  $n$  steps, the total number of exposures to the risk is equal to  $n$ . In other words, there are  $n$  policies each with a probability  $p$  of a claim; of the  $n$  policies, a number  $N_i$  settled at time  $i$  either with a claim or with no claim. For simplicity and without loss of meaning, we make the number of policies  $n$  coincide with the number of time steps (indeed, in our example, the run-off is complete on average roughly at  $\log n$ , as proved in Section 3.3). This type of loss model is similar to the standard collective model widely used in insurance. It can be encountered in reality, for example, for the so-called fixed-sum insurance contracts, for which fixed amounts are paid to the contract owner in case of accident, like the IDA Auto insurance in France, the MRH Multi-Risk Property insurance in France, the Personal Accident insurance in Japan, certain products linked to the Health insurance in Switzerland, etc.

With the help of such models, we can develop a better understanding of the dynamics at work in the claim triangles. It also allows us to compute explicitly the various variables of interest, like the Solvency Capital Requirement (SCR) based on the one-year change or the Risk Margin (RM) that will account for the cost of holding the capital for this risk during the whole period. It helps also to make more explicit the various variables that need to be considered in the problem, and have an interest on their own.

We then compare them with the capital obtained with existing risk estimation methods. In particular, our results show that the commonly used Merz–Wüthrich method can lead to very wrong estimates in cases where its assumptions are not fulfilled (which is the case for our example). This is due to a multiplicative error assumption. Instead, our example has an additive error propagation as is often the case in practice for many different types of insurance contracts. Indeed, reserving actuaries often estimate the reserves using the Chain–Ladder method for mature underwriting years and the Bornhuetter–Ferguson method (Bornhuetter and Ferguson 1972; Schmidt and Zocher 2008) for young underwriting years. In Mack (2008), Mack has used an additive model for estimating the prediction error of reserves. Hence, claim triangles which have dominant young underwriting years will manifest an additive error propagation structure. The method proposed by Ferriero in Ferriero (2016) (Capital Over Time method, i.e., COT) gives instead good estimates even when its assumptions are not fulfilled and so proves to be more robust than the Merz–Wüthrich method. Our example shows the need of alternative methods beyond Merz–Wüthrich.

The paper is organized as follows. We present in the second section the model defining and explaining the quantities of interest. We then study those quantities in the third section, among which the SCR, first analytically, then their estimates numerically. In the fourth section, we develop a framework for comparing various estimation methods using as a benchmark the regulatory capital requirements computed explicitly by our model. We compare it to the estimates obtained by the Merz–Wüthrich and the COT methods, respectively, and discuss the results. The conclusions are drawn in the last section.

## 2. The Probabilistic Model

Let us introduce a process described by random variables following binomials' distribution where the risk of a claim has a probability  $p$  and the exposure to this risk is  $n$ -times. We use the same probabilistic framework as in Busse et al. (2013). The reader can think of  $n$  as the total number of policies and of  $p$  as the probability of a claim occurring for any policy.

Let  $X$  be a Bernoulli random variable (note that all rv's introduced in this paper will be defined on the same probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ ) representing the loss obtained when throwing an unbiased dice, i.e., when obtaining a "6":

$$X = \begin{cases} 1, & \text{with probability } p = 1/6, \\ 0, & \text{with probability } 1 - p. \end{cases}$$

Recall that, for  $X \sim \mathcal{B}(p)$ ,  $\mathbb{E}(X) = p$  and  $\text{var}(X) = p(1 - p)$ .

Let  $(X_i, i = 1, \dots, n)$  be an  $n$ -sample with parent rv  $X$ , (corresponding to the sequence of exposures  $n$  independent exposures we interpret this as a one-step case). The number of losses after  $n$  exposures is modelled by  $S_n = \sum_{i=1}^n X_i$ , a binomial distribution  $\mathcal{B}(n, p)$ . Recall that

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, \dots, n. \tag{1}$$

$\mathbb{E}(S_n) = \mathbb{E}(X) = np$  and that, by independence,  $\text{var}(S_n) = np(1 - p)$ . Note that defining the loss as a Bernoulli variable clearly specifies each single loss amount as a fixed quantity. It is the accumulation of these losses that will make the final loss amount  $S_n = \sum_{i=1}^n X_i$  stochastic. One could define a different probability distribution for the  $X_i$ , but the Bernoulli distribution is the most appropriate choice for the type of exposures in our case, i.e., the case in which the policies offer a lump sum as compensation.

### 2.1. The Case of a Multi-Step $n$

We consider now the case where the risk is replicated over  $n$  steps, with given  $n \in \mathbb{N} \setminus \{0\}$ . At each step  $i$  (for  $i = 1, \dots, n$ ), the number of exposures is random and represented by a rv  $N_i$  that satisfies the condition

$$\sum_{i=1}^n N_i = n, \tag{H}$$

which implies that  $0 \leq N_1 \leq n$  and  $0 \leq N_i \leq n - \sum_{j=1}^{i-1} N_j, \forall 2 \leq i \leq n$ .

Condition (H) is part of the assumption that the ultimate loss distribution is fixed and known. This assumption facilitates our objective, which is to study how the risk can materialize over time given the knowledge of the ultimate loss distribution. Actuarial methods have been developed to estimate the ultimate risk. It is thus reasonable to assume that they are doing a good job at this. Our purpose is thus to find ways to decompose this ultimate risk over time (steps).

All over the process, we are exposed to the risk a random number of times at each step independently, and with the same probability  $p$ . We keep the same notation as in the one-step case,  $X$  being the Bernoulli rv  $\mathcal{B}(p)$  that represents the loss obtained when exposing the contract to the risk with the probability  $p$ .

At each intermediate step  $i$ , we expose the contract  $N_i$  times. For ease of notation, let us define a new rv  $A_i$ , which is the sum up to  $i$  of all the exposures:

$$A_i := \sum_{k=1}^i N_k, \quad \text{with } A_0 := 0. \tag{2}$$

For the process, we obtain losses represented by

$$\begin{aligned} & (X_1, \dots, X_{N_1}), && \text{for } i = 1, \\ & (X_{1+A_{i-1}}, \dots, X_{N_i+A_{i-1}}), && \text{for } i = 2, \dots, n, \end{aligned}$$

in particular, for  $i = n$ , after  $n$ -steps, the losses are  $(X_1, \dots, X_n)$ . The variable  $X$  will then denote the parent rv of the  $X_i$ s. Hence, when looking at the number of exposures and losses, it means that:

- (i) at an intermediate step  $i$ , the number of exposures is  $N_i$  and the number of losses obtained at this step is given by:

$$S_{N_i}^{(i)} := \sum_{j=1}^{N_i} X_{j+A_{i-1}} \stackrel{d}{=} \sum_{j=1}^{N_i} X_j =: S_{N_i}, \tag{3}$$

(the equality in distribution is discussed and proved in Appendix A) setting  $S := \sum_{j=1}^i X_j$ , and is, conditionally on  $N_i$ , a binomial rv  $(S_{N_i}^{(i)} \mid N_i = n_i) \sim \mathcal{B}(n_i, p)$ , which we denote as

$$(S_{N_i}^{(i)} \mid N_i) \sim \mathcal{B}(N_i, p).$$

Recall (see e.g., Mikosch 2004) that, by independence of  $N_i$  and  $X$ , we have

$$\mathbb{E} [S_{N_i}^{(i)}] = \mathbb{E}(X) \mathbb{E}(N_i) = p \mathbb{E}(N_i) \tag{4}$$

and

$$\text{var} (S_{N_i}^{(i)}) = \mathbb{E}(N_i) \text{var}(X) + \text{var}(N_i) (\mathbb{E}(X))^2 = p(1 - p) \mathbb{E}(N_i) + p^2 \text{var}(N_i) \tag{5}$$

(ii) up to an intermediate step  $i$ , the total number of exposures is  $A_i$  and the total number of losses is given by

$$S^{(i)} := \sum_{j=1}^i S_{N_j}^{(j)} = \sum_{j=1}^i \sum_{k=1}^{N_j} X_{k+A_{j-1}} = \sum_{k=1}^{A_i} X_k = S_{A_i}$$

and is, conditionally on  $(N_j, 1 \leq j \leq i)$ , a binomial rv:

$$(S_{A_i} \mid (N_j, 1 \leq j \leq i)) \sim \mathcal{B}(A_i, p).$$

Here, we have  $\mathbb{E} [S_{A_i}] = \mathbb{E}(X) \sum_{k=1}^i \mathbb{E}(N_k) = p \sum_{k=1}^i \mathbb{E}(N_k)$  and

$$\text{var} (S_{A_i}) = \text{var}(X) \sum_{k=1}^i \mathbb{E}(N_k) + (\mathbb{E}(X))^2 \text{var} (A_i) = p(1 - p) \sum_{k=1}^i \mathbb{E}(N_k) + p^2 \text{var} (A_i) .$$

(iii) at the end of the multi-steps process under condition (H), the total number of exposures is  $n$  and the total number of losses is, as in the one-step case

$$S_n = \sum_{i=1}^n X_i \sim \mathcal{B}(n, p) = U(n).$$

This is what is called the ultimate loss, once the process is completed.

Summarizing the notation,  $S_{N_i}^{(i)}$  is the rv of the losses at the intermediate step  $i$ ,  $S_{A_i}$  is the rv of the losses up to the intermediate step  $i$  and  $S_n$  is the rv of the ultimate losses  $U(n)$ .

For  $1 \leq i \leq n$ ,  $\mathcal{N}_i$  the  $\sigma$ -algebra (as usual,  $\mathcal{N}_0$  or  $\mathcal{F}_0$  denote the trivial  $\sigma$ -algebra  $\{\emptyset, \Omega\}$ ) generated by the sequence of the rvs  $(N_j, 1 \leq j \leq i)$

$$\mathcal{N}_i = \sigma \{N_j, 1 \leq j \leq i\} \tag{6}$$

and let  $\mathcal{F}_i$  denote the  $\sigma$ -algebra generated by the sequence of the rvs  $(N_j, 1 \leq j \leq i)$  and the corresponding losses  $(X_j, 1 \leq j \leq N_1 + \dots + N_i)$ :

$$\mathcal{F}_i = \sigma \left\{ (N_j, 1 \leq j \leq i), \left( X_j, 1 \leq j \leq \sum_{k=1}^i N_k \right) \right\}. \tag{7}$$

Note that  $\forall 1 \leq i \leq n, \mathcal{N}_i \subset \mathcal{F}_i$ .

The process described here would generate a loss triangle following a Bornhutter–Ferguson (Bornhuetter and Ferguson 1972) type of dynamic, which is essentially linear in the error propagation as

in our example, and would lead this way to the ultimate. This type of dynamic is typical of certain lines of business mentioned in the introduction and often used by reserving actuaries to estimate the ultimate loss. Such loss triangle will be generated assuming independency across different underwriting years. This assumption, together with the independency between development periods already mentioned in the introduction of Section 2, are simplifications which allow to calculate explicitly the quantities of interest (below) and are anyway two of the assumptions behind the Merz–Wüthrich methodology, which is the standard actuarial methodology object of our study.

### 2.2. Random Variables of Interest

Now, let us introduce some variables of interest, namely:

- the ultimate loss  $U(n)$ , which is  $S_n$ ,
- the expected ultimate loss, given the information up to the step  $i$  (for  $1 \leq i \leq n$ ):

$$U(i) := \mathbb{E} [ U(n) \mid \mathcal{F}_i ] = \sum_{j=1}^i S_{N_j}^{(j)} + \mathbb{E} \left[ \sum_{j=i+1}^n S_{N_j}^{(j)} \mid \mathcal{N}_i \right], \tag{8}$$

the  $X_i$  s being independent of the  $N_i$ s. Note that we can also define

$$U(0) := \mathbb{E} [U(n) \mid \mathcal{F}_0] = \mathbb{E}[U(n)] = \mathbb{E}[S_n] = np, \tag{9}$$

which corresponds to the expected loss at ultimate. Note that Equation (8) defines a martingale. Note also that it is a real number (although the rvs are integer valued).

- the variation  $D(i)$  of the expected ultimate loss between two successive steps defines exactly the one year change, when choosing yearly steps:

$$D(i) = U(i) - U(i - 1). \tag{10}$$

Here  $D(i)$  is also a real rv. When  $i = 1$ , the  $D(1)$  is closely related to the solvency capital required as defined in the Solvency II framework, which reflects the risk of changes in the technical provision in one year. The difference lies in that  $D(1)$  does not take into account the risk margin change. However, this is of minor importance in the SCR estimation because the risk margin change is of a smaller order of magnitude. Indeed, in practice, it is commonly accepted that the risk margin, which represents the risk loading for the market value of the liability, is approximately constant from one year to the other. We note here that the  $D(i)$ s are the innovation of the martingale defined in Equation (8).

For simplicity, all the quantities of interest are considered undiscounted in our paper. The discounting would anyway not invalidate our findings, but only make the calculations more complex.

Note that, as in (ii), we can express the expected ultimate loss using the conditional expectations, as

$$U(n) = U(0) + \sum_{i=1}^n D(i). \tag{11}$$

The goal of this study is to understand better the behavior of  $D(i)$ . This means looking for the distribution of  $U(i)$ .

Using (8) and (10), and the properties of the conditional expectation, we can write, after straightforward computations,

$$D(i) = S_{N_i}^{(i)} + \mathbb{E} \left[ \sum_{j=i+1}^n S_{N_j}^{(j)} \mid \mathcal{N}_i \right] - \mathbb{E} \left[ \sum_{j=i}^n S_{N_j}^{(j)} \mid \mathcal{N}_{i-1} \right] \tag{12}$$

so that

$$\mathbb{E} \left[ \sum_{j=i+1}^n S_{N_j}^{(j)} \mid \mathcal{N}_i \right] = \mathbb{E} \left[ \sum_{j=A_i+1}^{A_i+(n-A_i)} X_j \mid \mathcal{N}_i \right] = (n - A_i) \mathbb{E}[X].$$

Applying (12) provides

$$D(i) = S_{N_i}^{(i)} + p(n - A_i) - p(n - A_{i-1}) = S_{N_i}^{(i)} - N_i p. \tag{13}$$

We can see from (13) that  $D(i)$  depends on  $N_i$  but also on the past information via  $A_{i-1}$ . We can then write the probability distribution of  $D(i)$  conditional on this information:

$$\begin{aligned} \mathbb{P}(D(i) = x \mid A_{i-1} = a_{i-1}) &= \mathbb{P}(S_{N_i}^{(i)} - N_i p = x \mid A_{i-1} = a_{i-1}) \\ &= \sum_{j=0}^{n-a_{i-1}} \mathbb{P}(S_j^{(i)} - jp = x \mid A_{i-1} = a_{i-1}, N_i = j) \\ &\quad \cdot \mathbb{P}(N_i = j \mid A_{i-1} = a_{i-1}) \\ &= \sum_{j=0}^{n-a_{i-1}} \mathbb{P}(S_j - jp = x) \cdot \mathbb{P}(N_i = j \mid A_{i-1} = a_{i-1}) \end{aligned} \tag{14}$$

using the independence between the  $X_i$ s and the  $N_j$ s in the equation before the last, and (3) in the last one.

Let us now choose a probability distribution for the  $N_i$ s. This choice is arbitrary. We could, for instance, use various distributions typical for modelling frequency or emergence of claims in actuarial modelling, like the Poisson distribution. For simplicity, we will pick first the case of a uniform distribution conditionally to  $\mathcal{N}$  so that, for  $2 \leq i \leq n - 1$ , we have

$$\mathbb{P}[N_i = n_i \mid (N_j = n_j, 1 \leq j \leq i - 1)] = \frac{1}{1 + n - \sum_{j=1}^{i-1} n_j}.$$

We remind the reader that  $n = \sum_{i=1}^n N_i$ . Strictly speaking, we do not even need to have  $N_1, \dots, N_{i-1}$  explicitly. Only their sum matters so that

$$\mathbb{P}[N_i = n_i \mid A_{i-1} = a_{i-1}] = \frac{1}{1 + n - a_{i-1}}. \tag{15}$$

We can now proceed to compute the expectation of  $N_i$ .

**Proposition 1.** *The expectation of  $N_i$  as a function of  $i$  is equal to:*

$$\mathbb{E}(N_i) = \frac{n}{2^i} \text{ for } i = 1, \dots, n - 1. \tag{16}$$

**Proof of Proposition 1.** Using the property of the conditional expectation, then that  $N_1$  is uniformly distributed, and finally the linearity of the expectation, we obtain:

$$\mathbb{E}(N_i) = \mathbb{E} \left( \mathbb{E}(N_i \mid \mathcal{F}_{i-1}) \right) = \mathbb{E} \left( \frac{n - \sum_{j=1}^{i-1} N_j}{2} \right) = \frac{1}{2} n - \frac{1}{2} \sum_{j=1}^{i-1} \mathbb{E}(N_j). \tag{17}$$

Let us prove the proposition by induction. For  $i = 1$ , by uniformity of the distribution of the rv  $N$ , we have:

$$\mathbb{E}(N_1) = \sum_{n_1=0}^n n_1 \mathbb{P}[N_1 = n_1] = \frac{n}{2}.$$

Suppose (16) is true for any  $i \leq j-1$  and let us prove it for  $i = j$ . This is straightforward using (17). Thus, the result  $\mathbb{E}(N_j) = \frac{1}{2} n \left( 1 - \sum_{i=1}^{j-1} \frac{1}{2^i} \right) = \frac{n}{2^j}$ .  $\square$

The last step has actually the same expectation as the previous one since it is the complementary and is completely conditioned by the sum of the previous steps

$$\mathbb{E}(N_n) = \mathbb{E} \left( n - \sum_{j=1}^{n-1} N_j \right) = n - \sum_{j=1}^{n-1} \mathbb{E}(N_j) = n - \sum_{j=1}^{n-1} \frac{n}{2^j} = \frac{n}{2^{n-1}}. \quad (18)$$

We have thus fully characterized the expectation of the number of exposures at each step. (In the Appendix B one can find the distribution of the  $N_i$  and also of the  $D(i)$ .)

### 3. Analytical Expressions of Quantities to Be Studied

#### 3.1. Incremental Pattern and Capital

Another quantity of interest is the incremental pattern  $\gamma_i$  of the expected value of the losses. It can be written, using (4) and (9), as:

$$\gamma_i := \frac{\mathbb{E} \left( S_{N_i}^{(i)} \right)}{\mathbb{E}(U(n))} = \frac{1}{2^i}, \quad \text{for } i = 1, \dots, n-1. \quad (19)$$

For computing the RM, we need to compute the capital requirement (SCR) at each step, using the risk measures  $\text{VaR}_{99.5\%}$  and  $\text{TVaR}_{99\%}$ , which we will simply call  $\rho$ . We define the capital requirement  $C_i$  at step  $i$  as

$$C_i \equiv \mathbb{E} \left( \rho(D(i) \mid \mathcal{F}_{i-1}) \right). \quad (20)$$

Once we have this quantity, we can then write the RM,  $R_n$ , as a function of the number of steps  $n$ , as:

$$R_n = \eta \sum_{i=1}^n C_i, \quad (21)$$

where  $\eta$  designates the cost of capital.

#### 3.2. Moments of $D(i)$

The properties we present here are specific to our choice of process, but they can be easily generalized in the framework of martingale assumptions. Let us compute the first two conditional moments of the vector  $\mathbf{D} = (D(1), \dots, D(n))$  given  $\mathcal{F}$ .

**Proposition 2.** Let  $0 \leq i \leq n - 1$ . For  $k$  such that  $i < k \leq n$ , we have

(i) 
$$\mathbb{E}[D(k) | \mathcal{F}_i] = 0, \tag{22}$$

(ii) 
$$\text{var}(D(k)|\mathcal{F}_i) = \begin{cases} \frac{(n-A_i)p(1-p)}{2^{k-i}} & \text{if } k \neq n, \\ \frac{(n-A_i)p(1-p)}{2^{n-i-1}} & \text{if } k = n, \end{cases} \tag{23}$$

(iii) For  $k = i + 1$ ,

$$\text{Cov}(\mathbf{D}(i + 1) | \mathcal{F}_i) = p(1 - p)(n - A_i) \cdot \begin{pmatrix} 2^{-1} & 0 & \dots & 0 \\ 0 & 2^{-2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 2^{1+i-n} & 0 \\ & & 0 & 2^{1+i-n} \end{pmatrix}, \tag{24}$$

where  $\mathbf{D}(i + 1) := (D(i + 1), \dots, D(n))$ .

**Corollary 1.** From which, we can deduce the following two expressions as corollary:

(a) As a consequence of the Proposition 2, the moments of  $\mathbf{D}$  are given by

$$\mathbb{E}(\mathbf{D}) = \mathbf{0} \quad \text{and} \quad \text{Cov}(\mathbf{D}) = np(1 - p) \cdot \begin{pmatrix} 2^{-1} & 0 & \dots & 0 \\ 0 & 2^{-2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 2^{1-n} & 0 \\ & & 0 & 2^{1-n} \end{pmatrix}. \tag{25}$$

(b) The conditional variance of the ultimate is

$$\text{var}(U(n) | \mathcal{F}_i) = (n - A_i) p(1 - p).$$

**Proof of Proposition 2.** (i) Using the definitions (8) and (10) and the tower property (with  $k > i$ ) produces

$$\mathbb{E}[D(k) | \mathcal{F}_i] = \mathbb{E}(\mathbb{E}[U(n) | \mathcal{F}_k] | \mathcal{F}_i) - \mathbb{E}(\mathbb{E}[U(n) | \mathcal{F}_{k-1}] | \mathcal{F}_i) = \mathbb{E}[U(n) | \mathcal{F}_i] - \mathbb{E}[U(n) | \mathcal{F}_i] = 0.$$

(ii) The conditional variance of  $D(i)$  can be obtained in the following way: again, take  $k \in \{i + 1, \dots, n\}$ , and recall the general property of conditional variance, for two random variables  $X$  and  $Y$ ,  $\text{var}(X) = \mathbb{E}(\text{var}(X|Y)) + \text{var}(\mathbb{E}(X|Y))$ . By using this result on  $\text{var}(D(k)|\mathcal{F}_i)$ , we obtain

$$\text{var}(D(k) | \mathcal{F}_i) = \mathbb{E}(\text{var}(D(k)|N_k) | \mathcal{F}_i) + \text{var}(\mathbb{E}(D(k) | N_k) | \mathcal{F}_i). \tag{26}$$

We now use Equation (13) on both terms of the right side of the last equation. Since  $S_{N_k}^{(k)} \sim \mathcal{B}(N_k, p)$ ,

$$\mathbb{E}(D(k) | N_k) = \mathbb{E} \left( S_{N_k}^{(k)} - N_k p \mid N_k \right) = \underbrace{\mathbb{E} \left( S_{N_k}^{(k)} \mid N_k \right)}_{=N_k p} - N_k p = 0,$$

which sets the second term of Equation (26) to 0. Let us calculate the first term.

$$\begin{aligned} \mathbb{E}\left(\text{var}\left(D(k) \mid N_k\right) \mid \mathcal{F}_i\right) &= \mathbb{E}\left(\text{var}\left(S_{N_k}^{(k)} - N_k p \mid N_k\right) \mid \mathcal{F}_i\right) \\ &= \mathbb{E}\left(\text{var}\left(S_{N_k}^{(k)} \mid N_k\right) \mid \mathcal{F}_i\right) \\ &= \mathbb{E}(N_k p(1-p) \mid \mathcal{F}_i) \\ &= p(1-p) \mathbb{E}(N_k \mid \mathcal{F}_i). \end{aligned}$$

We now consider the following argument: Assuming  $\mathcal{F}_i$  known, is equivalent in terms of rvs to starting a new process with  $n - i$  steps and  $n - A_i$  rvs. A generalization of Equations (16) and (18) therefore gives

$$\mathbb{E}(N_k \mid \mathcal{F}_i) = \frac{n - A_i}{2^{k-i}}, \quad \text{if } k \in \{i + 1, \dots, n - 1\}$$

and

$$\mathbb{E}(N_n \mid \mathcal{F}_i) = \mathbb{E}(N_{n-1} \mid \mathcal{F}_i) = \frac{n - A_i}{2^{n-i-1}}.$$

It follows then that

$$\text{var}(D(k) \mid \mathcal{F}_i) = \frac{(n - A_i) p(1-p)}{2^{k-i}} \quad \text{if } k \neq n$$

and that

$$\text{var}(D(n) \mid \mathcal{F}_i) = \frac{(n - A_i) p(1-p)}{2^{n-i-1}}.$$

We also notice that, since  $N_n = n - A_{n-1}$  is  $\mathcal{F}_{n-1}$ -measurable,

$$\text{var}(D(n) \mid \mathcal{F}_{n-1}) = \text{var}\left(S_{N_n}^{(n)} \mid \mathcal{F}_{n-1}\right) = N_n p(1-p).$$

This finishes the calculation of the conditional variance.

(iii) Only calculation of the conditional covariance remains. Let  $i \in \{0, 1, \dots, n - 2\}$  and  $j, k \in \{i + 1, \dots, n\}$ , with  $j < k$ . Since  $\mathbb{E}(D_j \mid \mathcal{F}_i) = 0$ ,

$$\begin{aligned} \text{Cov}(D(j), D(k) \mid \mathcal{F}_i) &= \mathbb{E}(D(j)D(k) \mid \mathcal{F}_i) \\ &= \mathbb{E}(\mathbb{E}(D(j)D(k) \mid \mathcal{F}_j) \mid \mathcal{F}_i) \\ &= \mathbb{E}(D(j) \underbrace{\mathbb{E}(D(k) \mid \mathcal{F}_j)}_{=0} \mid \mathcal{F}_i) = 0. \end{aligned}$$

This finishes the characterization of the second moment of  $\mathbf{D}(i)$ . Indeed, we have shown that  $\forall i \in \{0, 1, \dots, n - 1\}$ ,

$$\mathbb{E}(\mathbf{D}(i + 1) \mid \mathcal{F}_i) = 0$$

and

$$\text{Cov}(\mathbf{D}(i + 1) \mid \mathcal{F}_i) = p(1-p)(n - A_i) \cdot \begin{pmatrix} 2^{-1} & 0 & \dots & 0 \\ 0 & 2^{-2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 2^{1+i-n} & 0 \\ & & 0 & 2^{1+i-n} \end{pmatrix}. \tag{27}$$

□

**Proof of Corollary 1.** Since the particular case  $i = 1$  corresponds to the unconditional moments of  $\mathbf{D}$ , we have:

$$\mathbb{E}(\mathbf{D}) = \mathbb{E}(\mathbf{D} \mid \mathcal{F}_0) \quad \text{and} \quad \text{Cov}(\mathbf{D}) = \text{Cov}(\mathbf{D} \mid \mathcal{F}_0).$$

(a) Thus, the unconditional moments of  $D$  are given by

$$\mathbb{E}(\mathbf{D}) = 0 \quad \text{and} \quad \text{Cov}(\mathbf{D}) = np(1-p) \cdot \begin{pmatrix} 2^{-1} & 0 & \dots & 0 \\ 0 & 2^{-2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 2^{1-n} & 0 \\ & & 0 & 2^{1-n} \end{pmatrix}, \quad (28)$$

(b) As an immediate consequence of Equation (24), one can write the conditional variance of the ultimate as

$$\begin{aligned} \text{var}(U(n) \mid \mathcal{F}_i) &= \text{var} \left( U(0) + \sum_{j=1}^i D(j) + \sum_{j=i+1}^n D(j) \mid \mathcal{F}_i \right) \\ &= \sum_{j=i+1}^n \text{var}(D(j) \mid \mathcal{F}_i) = (n - A_i) p(1 - p). \end{aligned}$$

(This formula can also be derived directly, arguing that  $U(n) \mid \mathcal{F}_i \sim \mathcal{B}(n - \sum_{j=1}^i N_j, p)$ .)  $\square$

### 3.3. Completion Time

Our process is defined on a finite time scale of  $n$ -steps. However, the process will, most of the time, finish much before the  $n$ -th step (the ultimate time). It is therefore of interest to know how fast the process reaches the maximum exposures allowed. This is also very important for computational reasons. Indeed, when simulating from the process, one can stop the simulation procedure earlier than the ultimate time thus sparing large amounts of calculations. In order to answer this question, we first consider a type of process that differs slightly from ours in the sense that it has an infinite time range (infinite steps), but is still only allowed a finite number of exposures,  $n$ . For an infinite time process with  $n$  rvs, let us denote by

$$T_n = \inf \left\{ t : \sum_{k=1}^t N_k = n \right\}$$

the completion time given by the step at which the last exposure is realized. Before trying to approximate a finite process with an infinite process, we need to show that the infinite process finishes with probability 1.

**Proposition 3.** *The probability of the completion time at infinity is 0:*

$$P(T_n = \infty) = 0. \quad (29)$$

**Proof of Proposition 3.** For  $n = 1$ , we can argue that

$$\begin{aligned} \mathbb{P}(T_1 = \infty) &= \underbrace{\mathbb{P}(T_1 = \infty \mid N_1 = 1)}_{=0} \underbrace{\mathbb{P}(N_1 = 1)}_{=\frac{1}{2}} + \underbrace{\mathbb{P}(T_1 = \infty \mid N_1 = 0)}_{=\mathbb{P}(T_1 = \infty)} \underbrace{\mathbb{P}(N_1 = 0)}_{=\frac{1}{2}} \\ &= \frac{1}{2} \mathbb{P}(T_1 = \infty), \end{aligned}$$

which implies that  $\mathbb{P}(T_1 = \infty) = 0$ . Note that  $\mathbb{P}(T_1 = \infty \mid N_1 = 0) = \mathbb{P}(T_1 = \infty)$  is justified by the fact that, since the process is infinite, assuming  $N_1 = k$  is equivalent to starting a new infinite process with  $n - k$  rvs. This argument is used here with  $k = 0$  but is explained with general  $k$  because we will use it further for other values of  $k$ . For  $n > 1$ , we can do very similar calculations by induction. Assume that  $\mathbb{P}(T_i = \infty) = 0, \forall i \in \{1, \dots, n - 1\}$ . Then,

$$\begin{aligned} \mathbb{P}(T_n = \infty) &= \underbrace{\mathbb{P}(T_n = \infty \mid N_1 = 0)}_{=\mathbb{P}(T_n = \infty)} \underbrace{\mathbb{P}(N_1 = 0)}_{=\frac{1}{n+1}} + \sum_{k=1}^n \underbrace{\mathbb{P}(T_n = \infty \mid N_1 = k)}_{=\mathbb{P}(T_{n-k} = \infty)} \underbrace{\mathbb{P}(N_1 = k)}_{=\frac{1}{n+1}} \\ &= \frac{1}{n+1} \mathbb{P}(T_n = \infty) + \frac{1}{n+1} \sum_{k=1}^n \underbrace{\mathbb{P}(T_{n-k} = \infty)}_{=0 \text{ by ind. hyp.}} \\ &= \frac{1}{n+1} \mathbb{P}(T_n = \infty), \end{aligned}$$

which again proves that  $\mathbb{P}(T_n = \infty) = 0$ .  $\square$

Since the process finishes with probability 1 and the probability of an infinite process with  $n$  rvs finishing after time  $n$  is very small for  $n$  large enough, the approximation is very reasonable.

We would like to find a formula for  $\mathbb{E}(T_n)$ . By conditioning on the first step, we obtain

$$\mathbb{E}(T_n) = \mathbb{E}\left(\mathbb{E}(T_n \mid \mathcal{F}_1)\right) = \sum_{k=0}^n \mathbb{E}(T_n \mid N_1 = k) \mathbb{P}(N_1 = k).$$

Again, assuming  $N_1 = k$  is equivalent to starting a new process with  $n - k$  rvs. Therefore,  $\mathbb{E}(T_n \mid N_1 = k) = 1 + \mathbb{E}(T_{n-k})$ , where the 1 takes into account the first step. As  $\mathbb{E}(T_n)$  only depends on  $n$ , let us denote it by  $\mathbb{E}(T_n) = f(n)$ . We can then write the following iterative formula

$$f(n) = \frac{1}{n+1} \sum_{k=0}^n (1 + f(n-k)) = 1 + \frac{f(0) + \dots + f(n)}{n+1}.$$

Solving for  $f(n)$  and realising that  $f(0) = 0$ , we obtain

$$f(n) = 1 + \frac{1 + f(1) + \dots + f(n-1)}{n},$$

which gives by iteration, for  $n \geq 2$ :

$$f(n) = \frac{1}{n} + f(n-1). \tag{30}$$

Note that

$$f(1) = \mathbb{E}(T_1) = \sum_{j=1}^{\infty} j \mathbb{P}(T_1 = j) = \sum_{j=1}^{\infty} j \frac{1}{2} \frac{1}{2^{j-1}} = 2.$$

Hence, by iterating Equation (30), we obtain the formula

$$f(n) = f(1) + \sum_{j=2}^n \frac{1}{j} = 1 + \sum_{j=1}^n \frac{1}{j}. \tag{31}$$

We have now an expression that gives the average completion time of an infinite process as 1 plus the truncated harmonic series, which can itself be approximated by

$$f(n) = 1 + \sum_{j=1}^n \frac{1}{j} = 1 + \gamma + \ln(n) + \mathcal{O}(n^{-1}) \approx \ln(n) + 1.57, \tag{32}$$

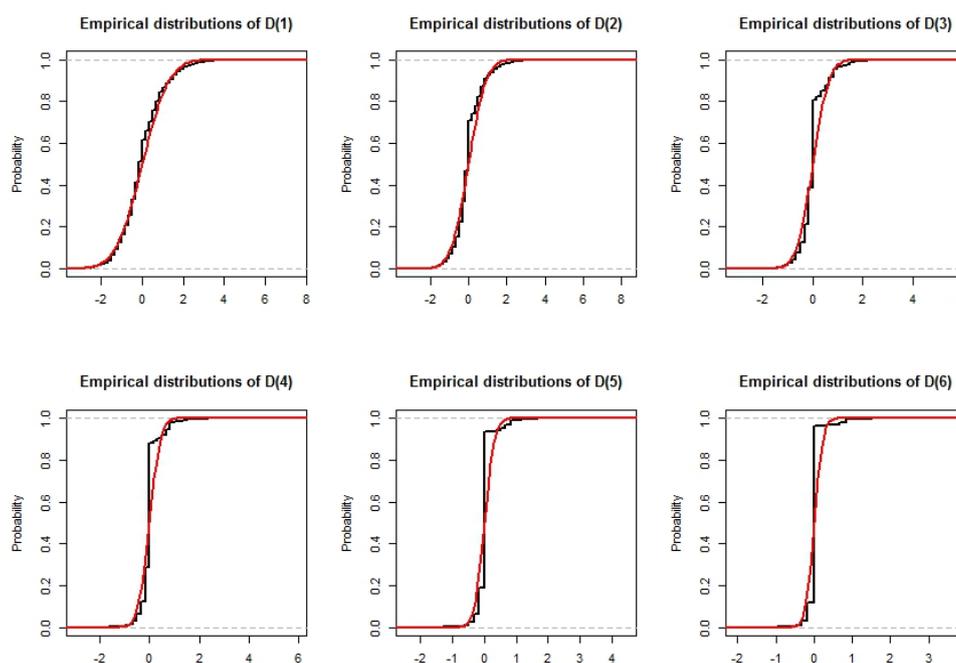
where  $\gamma \approx 0.5772$  is the Euler constant. If  $n$  is large enough, we therefore have a simple way to estimate approximately how many steps the process will last on average.

### 3.4. Distribution of the $D(i)$

In Section 3.2, we studied analytically the two first moments of the  $D(i)$ . Here, we simulate 200,000 times the process of size  $n = 15$  to estimate the empirical distribution of the  $D(i)$ . We also compute the normal distribution with same mean and variance to compare the distributions. The results are displayed in Figure 1. Let us recall Equation (13), which states  $D(i) = S_{N_i}^{(i)} - N_i p$ . Conditional on  $N_i$ , this random variable has centered binomial distribution (i.e., binomial distribution minus its average). The Central Limit Theorem applies to binomial rv  $X$  with parameters  $n, p$ . Thus, the distribution,  $f$ , of a sum of  $n$  independent Bernoulli rvs with parameter  $p$  will converge to a Gaussian distribution:  $f(X) \approx \mathcal{N}(np, np(1 - p))$ , for  $n$  large enough. Assume that  $F$  is a discrete mixture of normal distributions with mean 0, variances  $\sigma_1^2, \dots, \sigma_N^2$  and weights  $p_1, \dots, p_n$ . Then,  $\forall x \in \mathbb{R}$ ,

$$F(x) = \sum_{i=1}^n p_i \cdot \mathbb{P}[W \leq x] = \sum_{i=1}^n p_i \cdot \Phi(x/\sigma_i),$$

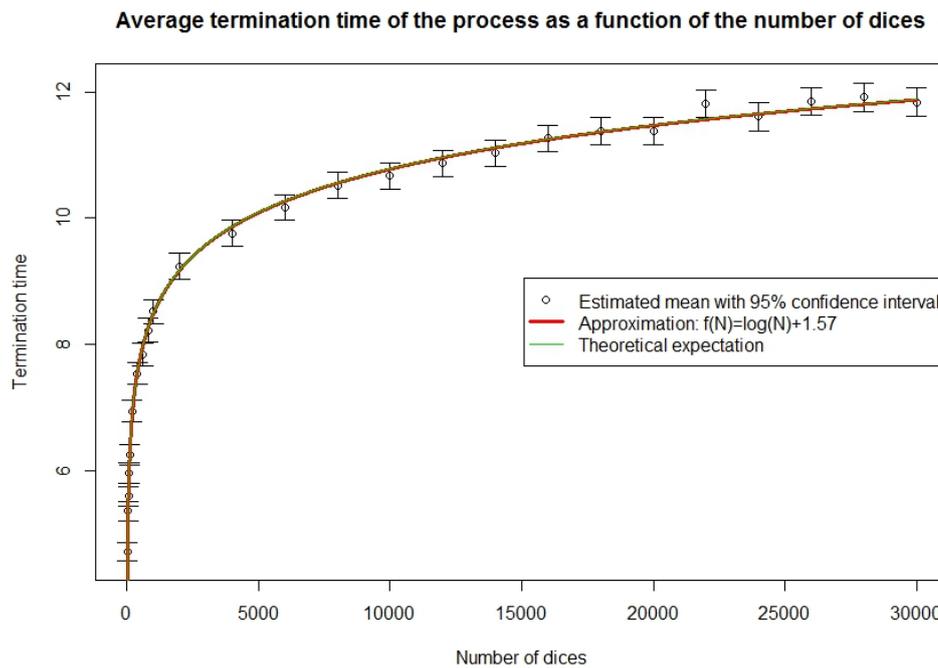
where  $W$  is a normal rv  $\mathcal{N}(0, \sigma_i^2)$ . In particular, if for some indices  $i$  and  $j$ ,  $1/\sigma_i$  and  $1/\sigma_j$  are close to each other, then  $\Phi(x/\sigma_i)$  and  $\Phi(x/\sigma_j)$  are also going to be close to each other. In our case,  $1/\sigma_i = (ip(1 - p))^{-1/2}$  and these values are going to be close when  $i$  is large enough. If a sufficiently large part of the weight is on large values of  $i$ , our mixture of centred binomial distributions is going to be close to a Normal distribution. In other words,  $D(i)$  is going to have a distribution that is close to Normal only if the probability of having a large  $N_i$  is high. For  $n = 15$ , as we can see in Figure 1, it is the case. For  $i = 2$ , the approximation is still reasonable despite a mass larger than normal around 0. For  $i > 2$ , the fit is rather good for large and small quantiles but not for the middle part of the distribution. We recall from Equation (16) that expectation of  $N_i$  is divided by 2 at every step so that small values of  $N_i$  and in particular 0 will have a larger probability. It should be noted that, due to the probability mass of small values of  $N_i$ , these mixture distributions, even if they are close to a normal distribution, will always have a higher than normal probability density around 0. To verify this intuition empirically, we also simulate 200,000 times the process for sizes  $n = 50$  and  $n = 100$ . The results do confirm our intuition. Indeed, for  $n = 50$ , the approximation is good for  $i = 1, 2, 3$  and starts failing for  $i = 4$ , where  $\mathbb{E}(N_4) = 3.125$ . For  $n = 100$ , we can of course go one step further, which makes the normal approximation good also for  $D(4)$ .



**Figure 1.** Simulated distribution of the  $D(i)$ ,  $i = 1, \dots, 6$  for a process of length  $n = 15$ . A normal distribution with mean 0 and variance set at the value calculated in Section 3 is plotted in red.

### 3.5. Simulation of the Completion Time

We have seen, in Section 3.3, Equations (31) and (32), how to approximate the average time until the end of the process. We test this result empirically by simulating 10,000 times the process of size  $n$  for  $n = 20, 40, \dots, 100, 200, 400, \dots, 1000, 2000, 4000, \dots, 30,000$ . For each  $n$ , we calculate the average time to complete the total number of exposures and the 95% Gaussian confidence interval given by adding or subtracting 1.96 times the empirical standard deviation of the sample. We also calculate the harmonic series and its approximation by a logarithm. The two approximations are almost identical and fit very well the results. Indeed, in 25 times, the 95% confidence interval misses once, which is what is expected. The approximations should however be expected not to be valid for very short processes because the logarithm approximation is based on convergence and because both the logarithm and harmonic series approximations are based on the approximation of a finite process by an infinite one, which is inaccurate for very short processes. In order to know how long a process must be for the approximation to be valid, we simulate 10,000 times the process of size  $n$  for  $n = 2, \dots, 20$ . From the results, displayed in Figure 2, we see that for  $n = 1, \dots, 6$ , the fit is not good. However, from  $n \geq 7$  onwards, our approximation is a reasonable and easy way to find the completion time. We shall use this result later when we build loss triangles.



**Figure 2.** Average run-off time of the process for short size processes.

#### 4. Capital Requirements

In this section, we propose a framework to test the accuracy of some of the methods to estimate the one year change. We construct triangles using the model presented in Section 3.4. From these triangles, we estimate the one year change using the classical Merz–Wüthrich method [Wüthrich and Merz \(2008\)](#) and the capital-over-time (COT) method, currently used at SCOR [Ferrero \(2016\)](#). We then compare the results obtained analytically and by simulation based on the specifications of our model to those estimated by the Merz–Wüthrich and with the capital-over-time (COT) method.

##### 4.1. Triangles

Until now, we have only considered the development properties of one process. Actual liability data are generally available in the form of triangles that represent the losses attributed to insurance contracts for each underwriting year after their current numbers of years of development. It is therefore reasonable to compare the results of our model in this framework. We first need to define the notation. Let us consider  $n$  representations of the process of size  $n$ . We denote their ultimate losses  $U_1(n), \dots, U_n(n)$ . The ultimate loss of the triangle is then

$$U(n) = \sum_{i=1}^n U_i(n).$$

We denote by  $N_{i,1}, \dots, N_{i,n}$  the number of exposures realized at each step of the  $i$ th process and  $S_{N_{i,j}}^{(j)}$  the losses due to  $N_{i,j}$ . The cumulated losses of row  $i$  up to column  $j$  are written  $S_i^{(j)}$ . We consider the discrete filtration  $(\mathcal{F}_k)_{0 \leq k \leq n}$  given by the information available at each calendar year. Formally,

$$\mathcal{F}_k = \sigma \left( N_{i,j}, S_{N_{i,j}}^{(j)} : i + j \leq n + 1 + k \right).$$

The individual filtrations of each representation of the process are denoted by

$$\mathcal{F}_{i,k} = \sigma \left( N_{i,j}, S_{N_{i,j}}^{(j)} : j \leq k \right).$$

We denote the  $j$ th one-year change of the  $i$ th individual process by  $D_i(j)$  and define the one-year changes of the triangle until its completion by

$$\begin{aligned} \Delta(k) &= \mathbb{E}(U(n) | \mathcal{F}_k) - \mathbb{E}(U(n) | \mathcal{F}_{k-1}) \\ &= D_{k+1}(n) + D_{k+2}(n-1) + \dots + D_n(k+1) \\ &= \sum_{i=k+1}^n D_i(n+k+1-i). \end{aligned} \quad (33)$$

Equation (33) simply writes the global one-year change as the sum of the individual one-year changes. Note that, due to linearity of expectation, Equation (33) also implies that the  $\Delta(i)$  have expectation 0 and are uncorrelated. Their variances can be calculated by summing the variances of the individual one-year changes that constitute them.

Similarly to the individual process situation, we define the capital requirement for calendar year  $i$  as

$$K_i = \mathbb{E} \left( \rho \left( \Delta(i) | \mathcal{F}_{i-1} \right) \right), \quad i = 1, \dots, n-1,$$

for some risk measure  $\rho$ . The risk-margin associated with these capitals is

$$R_n = \eta \sum_{i=1}^{n-1} K_i,$$

for the cost of capital  $\eta$ .

#### 4.2. Methodology and Results' Comparison

We now describe the methodology used to obtain our risk measure results and to compare them with the Merz–Wüthrich and the COT methods, both briefly explained in Appendix D. For convenience, we use the following notation

$$\tilde{N}_{i,j} = n - \sum_{k=1}^j N_{i,k}$$

for the number of exposures remaining to be realized for row  $i$  of the triangle after time  $j$ . In particular,  $N_{i,j} \sim \text{Unif}(\{0, 1, \dots, \tilde{N}_{i,j-1}\})$  for  $j = 1, \dots, n-1$ . In particular,  $\tilde{N}_{i,j}$  is an  $\mathcal{F}_{i,j}$ -measurable random variable.

##### 4.2.1. First Year Capital Comparison

A first value of interest is the required capital for the first year

$$K_1 = \text{TVaR}_\kappa(\Delta(1)|\mathcal{F}_0).$$

The Merz–Wüthrich method only provides  $\text{var}(\Delta(1)|\mathcal{F}_0)$ , while the COT method was originally designed for  $x\text{TVaR}$  (A4). An assumption concerning the link between  $\text{var}$  and  $\text{TVaR}$  is therefore to be made. Since  $\Delta(1)|\mathcal{F}_0$  follows a mixture of binomial distributions for a generally large number of rvs, its distribution can be approximated relatively well by a normal distribution (this approximation may lead up to 20% underestimation of the risk depending on the number of exposures  $n$ ). Note that the number of rvs from the binomial is not a uniform variable but a sum of uniform variables, which diminishes the probability of very large or very low values, thus making the normal approximation

better than for a simple uniform number of rvs. Normal distribution fixes the relation between var and TVaR:

$$\text{TVaR}_\kappa \left( \mathcal{N}(\mu, \sigma^2) \right) = \mu + \sigma \underbrace{\frac{\phi(\Phi^{-1}(\kappa))}{1 - \kappa}}_{\approx 2.665 \text{ for } 99\%}. \tag{34}$$

In our case,  $\mu = \mathbb{E}(\Delta(1)|\mathcal{F}_0) = 0$  and

$$\begin{aligned} \sigma^2 &= \text{var}(\Delta(1)|\mathcal{F}_0) \\ &= \mathbb{E}(\text{var}(\Delta(1)|N_{2,n}, \dots, N_{n,2})|\mathcal{F}_0) + \underbrace{\text{var}(\mathbb{E}(\Delta(1)|N_{2,n}, \dots, N_{n,2})|\mathcal{F}_0)}_{=0}) \\ &= \mathbb{E}((N_{2,n} + \dots + N_{n,2})p(1 - p)|\mathcal{F}_0) \\ &= \left( \tilde{N}_{2,n-1} + \sum_{j=3}^n \frac{\tilde{N}_{j,n+1-j}}{2} \right) \cdot p(1 - p). \end{aligned} \tag{35}$$

Combining Equations (34) and (35), we can obtain an approximate analytical value for  $K_1$ ,

$$K_1 \approx \frac{\phi(\Phi^{-1}(\kappa))}{1 - \kappa} \cdot (p(1 - p))^{1/2} \cdot \left( \tilde{N}_{2,n-1} + \sum_{j=3}^n \frac{\tilde{N}_{j,n+1-j}}{2} \right)^{1/2}. \tag{36}$$

The COT method, such as explained in Appendix D, was designed for real insurance data. In particular, parameter  $b$  models the dependence between relative loss increments. In the case of our model, the relative loss increments are uncorrelated, which points to the choice of parameter  $b = 0.5$  instead of the value chosen with mean time to payment. The choice of coefficient  $p_b$  is also arbitrary. Indeed,  $p_b$  determines the proportion of the risk that is due to the jump part of the process. For our process, there is no “special” type of behaviour that the model could have and that would increase the risk. Therefore, we choose  $p_b = 0$ . In general, the type of data is not known, in particular the dependency between loss increments is not known. Thus, we are also interested in the results given by the COT method applied the standard way. We therefore also compute the COT estimator with  $p_b$  chosen according to Formula (A5). In our case,  $b$  cannot be chosen like in the formula, as the pattern used in the mean time to payment computation is a paid pattern that we do not have for our model. For the jump case, we choose  $b = 0.75$  as for a long-tail process. Indeed, the (incurred) pattern of our  $n$ -step process corresponds generally to the type of patterns that one can find in long (or possibly medium) tail lines of business. We will refer to the two variations of the method as “COT method with jump part” for the version with standard  $p_b$  and  $b = 0.75$  and “COT method without jump part” for the version with  $b = 0.5$  and  $p_b = 0$ .

In the case of an  $n$ -step Bernoulli model triangle, we notice that the accident-year (incremental) patterns are given by

$$\gamma_i = \left( \sum_{j=2}^n \tilde{N}_{j,n+1-j} \right)^{-1} \cdot \left( \frac{\tilde{N}_{i+1,n-i}}{2^{i-1}} + \sum_{j=i+2}^n \frac{\tilde{N}_{j,n+1-j}}{2^i} \right). \tag{37}$$

The first factor is simply the total number of exposures remaining to be realized, the first row not being counted because it is finished. Since, at each step but the last, half of the current exposures of the process are expected to be realized, the number of exposures remaining for each unfinished is expected to be the number of exposures remaining in the original triangle divided by 2 for each past step that is not the last step. Hence, the expected remaining exposures are

$$\sum_{j=i+2}^n \frac{\tilde{N}_{j,n+1-j}}{2^i}$$

for the lines that are not finished after  $i$  steps, plus

$$\frac{\tilde{N}_{i+1,n-i}}{2^{i-1}}$$

for the line that finishes precisely after  $i$  steps. The pattern designates the results of the binomial random variable and not the number of exposures. However, since the rvs (random variable  $X$ ) are independent and have the same expectation, the numerator and denominator are both multiplied by  $p$  leaving the result unchanged.

In particular, with the help of Equation (37) and after few manipulations, we have that

$$\text{var}(\Delta(i)|\mathcal{F}_0) = \mathbb{E}(\text{var}(\Delta(i)|\mathcal{F}_{i-1})|\mathcal{F}_0) + \underbrace{\text{var}(\mathbb{E}(\Delta(i)|\mathcal{F}_{i-1})|\mathcal{F}_0)}_{=0} = \gamma_i \cdot \text{var}(U(n)|\mathcal{F}_0).$$

This result uses the property of martingales that the variance of the sum of martingale increments is equal to the sum of variances. An analogous property is false for the TVaR. However, we get for the COT model without jump part, an approximation (see [Ferriero 2016](#))

$$\text{TVaR}_\kappa(\Delta(1)|\mathcal{F}_0) \approx \gamma_1^{1/2} \cdot \text{TVaR}_\kappa(U(n)|\mathcal{F}_0).$$

Moreover, if we assume that the normal approximation is not an approximation but indeed an exact distribution, it can be shown through straightforward calculations that this expression becomes an exact result for the required capital for the first year (see [Ferriero 2016](#)).

The Merz–Wüthrich method is the one posing the most problems. Indeed, the triangles generated with our process are very noisy in the sense that the simulated triangles can quite often have many zeros. The Mack hypotheses, on which the Merz–Wüthrich method is based, are multiplicative in nature and, therefore, very sensitive to zeros. If there is a zero in the first column of a triangle, Mack’s estimation fails to compute the parameters  $\sigma_1$ . There are more robust ways, such as the one developed in [Busse et al. \(2010\)](#), to calculate these estimators. However, all of them (except removing the line) fail if an entire row of the triangle is 0. This happens quite often for  $n$  small. If  $n$  is large, the problem becomes, as we explain in Section 3.5, that the process terminates on average in time  $\log(n) + 1.57$ , which means that the largest part of the triangle shows no variation at all and gives  $\sigma_i = 0$ . In order to eliminate all these problems, we simulate our test triangles with  $n = 100,000$  and work on the truncated top side of the triangle of size  $m = \lceil 5 + \underbrace{\log(n) + 1.57}_{\approx \mathbb{E}(T_n)} \rceil = 19$ , where 5 is a safety margin to insure that the run-off of the process is finished or at least almost finished.

We set  $p = 0.1\%$ , which is a more realistic value given the high number of policies, simulate 500 triangles and, for each of them, calculate the first year capital  $K_1$  using the theoretical value, the COT method with and without jumps and the Merz–Wüthrich method. We display, in Table 1, the mean capital and the standard deviation of the capital around that mean over the 500 triangles. We also calculate the mean absolute deviation (MAD)

$$\text{MAD} = \mathbb{E}(|\hat{X} - X|)$$

and mean relative absolute deviation (MRAD)

$$\text{MRAD} = \mathbb{E}\left(\frac{|\hat{X} - X|}{X}\right)$$

with respect to the theoretical value using standard and robust mean estimation.

**Table 1.** Statistics for the first year capital on the 500 simulated triangles. The mean first year capital, the standard deviation of the capital around that mean and the mean absolute and relative deviations (MAD/MRAD) from the true value are displayed. The latter are computed using both a standard and the robust Huber M-estimator, Huber (1981). The mean reserves estimated with chain-ladder are 101.87, which are consistent with the reserves calculated with our model, i.e.,  $n(1 - 1/2^l)p = 100,000(1 - 2^{-19})0.001 = 100.00$ .

Method	Mean	Std. Dev.	MAD	MRAD	Rob. MAD	Rob. MRAD
Theoretical value	18.37	3.92	0	0%	0	0%
SCOR, without jumps	19.08	3.93	0.71	4.14%	0.71	3.93%
SCOR, with jumps	18.81	3.86	0.43	2.47%	0.44	2.42%
Merz–Wüthrich	252.89	149.6	234.5	1365.6%	213.9	1217.8%

Note that, in our example, the relative risk, i.e., the first year capital relative to the reserves volume, is about 18% (the reserves are approximately 100 and the first year capital approximately 18), which is a realistic value. The reserves in our model can be easily computed by the close formula  $n(1 - 1/2^m)p$ , as proved in Appendix C. As a point of comparison, using the prescription of the Solvency Standard Formula, we find a stand-alone capital intensity (SCR/Reserves) between 14% to 26% for the P&C reserves. Given the type of risks we are considering here, it is logical that the capital intensity should be at the lower end of the range. By the way, we also see, as expected, that the average claim is much smaller than the maximum claim (100,000) given the fact that the chances that 100,000 independent policies claim at the same time with such a low probability of claims ( $p = 1‰$ ) is practically nil.

The results presented in Table 1 are striking. While the COT method gives answers close to the theoretical value with a slight preference, as expected, for the COT without jumps, the Merz–Wüthrich method is way off (1356.6% off the true value), and the true result is not even within one standard deviation away. The coefficient of variation  $\sigma/\mu$  for this method is more than 59% while in all the other cases it hovers around 21%. There are many explanations for this. Looking at triangles and analysing the properties of the methods allows us to understand those results. The true risk depends on the number of rvs remaining to be realized. In most cases, only the few last underwriting years are truly important in that matter because the others will be almost fully developed. For Merz–Wüthrich, as most of the volatility of the process will appear on the first step, the most crucial part of the triangle is the last line of the triangle, which is the only process representation at this stage of development. If the latter is large, it influences the Merz–Wüthrich capital in the same direction. Merz–Wüthrich interprets a large value as: “Something happened on that accident year, there is going to be more to pay than expected”. The logic behind our model is different, through the “fixed number of rvs” property, it is: “What has been paid already needs not to be paid anymore”. A large value on the last line of the triangle is therefore likely to indicate that few rvs remain to be realized, which implies smaller remaining risk. This explains negative correlation because the same cause has the exact opposite effect on the result.

How can we explain the difference of magnitude in the estimated capitals? This may be due to the fact that our model is additive while Merz–Wüthrich is multiplicative. If a small number appears in the first column and then the situation reestablishes on the second step by realizing a larger number of exposures, we know that this is irrelevant for future risk. However, Mack and subsequently Merz–Wüthrich don’t consider the increase but the ratio. If the first value is small, the ratio may be large. However, the estimated ratio  $\hat{f}_j$  is the mean of the ratios weighted with the value of the first column, i.e., Equation (A2) can be rewritten as

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j} S_i^{(j)} \cdot \left( S_i^{(j+1)} / S_i^{(j)} \right)}{\sum_{i=1}^{n-j} S_i^{(j)}}.$$

Therefore, cases with large ratios, such as described before, will not appear in  $\hat{f}_1$  but in  $\hat{\sigma}_1^2$ . The Merz–Wüthrich (Mack) method considers that small and large values are as likely to be multiplied by a factor, which is not the case with our model for which small values are likely to be multiplied by large factors and large values are likely to be multiplied by small factors.

This is confirmed by plotting and comparing the distributions of the different capital measurements (Figure 3), we can notice that, while the true capital and the two COT capitals seem to follow a normal distribution, the distribution of the Merz–Wüthrich capital seems to follow rather a log-normal distribution.

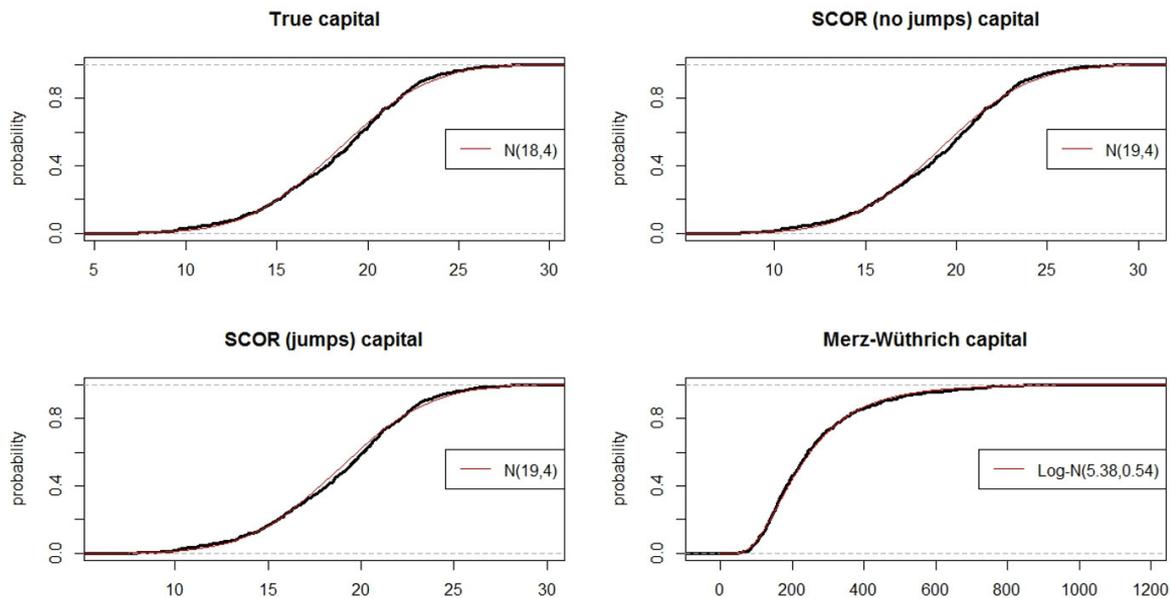


Figure 3. Simulated distribution of the capital for the different methods.

Another interesting statistic to understand how related these capital measurements are is the correlation between them. Computing the correlation matrix yields the results presented in Table 2, we see that the correlation is almost 100% for the two COT estimates and the true value. Indeed, with or without jumps, the COT method is very close to the theoretical result. This is partially due to the fact that the ultimate distribution is known and that all these methods simply multiply the ultimate risk by a constant. The correlation is not exactly 100% due to the stochasticity induced by the simulations used to calculate ultimate risk in the COT methods. The Merz–Wüthrich capital however shows a negative correlation. The standard and robust estimators are very different, which suggests the presence of very large values of Merz–Wüthrich capital and departure from normality. This is confirmed by Figure 3 where the distribution in the bottom right plot is very different from a Gaussian. It indicates in particular that the robust estimator is more representative of the data. A correlation of  $-46\%$  is rather strong. It is not true though that a small true capital implies a large Merz–Wüthrich capital, nor the opposite, but there is a real tendency among large values of true capital to coincide with relatively small values of Merz–Wüthrich capital.

**Table 2.** Correlation matrix of the different capital measures. Above, the standard correlation estimate and below, the robust “MVE” estimate Rousseeuw and Leroy (1987).

Standard Corr	True Value	SCOR, No Jumps	SCOR, Jumps	Merz–Wüthrich
True value	100%	99.98%	99.97%	−37.64%
SCOR, no jumps		100%	99.99%	−37.65%
SCOR, jumps			100%	−37.64%
Merz–Wüthrich				100%
MVE Corr	True Value	SCOR, No Jumps	SCOR, Jumps	Merz–Wüthrich
True value	100%	99.98%	99.97%	−46.56%
SCOR, no jumps		100%	99.99%	−46.61%
SCOR, jumps			100%	−46.60%
Merz–Wüthrich				100%

#### 4.2.2. Risk Margin Comparison

Another important quantity to study is the risk margin defined in Equation (21). We compare here the results obtained with the COT method described in Appendix D to those obtained from theory. Note that we cannot do this for the Merz–Wüthrich method as it is only giving the variation for the first year.

We want to estimate

$$K_i = \mathbb{E} [\text{TVaR}_\kappa(\Delta(i)|\mathcal{F}_{i-1})|\mathcal{F}_0] .$$

Our methodology is very close to the one for the first-year capital using normal approximation and Equations (34) and (35). Assume  $\mathcal{F}_{i-1}$  known, we can then generalise Equation (36), to get the following expression

$$\text{var}(\Delta(i)|\mathcal{F}_{i-1}) \approx \left( \tilde{N}_{i+1,n-1} + \sum_{j=i+2}^n \frac{\tilde{N}_{j,n+i-j}}{2} \right) \cdot p(1-p) . \tag{38}$$

We then use the normality assumption to write

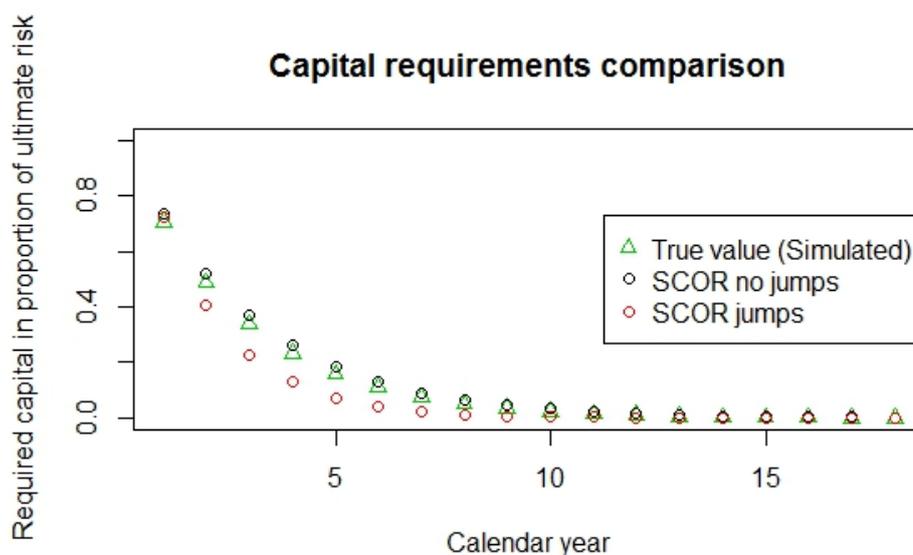
$$\text{TVaR}_\kappa(\Delta(i)|\mathcal{F}_{i-1}) = \frac{\phi(\Phi^{-1}(\kappa))}{1-\kappa} \sqrt{\text{var}(\Delta(i)|\mathcal{F}_{i-1})} , \tag{39}$$

thus obtaining a theoretical form for the tail value at risk, given the triangle developed up to calendar-year  $i - 1$ . Our methodology, starting from a triangle of realized rvs, is to complete it  $R$  times using the Bernoulli model and to calculate on each completed triangle  $\text{TVaR}_\kappa(\Delta(i)|\mathcal{F}_{i-1})$  according to the formula of Equation (39). By taking the mean over all  $R$  triangles, we obtain the required capital for calendar-year  $i$  that we sum up and multiply by the cost of capital (chosen here at 6%, as in the Solvency II directive) to obtain the risk margin.

In this case, we do not need to simulate truncated large triangles to make our comparison. Indeed, both the COT method and the theoretical simulation method work on small triangles. However, for the results to be similar and to avoid too frequent “zero risk left” situations, we still use a truncated large triangle like for the first-year capital comparison, i.e., triangles of size 19 and with  $n = 100,000$  rvs. Like for the first-year capital, we simulate 500 triangles from the process and, for each of them, calculate the capital required at each consecutive year and the risk margin using for the theoretical simulation method  $R = 10,000$  triangle completions and for the COT  $b = 0.5$  and  $p_b = 0$  (without jump part) and  $b = 0.75$  (long tail) and  $p_b$  from Equation (A5) (with “jump part”). In Table 3 and Figure 4, we can observe the results obtained on average and the measures of deviation over the 500 triangles.

**Table 3.** Statistics for the risk margin on the 500 simulated triangles. The average risk margin, the standard deviation of the risk margin around the average and the mean absolute and relative deviation (MAD/MRAD) from the true value are displayed.

Method	Mean	Std. Dev.	MAD	MRAD
True value (simulation)	5.89	1.27	0	0%
SCOR, without jumps	6.49	1.34	0.61	10.57%
SCOR, with jumps	4.32	0.89	1.57	26.52%



**Figure 4.** Comparison of the average yearly required capital over 500 triangles as proportion of the ultimate.

As we just saw, assuming normality, the first year capital of the COT method without jump part is an exact result. However, for  $i > 1$  (still assuming normality), the method gives

$$\mathbb{E}(\text{TVaR}_\kappa(\Delta(i)|\mathcal{F}_{i-1})|\mathcal{F}_0) = \frac{\phi(\Phi^{-1}(\kappa))}{1 - \kappa} \cdot \mathbb{E}\left(\sqrt{\text{var}(\Delta(i)|\mathcal{F}_{i-1})}\middle|\mathcal{F}_0\right). \tag{40}$$

However, from Schwarz inequality, which also holds for conditional expectation, for any positive integrable random variable  $Y$  and any  $\sigma$ -algebra  $\mathcal{F}$ , the COT method without jumps is systematically overestimating the true capital, as we can see in Figure 4. However, the overestimation is not very big (Table 3) and the method replicates reasonably well the form of the actual yearly capital. The average relative absolute error of the risk margin is 10.57% (see results in Table 3). We do not show the results for the capital at each year, but they lead to a similar message with the error increasing with the years as the capital itself decreases. The same method with jumps has less success with 26.52% of absolute error. This error is always underestimation, which is also true for each year. The error on the capital is always bigger with jumps. It is only at the end (calendar year 13 here) that the capital estimation is better with jumps and these values are almost 0, so they are not very relevant for the risk margin.

In general, one can see (Tables 1 and 3) that, for our  $n$ -steps model, the COT method without jumps is the one that performs the best. If we compute the autocorrelation of consecutive loss increments, we obtain 5% of mean correlation, which is close to independence. The independence situation corresponds to the calibration of the COT method with  $b = 0.5$ , thus explaining why the COT method without jumps provides the best results. This raises the question of what value of  $b$  would give the risk margin the closest to the benchmark. To answer this, we simulate another 100 triangles and calculate each time the risk margin with the benchmark method and with both COT methods with and without jumps, for all values of parameter  $b$  between 0.3 and 1 by steps of 0.01. For the COT method

without jumps, we find that the fitted values for  $b$  are between 0.52 and 0.53, which is very close to the  $b = 0.5$  that we have been using. For the COT method with jumps, the mean best  $b$  is also close to 0.5. However, we find some best  $b$  observations that are below 0.5, which stands for negative dependence between accident years and is not accepted by the COT method. In this case, the best  $b$  is much further than the one that has been used (0.75) by SCOR for real data. This is not unexpected since the method yields rather poor results (Table 3) for our  $n$ -steps model. (In the Appendix E we discuss the numerical stability of the above approximations. Furthermore, in the Appendix F we discuss the one-year capital for the first period as proportion to the sum of all the one-year capitals over all the periods.)

## 5. Conclusions

In this study, we have decomposed the various steps to reach the ultimate loss through a simple, but realistic example, which is used in a variety of line of business all over the World. The goal is to study the one year change required by the new risk based solvency regulations (Solvency II and the Swiss Solvency Test). Our example allows us to compute explicit analytical expressions for the variables of interest and thus test two methods used by actuaries to derive the one year change: the Merz–Wüthrich method and the COT method developed at SCOR. We find that the COT method is able to reproduce quite well the model properties while Merz–Wüthrich is not, even though the assumptions behind both methodologies are not satisfied in the case of our example (therefore the COT methodology is more robust). It is thus dangerous to use the Merz–Wüthrich method without making sure that its assumptions are met by the underlying data. Even though this seems obvious at first, the authors feel the need to warn about the risk of using the Merz–Wüthrich methodology acritically to any non-life portfolio, ignoring the fulfillment of the Merz–Wüthrich method’s assumptions, as this is the tendency among practitioners, regulators and auditors in the last years.

**Funding:** This research was funded by SCOR, General Guisan-Quai 26, 8022 Zurich, Switzerland.

**Acknowledgments:** The authors would like to thank Christoph Hummel for suggesting the binomial process and acknowledge his first contribution with a model of tossing a coin. The authors would also like to thank Arthur Charpentier for providing the R code of the Merz–Wüthrich method and for discussions. We also would like to thank anonymous referees for the useful remarks.

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

## Appendix A. Proof of an Equality in Distribution

In this appendix, we examine the equality in distribution formulated in Equation (3):

$$S_{N_k}^{(k)} = \sum_{j=1}^{N_k} X_{j+A_{k-1}} \stackrel{d}{=} \sum_{j=1}^{N_k} X_j = S_{N_k}.$$

The reason for it is that both sides are a sum of  $N_k$  Bernoulli random variables with the same parameter, so that no matter what the distribution of  $N_k$  is, both sides will have the same distribution. This can be proved rigorously by calculating both characteristic functions. The characteristic function of a random variable  $Y$  is defined as

$$\phi_Y(t) = \mathbb{E} \left( e^{itY} \right)$$

and determines uniquely the distribution of a function.

The characteristic function of  $S_{N_k}^{(k)}$  is

$$\begin{aligned} \phi_{S_{N_k}^{(k)}}(t) &= \mathbb{E} \left( e^{itS_{N_k}^{(k)}} \right) \quad \text{by definition of the characteristic function} \\ &= \mathbb{E} \left( \mathbb{E} \left( e^{itS_{N_k}^{(k)}} \mid N_k \right) \right) \quad \text{by definition of conditional expectation} \\ &= \sum_{j=0}^n \mathbb{E} \left( e^{itS_{N_k}^{(k)}} \mid N_k = j \right) \cdot \mathbb{P}(N_k = j) \quad \text{by developing the expectation} \\ &= \sum_{j=0}^n \mathbb{E} \left( e^{itS_j} \right) \cdot \mathbb{P}(N_k = j) \quad \text{because } X_j \perp N_k. \end{aligned}$$

The same calculations and arguments give the characteristic function of  $S_{N_k}$ :

$$\begin{aligned} \phi_{S_{N_k}}(t) &= \mathbb{E} \left( e^{itS_{N_k}} \right) \quad \text{by definition of the characteristic function} \\ &= \mathbb{E} \left( \mathbb{E} \left( e^{itS_{N_k}} \mid N_k \right) \right) \quad \text{by definition of conditional expectation} \\ &= \sum_{j=0}^n \mathbb{E} \left( e^{itS_{N_k}} \mid N_k = j \right) \cdot \mathbb{P}(N_k = j) \quad \text{by developing the expectation} \\ &= \sum_{j=0}^n \mathbb{E} \left( e^{itS_j} \right) \cdot \mathbb{P}(N_k = j) \quad \text{because } X_j \perp N_k \\ &= \phi_{S_{N_k}^{(k)}}(t), \end{aligned}$$

thus proving the equality in distribution.

The calculations can be pushed forward. Indeed,

$$\begin{aligned} \phi_{S_{N_k}}(t) &= \phi_{S_{N_k}^{(k)}}(t) = \sum_{j=0}^n \mathbb{E} \left( e^{itS_j} \right) \cdot \mathbb{P}(N_k = j) \\ &= \sum_{j=0}^n \mathbb{E} \left( e^{it \cdot \sum_{l=1}^j X_l} \right) \cdot \mathbb{P}(N_k = j) \\ &= \sum_{j=0}^n \mathbb{E}^j \left( e^{itX} \right) \cdot \mathbb{P}(N_k = j) \quad \text{because the } X_l \text{ are i.i.d,} \\ &= \sum_{j=0}^n \left( 1 - p + pe^{it} \right)^j \cdot \mathbb{P}(N_k = j) \quad \text{because } X \text{ is Bernoulli.} \end{aligned}$$

Note that, if  $k = 1$ , and  $N_1 \sim U(\{0, \dots, n\})$ , then we can write

$$\begin{aligned} \phi_{S_{N_1}}(t) &= \sum_{j=0}^n \left( 1 - p + pe^{it} \right)^j \cdot \mathbb{P}(N_1 = j) \\ &= \frac{1}{n+1} \sum_{j=0}^n \left( 1 - p + pe^{it} \right)^j \\ &= \begin{cases} \frac{1}{n+1} \frac{1 - (1 - p + pe^{it})^{n+1}}{p(1 - e^{it})}, & \text{if } t \neq 0, \\ \frac{1}{n+1} \sum_{j=0}^n 1 = 1, & \text{if } t = 0. \end{cases} \end{aligned}$$

We therefore have a close form for the characteristic function of  $S_{N_1}$ .

### Appendix B. The Distributions of $N_i$ and of $D(i)$

**Proposition A1.** *The distribution of  $N_i$ , for  $1 \leq i \leq n - 1$ , is*

$$\mathbb{P}[N_{n-1} = k] = \frac{1}{1+n} \sum_{a_{n-2}=0}^{n-k} \sum_{a_{n-3}=0}^{a_{n-2}} \cdots \sum_{a_1=0}^{a_2} \prod_{k=1}^{n-2} \frac{1}{1+n-a_k}, \tag{A1}$$

and  $\mathbb{P}[N_n = k] = \mathbb{P}[N_{n-1} = k]$ .

**Proof.** Indeed, the first value of the probability is

$$\mathbb{P}[N_1 = n_1] = \frac{1}{1+n}.$$

Then, we can write, for  $n = 2$ ,

$$\mathbb{P}[N_2 = n_2] = \sum_{a_1=0}^{n-n_2} \mathbb{P}[N_2 = n_2 \mid N_1 = a_1] \mathbb{P}[N_1 = a_1] = \frac{1}{1+n} \sum_{a_1=0}^{n-n_2} \frac{1}{1+n-a_1},$$

and, for  $n = 3$ ,

$$\mathbb{P}[N_3 = n_3] = \sum_{a_2=0}^{n-n_3} \mathbb{P}[N_3 = n_3 \mid A_2 = a_2] \mathbb{P}[A_2 = a_2] = \frac{1}{1+n} \sum_{a_2=0}^{n-n_3} \sum_{a_1=0}^{a_2} \frac{1}{1+n-a_2} \frac{1}{1+n-a_1}.$$

We can continue iteratively up to  $n - 1$  and obtain (A1).

We should note here that, because of Condition (H), for the last distribution, we have  $\mathbb{P}[N_n = k] = \mathbb{P}[N_{n-1} = k]$ . Alternatively, reminding that  $N_n + N_{n-1} = n - A_{n-2}$ , we come to the same conclusion.  $\square$

**Proposition A2.** *The distribution of  $D(i)$  conditioned to  $A_{i-1}$ ,  $2 \leq i \leq n$  is equal to*

$$\mathbb{P}(D(i) = x \mid A_{i-1} = a_{i-1}) = \frac{1}{1+n-a_{i-1}} \sum_{j=0}^{n-a_{i-1}} \binom{j}{x+jp} p^{x+jp} (1-p)^{j(1-p)-x} \mathbf{1}_{(x+jp \in \mathbb{N})}$$

for  $0 \leq a_{i-1} \leq n$ , and  $x \in \mathbb{R}$ .

**Proof.** The formula can be obtained directly from (14), by using (15) and that  $S_i$  has binomial distribution.  $\square$

### Appendix C. The Reserves in Our Model

For our model with only one row, i.e., one rv, the reserves at  $i < n$  years (steps) are simply

$$\sum_{j=i+1}^n \mathbb{E}(X_j) = \mathbb{E}(S_n) - \sum_{j=1}^i \mathbb{E}(X_j) = np \left( 1 - \sum_{j=1}^i \frac{1}{2^j} \right) = \frac{np}{2^i}.$$

In case we have a claims triangle with  $m$  rows and columns, where each row is our process with  $n$  exposures, the reserves at calendar year  $i$  are

$$\sum_{i=1}^m \sum_{j=i+1}^n \mathbb{E}(X_j) = np \sum_{i=1}^m \frac{1}{2^i} = np \left( 1 - \frac{1}{2^m} \right).$$

### Appendix D. Presentation of the Methods to Compute the One-Year Change Volatility

We briefly present here the methods we use in Section 4.2.

#### Appendix D.1. Merz–Wüthrich Method

The Merz–Wüthrich method [Wüthrich and Merz \(2008\)](#) is the most commonly used approach to model one-year change volatility. It is based on the distribution free method developed by [Mack \(1993\)](#) to estimate the ultimate uncertainty of the reserves on a triangle of claims. Using our notation, Mack makes the following assumptions:

1. Independence across rows of the triangle.
2. There exists a sequence of factors  $f_1, \dots, f_{n-1}$ , such that

$$\mathbb{E} \left( S_i^{(j+1)} \mid \mathcal{F}_{i,j} \right) = f_j S_i^{(j)}.$$

3. There exists a sequence of factors  $\sigma_1^2, \dots, \sigma_{n-1}^2$ , such that

$$\text{var} \left( S_i^{(j+1)} \mid \mathcal{F}_{i,j} \right) = \sigma_j^2 S_i^{(j)}.$$

Under these assumptions, Mack proposes the following unbiased estimators for the factors:

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j} S_i^{(j+1)}}{\sum_{i=1}^{n-j} S_i^{(j)}}, \quad j = 1, \dots, n - 1, \tag{A2}$$

and

$$\hat{\sigma}_j^2 = \frac{1}{k-j-1} \sum_{i=1}^{k-j} S_i^{(j)} \left( \frac{S_i^{(j+1)}}{S_i^{(j)}} - \hat{f}_j \right)^2, \quad j = 1, \dots, k - 2.$$

For  $j = n - 1$ , Mack uses  $\hat{\sigma}_{n-1}^2 = \min(\hat{\sigma}_{n-2}^2, \hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-2}^2 / \hat{\sigma}_{n-3}^2)$ . Furthermore, he uses the estimates  $\hat{f}_j$  to estimate the future parts of the triangle with the estimator

$$\hat{S}_i^{(j)} = S_i^{(n+1-i)} \cdot \hat{f}_{n+1-i} \cdots \hat{f}_{j-1}, \quad i + j > n + 1.$$

Based on these estimators, [Wüthrich and Merz \(2008\)](#), among other risk measures, calculate

$$\widehat{\text{var}}(D_i(n+2-i) \mid \mathcal{F}_{i,n+1-i}) = \left( \hat{S}_i^{(n)} \right)^2 \Psi_i,$$

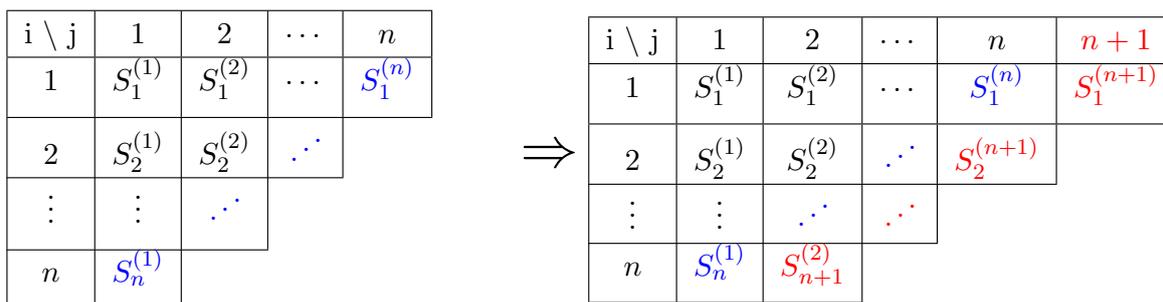
where

$$\Psi_i = \frac{\hat{\sigma}_{n+1-i}^2}{\hat{f}_{n+1-i}^2 S_i^{(n+1-i)}}.$$

Summing these conditional variances, they obtain an estimator for the first one year change:

$$\widehat{\text{var}}(\Delta(1) \mid \mathcal{F}_0).$$

The idea behind the Merz–Wüthrich method is to calculate the Chain–Ladder estimation of the ultimate uncertainty at time 0 and at time 1 after adding the next diagonal. They then calculate the uncertainty of the difference to obtain the one-year uncertainty. We illustrate this in [Figure A1](#) by showing the addition of one diagonal.



**Figure A1.** The current triangle (left) and the next year triangle (right). The present is outlined in blue and the one-year ahead future in red. The Merz–Wüthrich method allows for computing the ultimate risk on both triangles and calculate the uncertainty of the difference.

Appendix D.2. The COT Method

The COT formula used at SCOR [Ferriero \(2016\)](#) consists of computing first the ultimate risk, in our case, the TVaR<sub>99%</sub> of a  $\mathcal{B}(\tilde{N}, p)$  distribution, and taking a part of it, to be determined, as the required capital for each year. The idea behind the COT formula is to look at the evolution of the risk over time till the ultimate, and thus obtain the one-year period risks as a portion of the ultimate risk. In what follows, we give a brief description of the COT formula; however, the detailed derivation is presented in [Ferriero \(2016\)](#). Here,  $\tilde{N}$  designates the number of exposures that remain to be included in the whole triangle. This can be written as

$$K_i = \delta_i \cdot \rho_\kappa(U(n) \mid \mathcal{F}_0), \tag{A3}$$

where

$$\rho_\kappa(X) = x\text{TVaR}_\kappa(X) = \text{TVaR}_\kappa(X) - \mathbb{E}(X), \quad \kappa = 99\%. \tag{A4}$$

The vector  $\delta = (\delta_1, \dots, \delta_{n-1})$  is called the COT-pattern and is obtained through the following relation:

$$\delta_k = \gamma_k^b(1 - p_b) + p_b \sum_{i=k}^n \gamma_i,$$

where  $\gamma = (\gamma_1, \dots, \gamma_{n-1})$  designates the incremental calendar year pattern,

$$p_b = (1 - \lambda) \left[ 1 - \frac{\text{VaR}_{1-\alpha}(U(n)) - \mathbb{E}(U(n))}{\rho_\kappa(U(n))} \right], \tag{A5}$$

and parameters are fixed to the values  $\lambda = x^{\frac{-(1-b)}{1-\alpha}}$ ,  $x = 10$ ,  $\alpha = 10\%$ . Parameter  $b$  is set to 0.6 for short-tailed Lines of Business, 0.65 for medium-tailed LoB and 0.75 for long-tailed LoB, where the duration is determined by the notion of mean time to payment  $\tau$ . It is defined as

$$\tau = \sum_{i=1}^n \tilde{\gamma}_i(i - 0.5) = \sum_{i=1}^n i\tilde{\gamma}_i - \frac{1}{2},$$

where  $\tilde{\gamma}$  designates the incremental calendar year paid pattern. A short-tail is defined as  $\tau \leq 2$ , medium-tail by  $\tau \in [2, 4]$  and long-tail by  $\tau > 4$ . Once we have the  $K_i$ , we can sum them and multiply them by the cost of capital to obtain the risk margin

$$\begin{aligned} R_n &= \eta \left( \sum_{k=1}^n \delta_k \right) \cdot \rho_\kappa(U(n)) \\ &= \eta \left( (1 - p_b) \sum_{k=1}^n \gamma_k^b + p_b \sum_{k=1}^n \sum_{i=k}^n \gamma_i \right) \cdot \rho_\kappa(U(n)) \\ &= \eta \left( (1 - p_b) \sum_{k=1}^n \gamma_k^b + p_b \sum_{k=1}^n k\gamma_k \right) \cdot \rho_\kappa(U(n)). \end{aligned} \tag{A6}$$

In short, this model is based on the idea that claims will develop partially with a “good” continuous part and partially according to a “bad” part characterized by sudden jumps, the bad part being modelled as the total rest of the claims realising at once. The variable  $p_b$  is a coefficient between  $[0, 1]$  that determines in which proportion the evolution is going to be continuous or discrete and  $b \in [0.5, 1]$  models the dependence between different calendar years.

In order for the COT method to be exact, the following assumptions must be true:

1. The evolutions of the claims losses and of the best estimates are stochastic processes as described in [Ferriero \(2016\)](#); roughly speaking, the relative losses evolve from the start to the end as a Brownian motion, except during a random time interval in which they evolve as a fractional Brownian motion, and the consequently best estimates evolve as the conditional expectation of the ultimate loss plus a sudden reserves jump, which may happen as a result of systematic under-estimations of the losses.
2. The volatility, measured in standard deviations, of the attritional claims losses is small relative to the ultimate loss size.

However, the COT method is robust in the sense that gives good estimates even when the assumptions are not fulfilled as we show here with our example.

### Appendix E. Numerical Stability

In Section 4.2, we calculate the risk margin by simulating triangle completions according to our  $n$ -steps model. It is therefore legitimate to ask if the  $R = 10,000$  simulations we use are enough to obtain stable results. To investigate this question, we simulate an  $n$ -steps model triangle. We then calculate its risk margin 200 times using our simulation method with a grid of values of  $R$ . From these, we calculate the mean and the standard deviation of the computed risk margins for each value of  $R$ . The distribution of the calculated risk margin being approximatively normal due to the central limit theorem, the mean and standard deviation fully characterize the distribution allowing us in particular to draw confidence intervals. We chose for this test  $R \in \{10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10,000, 20,000\}$  and obtained the values presented in Table A1.

The results seem to indicate that the standard deviation of the risk margin calculation is inversely proportional to  $\sqrt{R}$  as one would expect. The mean prediction is almost the same no matter what the number of simulations is even though the variation of this mean around 3.28 diminishes as  $R$  grows. The value  $R = 10,000$  that we have used seems in any case sufficient as, for this value, the 95% confidence interval represents only a variation of  $\pm 0.34\%$  around the mean.

**Table A1.** Test of the number  $R$  of random triangle completions for the risk margin calculation. The mean and standard deviation allow for constructing a Gaussian 95% confidence interval by adding (resp. subtracting) 1.96 times the standard deviation to the mean to obtain the upper (resp. lower) bound. The “Variation” column designates the variation around the mean that represents the 95% confidence interval, i.e., 1.96 times the standard deviation.

$R$	Mean	Standard Dev.	Confidence Interval	Variation
10	3.279	0.176	[2.935, 3.624]	$\pm 10.50\%$
20	3.294	0.122	[3.055, 3.533]	$\pm 7.28\%$
50	3.287	0.077	[3.136, 3.438]	$\pm 4.60\%$
100	3.291	0.057	[3.180, 3.402]	$\pm 3.38\%$
200	3.289	0.042	[3.206, 3.372]	$\pm 2.53\%$
500	3.281	0.027	[3.228, 3.333]	$\pm 1.60\%$
1000	3.284	0.018	[3.248, 3.320]	$\pm 1.09\%$
2000	3.283	0.013	[3.258, 3.308]	$\pm 0.76\%$
5000	3.282	0.009	[3.265, 3.300]	$\pm 0.53\%$
10,000	3.283	0.006	[3.272, 3.294]	$\pm 0.34\%$
20,000	3.283	0.004	[3.275, 3.291]	$\pm 0.24\%$

### Appendix F. Capital Properties of the Model

In this section, we discuss the properties of the ratio first year capital divided by the sum of all required capitals. That is the ratio

$$\frac{K_1}{\sum_{j=1}^{m-1} K_j}$$

where  $m$  represents the size of the triangle. We would like to know in particular how a variation of the parameter  $n$  will influence that ratio. To investigate this question, we simulate 100 triangles with  $n$  exposures per line, for values of  $n$  in the set  $\{k \cdot 10^4 : k = 1, \dots, 10\}$ . Like for the previous capital requirement calculations, we truncate the triangles to size  $m = \lceil 5 + \log(n) + 1.57 \rceil = 19$ . This avoids considering triangles of very large sizes without changing the result, as, in most cases, almost all exposures will have been realized before this step. For each triangle, we compute the capital requirements using 2000 random triangle completions, as the study of Appendix E has shown that this is sufficient to have less than 1% error. For convenience reasons, we only show some statistics of the results that are representative of the spread of the whole sample (Table A2), the mean, standard deviation and the two extremal values observed.

**Table A2.** Statistics of the proportion of the capital represented by the first year as a function of the number of rvs  $n$ . Note that the number of rvs modifies the number of steps  $l = \lceil 5 + \log(n) + 1.57 \rceil$ .

Number of rvs	Number of Steps	Mean	Standard Dev.	Min. Obs.	Max. Obs.
10,000	16	0.3137	0.00395	0.3067	0.3267
20,000	17	0.3129	0.00472	0.3049	0.3286
30,000	17	0.3130	0.00444	0.3056	0.3278
40,000	18	0.3125	0.00372	0.3064	0.3257
50,000	18	0.3129	0.00465	0.3040	0.3244
60,000	18	0.3123	0.00412	0.3041	0.3240
70,000	18	0.3123	0.00422	0.3057	0.3279
80,000	18	0.3118	0.00440	0.3039	0.3264
90,000	18	0.3125	0.00465	0.3041	0.3284
100,000	19	0.3118	0.00383	0.3033	0.3231

The results are quite independent of the number of rvs and show no obvious pattern of development. The size of the changes between different values of  $n$  is much smaller than the standard deviation indicating that  $n$  has no (or non-significant) effect on the ratio of interest. The standard deviation shows no sign of correlation to the number of realized rvs per line, and the maximal observation is slightly more volatile than the minimal. However, both are very stable, giving no indication that  $n$  might have any significant effect on the ratio.

Let us give some intuition behind these results. In our model, with the exception of the move from penultimate column to ultimate, for which all not yet realized exposures are forced to be realized, each move forward in the triangle means, in expectation, dividing by two the number of remaining rvs to be realized. Since we use 2000 triangle completion, we can assume that we are at the expectation. The variance is proportional to the number of rvs remaining. This means that the TVaR, which is proportional to the standard deviation (under normality assumption), is proportional to the square root of the same number. The crucial number in calculating capital is therefore the expectation of the square root of the number of rvs as described in Equations (38) and (40). The expectation of the square root is divided every calendar year by a factor that is almost the same, except near the end. This factor depends on the triangle, unlike the square root of the expected variance (the square root of the expected variance is approximately divided by  $\sqrt{2}$  at every step). However, it is generally close to 0.69. Therefore, the ratio is approximately

$$\frac{K_1}{\sum_{j=1}^{m-1} K_j} \approx \frac{K_1}{\sum_{j=1}^{m-1} K_1 \cdot 0.69^{j-1}} \approx \frac{1}{\sum_{j=1}^{\infty} 0.69^{j-1}} = 1 - 0.69 = 0.31.$$

This is obviously not an exact result but shows that this quantity has a small sensitivity to  $m$ ,  $p$ , and gives an insight into why the results are so similar every time.

## References

- Bornhuetter, Ronald L., and Ronald E. Ferguson. 1972. The Actuary and IBNR. *Proceedings of the Casualty Actuarial Society* 59: 181–95.
- Busse, Marc, Michel M. Dacorogna, and Marie Kratz. 2013. Does Risk Diversification Always Work? The Answer through Simple Modelling. SCOR Paper No 24. Available online: <http://www.scor.com/en/sgrc/scor-publications/scor-papers.html> (accessed on 10 May 2013).
- Busse, Marc, Ulrich Müller, and Michel Dacorogna. 2010. Robust estimation of reserve risk. *Astin Bulletin* 40: 453–89.
- Dacorogna, Michel M., and Christoph Hummel. 2008. *Alea Jacta Est, An Illustrative Example of Pricing Risk*. SCOR Technical Newsletter. Zurich: SCOR Global P&C.
- Ferriero, Alessandro. 2016. Solvency capital estimation, reserving cycle and ultimate risk. *Insurance: Mathematics and Economics* 68: 162–68. [[CrossRef](#)]
- Huber, Peter J. 1981. *Robust Statistics*. Chichester: Wiley.
- Mack, Thomas. 1993. Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin* 23: 213–55. [[CrossRef](#)]
- Mack, Thomas. 2008. The Prediction Error of Bornhuetter-Ferguson. *Astin Bulletin* 38: 87–103. [[CrossRef](#)]
- Merz, Michael, and Mario V. Wüthrich. 2008. Modelling the claims development result for solvency purposes. *Casualty Actuarial Society* 542–68.
- Mikosch, Thomas. 2004. *Non-Life Insurance Mathematics: An Introduction with Stochastic Processes*. Berlin: Springer Verlag.
- Rousseeuw, Peter J., and Annick M. Leroy. 1987. *Robust Regression and Outlier Detection*. New York: Wiley.
- Schmidt, Klaus D., and Mathias Zocher. 2008. The Bornhuetter-Ferguson principle. *Variance* 2: 85–110.
- Wüthrich, Mario V., and Michael Merz. 2008. *Stochastic Claims Reserving Methods in Insurance*. Chichester: Wiley-Finance.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).