Astin Bulletin (2026), **56**, pp. 317–343
doi:10.1017/asb.2026.10087

International Actuarial Association
Association Actuarielle Internationale

RESEARCH ARTICLE

On a risk model with tree-structured Poisson Markov random field frequency, with application to rainfall events

Hélène Cossette[1], Benjamin Côté[2], Alexandre Dubeau[1] and Etienne Marceau[1]

[1]École d'actuariat, Université Laval, Canada
[2]University of Waterloo, Canada
Corresponding author: Etienne Marceau; Email: etienne.marceau@act.ulaval.ca

Received: 26 September 2025; **Revised:** 27 January 2026; **Accepted:** 28 January 2026; **First published online:** 23 March 2026

Keywords: Undirected graphical models; multivariate Poisson distribution; risk sharing; weak convergence; Bethe lattice

Abstract

In many insurance contexts, dependence between risks of a portfolio may arise from their frequencies. We investigate a dependent risk model in which we assume the vector of count variables to be a tree-structured Markov random field with Poisson marginals. The tree structure translates into a wide variety of dependence schemes. We study the global risk of the portfolio and the risk allocation to all its constituents. We provide asymptotic results for portfolios defined on infinitely growing trees. To illustrate its flexibility and computational scalability to higher dimensions, we calibrate the risk model on real-world extreme rainfall data and perform a risk analysis.

1. Introduction

Compound Poisson distributions serve as a basis for building risk models with applications in property and casualty insurance, such as risk management, pricing, and reserves. In these contexts, the loss X_v for a given risk v is often assumed to follow a compound Poisson distribution, and the portfolio's aggregate loss is thereby defined as

$$S = X_1 + X_2 + \cdots + X_d, \quad \text{with} \quad X_v = \sum_{j=1}^{N_v} B_{v,j}, \quad \text{for every} \quad v \in \mathcal{V} = \{1, \ldots, d\}, \quad d \in \mathbb{N}_1 = \mathbb{N}\backslash\{0\}, \quad (1.1)$$

where $N_v \sim \text{Poisson}(\lambda_v)$ is referred to as the risk's frequency and $\{B_{v,j}, j \in \mathbb{N}_1\}$ as its sequence of severities, with the convention $\sum_{j=1}^{0} B_{v,j} = 0$.

Dependence between risks of a portfolio may arise from their frequencies. The model in (1.1) has the advantage, as discussed in Cummins and Wiltbank (1983), of allowing for proper accommodation of this dependence while explicitly accounting for events of different sources, which may have distinct claim amount distributions. Mathematically, this translates into components of the frequency random vector $\boldsymbol{N} = (N_v,\ v \in \mathcal{V})$ being dependent in (1.1). In this paper, the sequences of claim amounts $\{B_{1,j},\ j \in \mathbb{N}_1\}, \ldots, \{B_{d,j},\ j \in \mathbb{N}_1\}$ are assumed to be mutually independent and independent of $\boldsymbol{N}$. Let the random variables within each sequence be identically distributed, and thus we may refer to claim amounts with stand-in random variables $B_1, \ldots, B_d$ for convenience. The portfolio $\boldsymbol{X} = (X_v,\ v \in \mathcal{V})$ thus follows a multivariate compound distribution of Type 2 according to the terminology in Sundt and Vernic (2009); see Chapters 19–20. In a risk modeling setting, multivariate compound Poisson distributions of Type 2 have been studied notably in Cossette *et al*. (2012) and Kim *et al*. (2019).

One may rely on three approaches to conceive a multivariate Poisson distribution for the random vector $\boldsymbol{N}$: copulas, common shocks, and binomial thinning; see for instance Inouye *et al*. (2017) and Liu *et al*. (2024). The copula approach allows separate modeling of marginals and dependence but

faces theoretical and computational challenges in a discrete setting (Genest and Nešlehová, 2007; Henn, 2022). The common-shock approach, dating back to M'Kendrick (1925) and later extended to higher dimensions (Krishnamoorthy, 1951; Teicher, 1954), offers a clear stochastic interpretation but quickly becomes intractable due to the exponential growth in parameters (Karlis, 2003). The family of common-shock-based models, while widely referred to as *multivariate Poisson*, does not encompass all Poisson-marginal distributions; see Çekyay *et al*. (2023) for a historical recap.

The third approach is to rely on stochastic representations employing binomial thinning. Binomial thinning was introduced in Steutel *et al*. (1983) and first employed to incorporate dependence between Poisson random variables in McKenzie (1985) and McKenzie (1988). Such an approach has been used for risk modeling in Yuen and Wang (2002), Lindskog and McNeil (2003), and Wang and Yuen (2005). In Côté *et al*. (2025b), the authors encapsulate binomial thinning operations within a tree structure to provide a much wider variety of dependence schemes under this approach. The resulting tree-structured Markov random field (MRF) has explicit probability mass function (pmf) and probability generating function (pgf) expressions, enabling efficient computation in high dimensions.

In a risk modeling context, scalability of computations and estimation methods to high dimensions is of the most important consideration. In particular, to introduce dependence between Poisson risk frequencies, the current methods based on common shocks are either little flexible or computationally intensive. This work draws inspiration from the MRF in Côté *et al*. (2025b) to reconcile flexibility and scalability without recourse to copulas. In this paper, the authors' main focus is the MRF's distributional properties. Here, we put forth its relevance for risk modeling. To do so, we begin by allowing more flexibility for the modeling of marginals by letting the MRF's means be heterogeneous, despite a slight trade-off in the disconnect between marginal and dependence. We also show the MRF's connection to the multivariate Poisson and discuss methods for the model's estimation.

We put forth the MRF's relevance for risk modeling through two objectives. One of our objectives is to highlight the computational methods' practicality and their applicability to actuarial science. We will discuss this through two tasks. First, we aim to evaluate the aggregate risk of the portfolio by studying the distribution of S in (1.1) and developing efficient methods to evaluate its pmf without resorting to approximations. Second, we aim to assess the contribution of every component of the portfolio $\boldsymbol{X}$ to the aggregate claim amount, and we perform this risk allocation twofold. For an allocation *ex ante*, we resort to the contribution to the Tail Value-at-Risk (TVaR) under Euler's principle, see Tasche (2007); for an allocation *ex post*, we turn to conditional-mean risk sharing, see Denuit and Dhaene (2012) and subsequent work. Algorithms for its exact computation are developed, inspired from the methods put forth in Blier-Wong *et al*. (2025). We furthermore provide results allowing for a better understanding of the portfolio's asymptotic behavior by defining the model on infinite-dimensional trees.

Another objective is to illustrate the practical relevance and effectiveness of the proposed risk model in real-world applications. Notably, from Theorem 7.1 of Coles (2001), counts of extreme events follow Poisson distributions. This insight provides a natural avenue for applying our model. We perform a detailed risk management study on extreme rainfall events data, in which we evaluate tail risk measures (e.g., TVaR) and assess the allocation of extreme losses across different station locations. Our study showcases the risk model's ability to capture dependence structures in multivariate extreme events while maintaining interpretability. This paper contributes to the growing use of graphical models in actuarial science, including Oberoi *et al*. (2020), Denuit and Robert (2022), and Boucher *et al*. (2024).

The structure of the paper is as follows. In Section 2, we present the tree-structured MRF with Poisson marginal distributions and discuss its connection with the distributions obtained through the common-shock approach. In Section 3, we perform our first risk management task, evaluating the risk associated with S. In Section 4, we perform our second risk management task, allocating that risk to the components of $\boldsymbol{X}$; we also provide results for asymptotic cases of infinitely large portfolios. In Section 5, we discuss the MRF's parameters' estimation. Section 6 comprises the application to extreme rainfall data. Section 7 provides concluding remarks. All proofs are relegated to Appendix A.

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

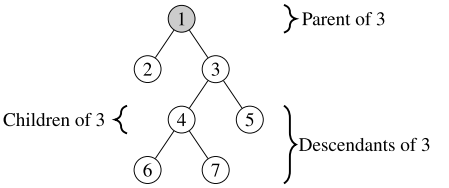


Figure 1. *Filial relations in a rooted tree.*

2. Tree-structured MRFs with Poisson marginal distributions

In this section, we present the tree-structured MRF with Poisson marginal distributions, which will be the center of consideration in the following sections. Distributions of this family describe tree-structured MRFs. We recall below the definition of a MRF (Cressie and Wikle, 2015, Chapter 4.2) and some elementary notions pertaining to trees. Let $\mathcal{V} = \{1, 2, \ldots, d\}$, with $d \in \mathbb{N}_1$, represent a set of vertices, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ be a set of edges.

Definition 2.1 (MRF). *A vector of random variables $\boldsymbol{X} = (X_v, v \in \mathcal{V})$ is a MRF if it satisfies the local Markov property with respect to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$; that is, for any two of its components, say X_u and X_w, such that $(u, w) \notin \mathcal{E}$,*

$$X_u \perp\!\!\!\perp X_w \mid \{X_j,\ (u, j) \in \mathcal{E}\}, \quad u, w \in \mathcal{V}, \tag{2.1}$$

where $\perp\!\!\!\perp$ denotes conditional independence. A MRF is tree-structured if its underlying graph is a tree.

A tree, denoted by $\mathcal{T}$, is a simple and connected undirected graph such that no path from a vertex to itself exists. A path from vertex u to vertex v, written path(u, v), is the set of edges $e \in \mathcal{E}$ such that u and v participate in an edge once and any other involved vertices twice. All graphs considered in this paper are trees. Labeling a specific vertex $r \in \mathcal{V}$ as the root of a tree, we define $\mathcal{T}_r$ as the r-rooted version of $\mathcal{T}$. A root for the tree defines a parent pa(v) (except for the root), children ch(v), and descendants dsc(v) for each $v \in \mathcal{V}$. An example of this notation is provided in Figure 1, where vertex 1 acts as the tree's root. We refer to Section 3.3 of Saoub (2021) for further insight on the terminology surrounding rooted trees.

By encrypting a MRF on a tree, the random vector's dependence scheme inherits the structural properties of the latter. For instance, the correlation between two of its components decays multiplicatively along the path between them. Although constraining feasible dependence schemes, this comes with the benefit of a greater interpretability: this is the idea underlying probabilistic graphical models.

The construction of tree-structured MRFs with Poisson marginal distributions relies on the binomial thinning operator, denoted by $\circ$, defined in terms of a random variable Y taking values in $\mathbb{N}$ as $\theta \circ Y := \sum_{j=1}^{Y} I_j^{(\theta)},\ \theta \in [0, 1]$, where $\{I_j^{(\theta)}, j \in \mathbb{N}_1\}$ is a sequence of independent Bernoulli random variables with a probability of success θ; see Steutel *et al*. (1983). This operation can be interpreted as sampling among Y elements, each with probability θ of being selected independently. We refer the interested reader to Weiß (2008) for further insight on the binomial thinning operator.

2.1. Main characteristics of tree-structured MRFs with Poisson marginal distributions

To construct a risk model with heterogeneous marginal behaviors in Section 3, we define tree-structured MRFs with Poisson marginals where each component has its own mean parameter λ_v, $v \in \mathcal{V}$.

Theorem 2.2. *Consider a tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$, and let $\mathcal{T}_r$ be its rooted version, for some $r \in \mathcal{V}$. Given a vector of mean parameters $\boldsymbol{\lambda} = (\lambda_v,\ v \in \mathcal{V})$ where $\lambda_v > 0$ for every $v \in \mathcal{V}$ and a vector of dependence parameters $\boldsymbol{\alpha} = (\alpha_e,\ e \in \mathcal{E})$ where $\alpha_{(\mathrm{pa}(v),v)} \in (0, \min(\sqrt{\lambda_v/\lambda_{\mathrm{pa}(v)}}, \sqrt{\lambda_{\mathrm{pa}(v)}/\lambda_v})]$ for every $(\mathrm{pa}(v), v) \in \mathcal{E}$. Let $\boldsymbol{L} = (L_v,\ v \in \mathcal{V})$ be a vector of independent random variables such that $L_v \sim$ Poisson$(\lambda_v(1 - \alpha_{(\mathrm{pa}(v),v)}\sqrt{\lambda_{\mathrm{pa}(v)}/\lambda_v}))$ for every $v \in \mathcal{V}$, with $\alpha_{(\mathrm{pa}(r),r)} = 0$ since the root has no parent. Define $\boldsymbol{N} = (N_v,\ v \in \mathcal{V})$ as a vector of random*

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

variables such that

$$N_v = \begin{cases} L_r, & \text{if } v = r \\ \left(\alpha_{(\mathrm{pa}(v),v)}\sqrt{\frac{\lambda_v}{\lambda_{\mathrm{pa}(v)}}}\right) \circ N_{\mathrm{pa}(v)} + L_v, & \text{if } v \in \mathrm{dsc}(r) \end{cases}, \quad \text{for every } v \in \mathcal{V}. \tag{2.2}$$

Then, $\boldsymbol{N}$ is a MRF with a unique joint distribution whichever the chosen root of $\mathcal{T}$, where the random variable N_v follows a Poisson distribution of parameter λ_v, for all $v \in \mathcal{V}$.

Henceforth, we write $\boldsymbol{N} \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T})$ to signify $\boldsymbol{N}$ admits the stochastic representation in (2.2), and we let $\boldsymbol{\Lambda}$ denote the set of admissible parameters $(\boldsymbol{\lambda}, \boldsymbol{\alpha})$. We write $\mathbb{MPMRF}$ the family of all such distributions for $\boldsymbol{N}$.

The MRF studied in Côté *et al*. (2025b) is a special case of the MRF constructed in Theorem 2.2, where the mean parameters are homogeneous. In (2.2), the components of $\boldsymbol{N}$, except the root, are defined as the sum of two independent random variables. We interpret them as the *propagation* and the *innovation* random variables, respectively. The propagation random variable $(\alpha_{(\mathrm{pa}(v),v)}\sqrt{\lambda_v/\lambda_{\mathrm{pa}(v)}}) \circ N_{\mathrm{pa}(v)}$ expresses the number of events that have duplicated by binomial thinning from $N_{\mathrm{pa}(v)}$ to N_v. The innovation random variable L_v expresses the number of events occurring at vertex v that have not propagated from vertex $\mathrm{pa}(v)$. The thinning parameter $\alpha_{(\mathrm{pa}(v),v)}\sqrt{\lambda_v/\lambda_{\mathrm{pa}(v)}}$ for the propagation random variable is an adjustment of the dependence parameter $\alpha_{(\mathrm{pa}(v),v)}$ taking into account the heterogeneous means, so that the correlation between $N_{\mathrm{pa}(v)}$ and N_v is $\alpha_{(\mathrm{pa}(v),v)}$. The upper bound $\alpha_{(\mathrm{pa}(v),v)} < \min\left(\sqrt{\lambda_v/\lambda_{\mathrm{pa}(v)}}, \sqrt{\lambda_{\mathrm{pa}(v)}/\lambda_v}\right)$ stems from the fact that dependence is characterized by this thinning mechanism. The fraction of events propagating from one random variable to another cannot exceed the root of the ratio of the marginal rates; otherwise, the innovation component would require more inherited events than the number of events possible on the parent. The admissible parameter space therefore reflects a structural feature of event-sharing mechanisms in count data. Note that any vertex can be chosen as the root of the tree; accordingly, the constraint on the dependence parameters takes into account the reversibility of parent–child relationships that may occur, thus preserving the root-free formulation of the distribution; see Remark 2.3. For illustration, if $(\lambda_1, \lambda_2) = (5, 5)$, the upper bound is 1, maximal correlation. If $(\lambda_1, \lambda_2) = (3, 5)$, the upper bound is $\sqrt{0.6} \approx 0.77$, compared to the Fréchet upper bound of ≈ 0.97.

Remark 2.3. As put forth in the proof of Theorem 2.2, the symmetry of pairwise stochastic dynamics, combined with the local Markov property, yields a root-free distribution of the MRF. The directionality seemingly implied by the necessity to choose a root is illusory: to define a parent–child dynamic is necessary to have a binomial thinning representation, but it has no impact on the joint distribution of $\boldsymbol{N}$. The model is thus truly undirected; we will provide in Section 2.2 a root-free stochastic construction. The artificial directionality, required for the binomial thinning representation, is what allows an iterative way of proceeding along trees, resulting in algorithmic efficiency. A similar situation occurs for tree-structured Ising models; see Côté et al. (2025a).

Remark 2.4. Compared to the homogeneous-mean model in Côté et al. (2025b), $\boldsymbol{\alpha}$ is not a true dependence parameter, as it is constrained by $\boldsymbol{\lambda}$. To circumvent this, one could let $\theta_v = \alpha_{(\mathrm{pa}(v),v)}\sqrt{\lambda_v/\lambda_{\mathrm{pa}(v)}}$, and $\boldsymbol{\theta}$ would be a true dependence parameter, ranging from 0 to 1. The values of $\boldsymbol{\theta}$ would however vary with the root, so this parameter change would come at the cost of the undirectionality of the model. Also note that, because the constraints on $\boldsymbol{\alpha}$ come directly from the binomial thinning operations, they should not be limiting if the stochastic dynamics underlying the data are effectively of propagation and innovation.

The following sequence of joint pgfs $\{\eta_v^{\mathcal{T}_r},\ v \in \mathcal{V}\}$ will prove useful throughout the paper.

Definition 2.5. *Consider a tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$, and let $\mathcal{T}_r$ be its rooted version, $r \in \mathcal{V}$. For any vector $\boldsymbol{\theta}$ of thinning parameters, let $\boldsymbol{\theta}_{\mathrm{dsc}(v)} = (\theta_j,\ j \in \mathrm{dsc}(v))$ for all $v \in \mathcal{V}$. We define $\{\eta_v^{\mathcal{T}_r},\ v \in \mathcal{V}\}$ as a sequence of joint pgfs through the recursive relation*

$$\eta_v^{\mathcal{T}_r}(\boldsymbol{t}_{v\mathrm{dsc}(v)};\ \boldsymbol{\theta}_{\mathrm{dsc}(v)}) := t_v \prod_{j \in \mathrm{ch}(v)} \left(1 - \theta_j + \theta_j \eta_j^{\mathcal{T}_r}(\boldsymbol{t}_{j\mathrm{dsc}(j)};\ \boldsymbol{\theta}_{\mathrm{dsc}(j)})\right), \quad \boldsymbol{t} \in [-1, 1]^d, \tag{2.3}$$

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

where $\boldsymbol{t}_{v\mathrm{dsc}(v)}$ *is a short-hand notation for the vector* $(t_j,\, j \in \{v\} \cup \mathrm{dsc}(v))$*, and with the convention* $\eta_j^{\mathcal{T}_r}(\boldsymbol{t}_{j\mathrm{dsc}(j)}; \boldsymbol{\theta}_{\mathrm{dsc}(j)}) = t_j$ *for vertices* j *that have no children according to the rooting in* r*.*

In the following proposition, we present the joint pmf and joint pgf of $\boldsymbol{N}$ as in (2.2).

Proposition 2.6. *Let* $\boldsymbol{N} \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T})$*, where* $(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \in \boldsymbol{\Lambda}$*. For a chosen root* $r \in \mathcal{V}$*, let* $\mathcal{T}_r$ *be the rooted version of* $\mathcal{T}$ *and* $\zeta_{L_v} = \lambda_v(1 - \alpha_{(\mathrm{pa}(v),v)}\sqrt{\lambda_{\mathrm{pa}(v)}/\lambda_v})$ *for* $v \in \mathcal{V}\backslash\{r\}$*. Then,*

(i) the joint pmf of $\boldsymbol{N}$ *is given by*

$$p_{\boldsymbol{N}}(\boldsymbol{x}) = \frac{\mathrm{e}^{-\lambda_r}\lambda_r^{x_r}}{x_r!} \prod_{v \in \mathrm{dsc}(r)} \sum_{k=0}^{\min(x_{\mathrm{pa}(v)}, x_v)} \frac{\mathrm{e}^{-\zeta_{L_v}}(\zeta_{L_v})^{x_v - k}}{(x_v - k)!} \binom{x_{\mathrm{pa}(v)}}{k} (\theta_v)^k \, (1 - \theta_v)^{x_{\mathrm{pa}(v)} - k}, \tag{2.4}$$

for $\boldsymbol{x} \in \mathbb{N}^d$*, where* $\theta_v = \alpha_{(\mathrm{pa}(v),v)}\sqrt{\lambda_v/\lambda_{\mathrm{pa}(v)}}$ *for all* $v \in \mathrm{dsc}(r)$*;*

(ii) the joint pgf of $\boldsymbol{N}$ *is given by*

$$\mathcal{P}_{\boldsymbol{N}}(\boldsymbol{t}) = \prod_{v \in \mathcal{V}} \mathrm{e}^{\zeta_{L_v}\left(\eta_v^{\mathcal{T}_r}\left(\boldsymbol{t}_{v\mathrm{dsc}(v)}; \boldsymbol{\theta}_{\mathrm{dsc}(v)}\right) - 1\right)}, \quad \boldsymbol{t} \in [-1, 1]^d, \tag{2.5}$$

where $\boldsymbol{\theta}_{\mathrm{dsc}(v)} = (\alpha_{(\mathrm{pa}(k),k)}\sqrt{\lambda_k/\lambda_{\mathrm{pa}(k)}},\, k \in \mathrm{dsc}(v))$ *is the vector of thinning parameters for the propagation random variables according to a rooting in* r*.*

The joint pmf in (2.4) factorizes along the tree, underscoring that $\boldsymbol{N}$ is a MRF. For further discussions on Gibbs distributions and pmf factorization, we refer to Appendix B, as well as Chapter 4.2 of Cressie and Wikle (2015) and Côté *et al*. (2025b). The analytical form of the joint pmf in (2.4) enables efficient numerical evaluation of the likelihood, making it particularly well suited for parameter estimation procedures, as shown in Sections 5 and 6.

In the upcoming subsection, we show that every distribution of $\mathbb{MPMRF}$ may be reparameterized such that $\boldsymbol{N}$ admits an alternative stochastic representation based on common shocks. Whereas models based on the common-shock approach, whose family of distributions we write $\mathbb{MPCS}$, may become intractable in high dimensions due to the exponential growth in the number of possible shock configurations, the $\mathbb{MPMRF}$ family scales conveniently to high dimensions.

2.2. A subfamily of the multivariate Poisson distribution based on common shocks

Let $\mathcal{V} = \{1, 2, \ldots, d\}$ be a set of indices and let $\mathscr{P}(\mathcal{V})$ be the power set of $\mathcal{V}$, that is, the set of all subsets of $\mathcal{V}$, including the empty set and $\mathcal{V}$ itself. For every $v \in \mathcal{V}$, let $\mathscr{P}(\mathcal{V}; v) = \{\mathcal{W} \in \mathscr{P}(\mathcal{V}) : v \in \mathcal{W}\}$, that is, $\mathscr{P}(\mathcal{V}; v)$ comprises the elements of $\mathscr{P}(\mathcal{V})$ in which v participates. Hence, $\bigcup_{v \in \mathcal{V}} \mathscr{P}(\mathcal{V}; v) = \mathscr{P}(\mathcal{V})$. We define $\boldsymbol{Y} = (Y_{\mathcal{W}}, \mathcal{W} \in \mathscr{P}(\mathcal{V}))$ as a vector of independent Poisson distributed random variables with a corresponding mean-parameter vector $\boldsymbol{\gamma} = (\gamma_{\mathcal{W}},\, \mathcal{W} \in \mathscr{P}(\mathcal{V}))$, with $\gamma_{\mathcal{W}} \geq 0$ for every $\mathcal{W} \in \mathscr{P}(\mathcal{V})$. We use the convention $Y_{\mathcal{W}} = 0$ whenever $\gamma_{\mathcal{W}} = 0$. Letting $\boldsymbol{D} = (D_v,\, v \in \mathcal{V}) \sim \mathrm{MPCS}(\boldsymbol{\lambda})$, we have $D_v = \sum_{\mathcal{W} \in \mathscr{P}(\mathcal{V}; v)} Y_{\mathcal{W}},\, v \in \mathcal{V}$, where, from the closure on convolution of the Poisson distribution, each component of $\boldsymbol{D}$ is Poisson distributed with parameter $\lambda_v = \sum_{\mathcal{W} \in \mathscr{P}(\mathcal{V}; v)} \gamma_{\mathcal{W}}$. The joint pgf of $\boldsymbol{D}$ is given by

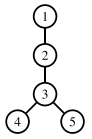
$$\mathcal{P}_{\boldsymbol{D}}(\boldsymbol{t}) = \mathrm{e}^{\left(\gamma_0 + \sum_{\mathcal{W} \in \mathscr{P}(\mathcal{V})} \gamma_{\mathcal{W}} \prod_{v \in \mathcal{W}} t_v\right)}, \quad \boldsymbol{t} \in [-1, 1]^d, \tag{2.6}$$

with $\gamma_0 = -\sum_{\mathcal{W} \in \mathscr{P}(\mathcal{V})} \gamma_{\mathcal{W}}$. We recall that the parameter vector $\boldsymbol{\gamma}$ is of length $|\mathscr{P}(\mathcal{V})| = 2^d$. This may make computations regarding the multivariate Poisson distribution cumbersome, as discussed earlier.

The following proposition provides an alternative parameterization and stochastic representation of $\boldsymbol{N}$ in terms of common shocks.

Proposition 2.7. *Consider a tree* $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ *and, for every* $v \in \mathcal{V}$*, let* Θ_v *be the set of all subtrees of* $\mathcal{T}$ *in which* v *participates, meaning* $\Theta_v = \{\mathcal{W} \in \mathscr{P}(\mathcal{V}; v)$*: for every* $i, j \in \mathcal{W},\, k, l \in \mathcal{W}$ *for every* $(k, l) \in \mathrm{path}(i, j)\}$*. If* $\boldsymbol{N} \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T})$*, with* $(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \in \boldsymbol{\Lambda}$*, then* $\boldsymbol{N}$ *admits the following alternative stochastic*

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press



$$
\begin{aligned}
N_1 &= Y_{\{1\}} + Y_{\{1,2\}} + Y_{\{1,2,3\}} + Y_{\{1,2,3,4\}} + Y_{\{1,2,3,5\}} + Y_{\{1,2,3,4,5\}};\\
N_2 &= Y_{\{2\}} + Y_{\{1,2\}} + Y_{\{2,3\}} + Y_{\{1,2,3\}} + Y_{\{2,3,4\}} + Y_{\{2,3,5\}} + Y_{\{1,2,3,4\}} + Y_{\{1,2,3,5\}} + Y_{\{2,3,4,5\}} + Y_{\{1,2,3,4,5\}};\\
N_3 &= Y_{\{3\}} + Y_{\{2,3\}} + Y_{\{3,4\}} + Y_{\{3,5\}} + Y_{\{1,2,3\}} + Y_{\{2,3,4\}} + Y_{\{2,3,5\}} + Y_{\{3,4,5\}} + Y_{\{1,2,3,4\}} + Y_{\{1,2,3,5\}} + Y_{\{2,3,4,5\}} + Y_{\{1,2,3,4,5\}};\\
N_4 &= Y_{\{4\}} + Y_{\{3,4\}} + Y_{\{2,3,4\}} + Y_{\{3,4,5\}} + Y_{\{1,2,3,4\}} + Y_{\{2,3,4,5\}} + Y_{\{1,2,3,4,5\}};\\
N_5 &= Y_{\{5\}} + Y_{\{3,5\}} + Y_{\{2,3,5\}} + Y_{\{3,4,5\}} + Y_{\{1,2,3,5\}} + Y_{\{2,3,4,5\}} + Y_{\{1,2,3,4,5\}}.
\end{aligned}
$$

Figure 2. *Tree $\mathcal{T}$ of Example 2.8 and **N** components' common shock representations.*

representation:

$$N_v = \sum_{\mathcal{W}\in\Theta_v} Y_{\mathcal{W}}, \quad v \in \mathcal{V}, \tag{2.7}$$

where $\{Y_{\mathcal{W}},\ \mathcal{W} \in \bigcup_{v\in\mathcal{V}} \Theta_v\}$ are independent Poisson random variables of respective parameters

$$\gamma_{\mathcal{W}} = \left(\prod_{w\in\mathcal{W}} \lambda_w\right)\left(\prod_{(i,j)\in\mathcal{E}_{\mathcal{W}}} \frac{\alpha_{(i,j)}}{\sqrt{\lambda_i\lambda_j}}\right)\left(\prod_{(i,j)\in\mathcal{E}^{\dagger}_{\mathcal{W}}} \left(1-\alpha_{(i,j)}\sqrt{\frac{\lambda_j}{\lambda_i}}\right)\right), \quad \mathcal{W} \in \bigcup_{v\in\mathcal{V}} \Theta_v, \tag{2.8}$$

with $\mathcal{E}_{\mathcal{W}} = \{(i,j) \in \mathcal{E}\colon i,j \in \mathcal{W}\}$ and $\mathcal{E}^{\dagger}_{\mathcal{W}} = \{(i,j) \in \mathcal{E}\colon i \in \mathcal{W}, j \notin \mathcal{W}\}$.

The upper limit for α_e, $e \in \mathcal{E}$, for $(\boldsymbol{\lambda}, \boldsymbol{\alpha})$ to be in $\boldsymbol{\Lambda}$, ensures $\gamma_{\mathcal{W}} \geq 0$ for every $\mathcal{W} \in \bigcup_{v\in\mathcal{V}} \Theta_v$.

Given Proposition 2.7, one easily sees that $\boldsymbol{N}$ follows a multivariate Poisson with vector of parameters $\boldsymbol{\gamma} = (\gamma_V,\ V \in \mathscr{P}(\mathcal{V}))$ such that

$$\gamma_V = \begin{cases} \gamma_{\mathcal{W}}, & \text{if } \mathcal{W} \in \bigcup_{v\in\mathcal{V}} \Theta_v \\ 0, & \text{else} \end{cases}, \quad V \in \mathscr{P}(\mathcal{V}).$$

Hence, Proposition 2.7 shows $\mathbb{MPMRF} \subseteq \mathbb{MPCS}$. For a further discussion on the connection between the thinning and the common-shock approaches for Poisson random variables, see Liu *et al*. (2024) and their Remark 2.3 in particular. Although the number of non-zero parameters in the common-shock representation of $\mathbb{MPMRF}$ is lower than 2^d (as for $\mathbb{MPCS}$), the reduction is not substantial enough to overcome computation challenges. Moreover, the parameterization in terms of $\boldsymbol{\gamma}$ intertwines the dependencies and the marginals, thereby removing their intended parametric disconnection. Theorem 2.2 remains a simpler representation, as put forth in Example 2.8. The family $\mathbb{MPMRF}$ being a subset of $\mathbb{MPCS}$, it cannot grow out of the latters' limitations in terms of achievable dependence structures. The advantage of $\mathbb{MPMRF}$ over generic elements of $\mathbb{MPCS}$ comes from its binomial thinning stochastic representation (2.2) and the computational efficiency ensuing.

Example 2.8. *A 5-variate distribution in $\mathbb{MPCS}$ generally requires $2^5 = 32$ parameters. Consider $\boldsymbol{N} \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T})$ where $\mathcal{T}$ is structured as in Figure 2. Using (2.7), we develop **N** into its common shock representation in Figure 2. One notices that constructing $\boldsymbol{\gamma}$ demands $|\bigcup_{v\in\mathcal{V}} \Theta_v| = 17$ non-zero parameters, which is a meaningful diminution, but still much higher than the nine parameters required by the representation in Theorem 2.2. A comparison of N_1 and N_2 in Figure 2 reveals that a change in $\gamma_{\{1,2\}}$ affects both mean parameters of the random variables N_1 and N_2. The parameters $\gamma_{\mathcal{W}}$ associated with each $Y_{\mathcal{W}}$ in Figure 2.8 are given in Table 1. We verify easily that $N_v \sim Poisson(\lambda_v)$ for every $v \in \{1,\ldots,5\}$. In Table 1 of Supplement A, a numerical instance illustrates the equivalence between $\mathbb{MRMRF}$ and $\mathbb{MPCS}$.*

Proposition 2.7 makes clear the difference between $\mathbb{MPMRF}$ and the tree-structured multivariate Poisson distribution examined in Kızıldemir and Privault (2017). In the latter, there are only random variables $Y_{\mathcal{W}}$ from the representation in (2.7) for $\mathcal{W} \in \mathscr{P}(\mathcal{V}; v)$ comprising two elements, given by the set of edges $\mathcal{E}$ of the graph. There are no shock random variables $Y_{\mathcal{W}}$ for $|\mathcal{W}| \geq 3$. As a consequence, the multivariate distribution does not exhibit the conditional independence relations from Definition 2.1 to render a MRF.

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

Table 1. *Parameters $\gamma_{\mathcal{W}}$ for each set $\mathcal{W}$ of vertices in Figure 2.*

Set W	Parameter γ_W
$\{1\}$	$\lambda_1(1-\alpha_{(1,2)}\sqrt{\lambda_2/\lambda_1})$
$\{2\}$	$\sqrt{\lambda_2}(1-\alpha_{(1,2)}\sqrt{\lambda_1/\lambda_2})\,(1-\alpha_{(2,3)}\sqrt{\lambda_3/\lambda_2})$
$\{3\}$	$\sqrt{\lambda_3}(1-\alpha_{(2,3)}\sqrt{\lambda_2/\lambda_3})\,(1-\alpha_{(3,4)}\sqrt{\lambda_4/\lambda_3})(1-\alpha_{(3,5)}\sqrt{\lambda_5/\lambda_3})$
$\{4\}$	$\sqrt{\lambda_4}(1-\alpha_{(3,4)}\sqrt{\lambda_3/\lambda_4})$
$\{5\}$	$\sqrt{\lambda_5}(1-\alpha_{(3,5)}\sqrt{\lambda_3/\lambda_5})$
$\{1,2\}$	$\sqrt{\lambda_1\lambda_2}\alpha_{(1,2)}(1-\alpha_{(2,3)}\sqrt{\lambda_3/\lambda_2})$
$\{2,3\}$	$\sqrt{\lambda_2\lambda_3}\alpha_{(2,3)}(1-\alpha_{(1,2)}\sqrt{\lambda_1/\lambda_2})\,(1-\alpha_{(3,4)}\sqrt{\lambda_4/\lambda_3})(1-\alpha_{(3,5)})$
$\{3,4\}$	$\sqrt{\lambda_3\lambda_4}\alpha_{(3,4)}(1-\alpha_{(2,3)}\sqrt{\lambda_2/\lambda_3})\,(1-\alpha_{(3,5)}\sqrt{\lambda_5/\lambda_3})$
$\{3,5\}$	$\sqrt{\lambda_3\lambda_5}\alpha_{(3,5)}(1-\alpha_{(2,3)}\sqrt{\lambda_2/\lambda_3})\,(1-\alpha_{(3,4)}\sqrt{\lambda_4/\lambda_3})$
$\{1,2,3\}$	$\sqrt{\lambda_1\lambda_3}\alpha_{(1,2)}\alpha_{(2,3)}\,(1-\alpha_{(3,4)}\sqrt{\lambda_4/\lambda_3})(1-\alpha_{(3,5)}\sqrt{\lambda_5/\lambda_3})$
$\{2,3,4\}$	$\sqrt{\lambda_2\lambda_4}\alpha_{(2,3)}\alpha_{(3,4)}\,(1-\alpha_{(1,2)}\sqrt{\lambda_1/\lambda_2})(1-\alpha_{(3,5)}\sqrt{\lambda_5/\lambda_3})$
$\{2,3,5\}$	$\sqrt{\lambda_2\lambda_5}\alpha_{(2,3)}\alpha_{(3,5)}\,(1-\alpha_{(1,2)}\sqrt{\lambda_1/\lambda_2})(1-\alpha_{(3,4)}\sqrt{\lambda_4/\lambda_3})$
$\{3,4,5\}$	$\sqrt{\lambda_3\lambda_4\lambda_5/\lambda_3}\alpha_{(3,4)}\alpha_{(3,5)}\,(1-\alpha_{(2,3)}\sqrt{\lambda_2/\lambda_3})$
$\{1,2,3,4\}$	$\sqrt{\lambda_1\lambda_4}\alpha_{(1,2)}\alpha_{(2,3)}\alpha_{(3,4)}\,(1-\alpha_{(3,5)}\sqrt{\lambda_5/\lambda_3})$
$\{1,2,3,5\}$	$\sqrt{\lambda_1\lambda_5}\alpha_{(1,2)}\alpha_{(2,3)}\alpha_{(3,5)}\,(1-\alpha_{(3,4)}\sqrt{\lambda_4/\lambda_3})$
$\{2,3,4,5\}$	$\sqrt{\lambda_2\lambda_4\lambda_5/\lambda_3}\alpha_{(2,3)}\alpha_{(3,4)}\alpha_{(3,5)}\,(1-\alpha_{(1,2)}\sqrt{\lambda_1/\lambda_2})$
$\{1,2,3,4,5\}$	$\sqrt{\lambda_1\lambda_4\lambda_5/\lambda_3}\alpha_{(1,2)}\alpha_{(2,3)}\alpha_{(3,4)}\alpha_{(3,5)}$

While previous work has extended the multivariate Poisson distribution based on common shocks to higher dimensions, no method combines minimal parameters with the wide variety of dependence structures achievable by $\mathbb{MPMRF}$. For instance, Schulz *et al*. (2021) generalize the bivariate Poisson model from Genest *et al*. (2018) to higher dimensions, requiring only $d+1$ parameters, but this approach imposes limitations on the correlation structure by restricting dependence to a single parameter. Murphy and Schulz (2025) address this limitation with the multivariate Poisson distribution based on triangular comonotonic shocks but requires $d+d(d-1)/2=\mathcal{O}(d^2)$ parameters, still computationally intensive in high-dimensional settings. The $\mathbb{MPMRF}$ family, by comparison, achieves complex dependence structures with only $2d-1$ parameters, scaling more efficiently at $\mathcal{O}(d)$. This allows for convenient estimation in higher dimensions; further discussion is provided in Section 6. The comonotonic shock approach in the above papers accommodates a broader range of correlations, but it cannot be expressed as a thinning operation. One could modify $\mathbb{MPMRF}$ by letting the construction in (2.7) be in terms of comonotonic shocks but would therefore be unable to reexpress it in terms of an iterative stochastic construction as in (2.2). The latter is what enables efficient computational methods, as we discuss in the upcoming sections.

3. Aggregate analysis of the portfolio

A risk model $\boldsymbol{X}=(X_v=\sum_{j=1}^{N_v}B_{v,j},\ v\in\mathcal{V})$ as defined in (1.1) with $\boldsymbol{N}$ from Theorem 2.2 benefits from analytical and computable expressions, even if the dimension $d=|\mathcal{V}|$ is high. The flexibility in choosing parameters $(\boldsymbol{\lambda},\boldsymbol{\alpha})\in\boldsymbol{\Lambda}$ and the underlying tree $\mathcal{T}$ provides a diversity of dependence structures.

The joint Laplace–Stieltjes transform (LST) of $\boldsymbol{X}$, denoted $\mathcal{L}_{\boldsymbol{X}}$, used to obtain the distribution of the aggregate claim amount for the portfolio, is given by

$$\mathcal{L}_{\boldsymbol{X}}(\boldsymbol{t})=\mathbb{E}\left[\prod_{v\in\mathcal{V}}\mathrm{e}^{-t_vX_v}\right]=\mathcal{P}_N(\mathcal{L}_{B_1}(t_1),\ldots,\mathcal{L}_{B_d}(t_d))=\prod_{v\in\mathcal{V}}\mathrm{e}^{\zeta_{L_v}\left(\eta_v^{\mathcal{T}_r}(\boldsymbol{\mathcal{L}}_{B_v}(\boldsymbol{t}_{\mathrm{vdsc}(v)});\boldsymbol{\theta}_{\mathrm{dsc}(v)})-1\right)},\tag{3.1}$$

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

for $\boldsymbol{t} \in \mathbb{R}_+^d$, with the sequence of joint pgfs $\{\eta_v^{\mathcal{T}_r},\ v \in \mathcal{V}\}$ defined by the recursive relation in (2.3), and with the vectors $\boldsymbol{\mathcal{L}}_{B_v}(\boldsymbol{t}_{\text{vdsc}(v)}) = (\mathcal{L}_{B_j}(t_j), j \in \{v\} \cup \text{dsc}(v))$ and $\boldsymbol{\theta}_{\text{dsc}(v)} = (\alpha_{(\text{pa}(k),k)}\sqrt{\lambda_k/\lambda_{\text{pa}(k)}}, k \in \text{dsc}(v))$ for every $v \in \mathcal{V}$.

Given $\mathcal{L}_S(t) = \mathcal{L}_{\boldsymbol{X}}(t,\ldots,t) = \mathcal{P}_N(\mathcal{L}_{B_1}(t),\ldots,\mathcal{L}_{B_d}(t)), t \geq 0$ (Theorem 1 of Wang, 1998), the joint LST in (3.1) leads to the following LST of the aggregate claim S:

$$\mathcal{L}_S(t) = \mathrm{e}^{\sum_{v\in\mathcal{V}} \zeta_{L_v}\left(\sum_{v\in\mathcal{V}} \frac{\zeta_{L_v}}{\sum_{v\in\mathcal{V}}\zeta_{L_v}} \eta_v^{\mathcal{T}_r}(\boldsymbol{\mathcal{L}}_{B_v}(t\,\mathbf{1}_{\text{vdsc}(v)});\boldsymbol{\theta}_{\text{dsc}(v)})-1\right)} = \mathrm{e}^{\lambda_S(\mathcal{L}_{C_S}(t)-1)}, \quad t \geq 0, \tag{3.2}$$

implying that S follows a compound Poisson distribution with primary mean parameter $\lambda_S = \sum_{v\in\mathcal{V}} \zeta_{L_v}$ and secondary LST given by $\mathcal{L}_{C_S}(t) = \sum_{v\in\mathcal{V}} \left(\zeta_{L_v}/\lambda_S\right) \eta_v^{\mathcal{T}_r}(\boldsymbol{\mathcal{L}}_{B_v}(t\,\mathbf{1}_{\text{vdsc}(v)});\boldsymbol{\theta}_{\text{dsc}(v)}), t \geq 0$.

Generating realizations of $\boldsymbol{X}$ is straightforward, given that, the stochastic representation of $\boldsymbol{N}$ allows for an easily scalable sampling method. One generates realizations for each component of $\boldsymbol{N}$ successively; this is well suited for high-dimensional contexts. In this vein, by adapting Algorithm 2 from Côté *et al*. (2025b) to accommodate flexible mean parameters, one can efficiently produce a realization of $\boldsymbol{X}$ by independently producing a realization of $\boldsymbol{N}$ and of the claim amounts.

For discrete claim amount random variables, values of the pmf of S can directly be computed using the fast Fourier transform (FFT) algorithm or Panjer's recursion. Algorithm 1 in Supplement B illustrates a procedure for our context using FFT. Discretization methods or mixed Erlang approximations may be used for continuous claim amounts. Mixed Erlang distributions are known to approximate any continuous positive distribution effectively; see for instance, Tijms (1994).

Remark 3.1. Let $\boldsymbol{X}$ be a multivariate compound Poisson with $\boldsymbol{N} \sim \text{MPMRF}(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T})$. We assume each B_v, $v \in \mathcal{V}$, follows a mixed Erlang distribution with parameters $(\boldsymbol{\pi}_v, \beta_v)$ where $\boldsymbol{\pi}_v = (\pi_{v,k}, k \in \mathbb{N}_1)$ is a vector of non-negative weight parameters, $\sum_{k=1}^{\infty} \pi_{v,k} = 1$, and $\beta_v > 0$. The LST of S in (3.2) becomes

$$\mathcal{L}_S(t) = \exp\left\{\lambda_S\left(\sum_{v\in\mathcal{V}} \frac{\zeta_{L_v}}{\lambda_S} \mathcal{P}_{\boldsymbol{G}_v^{\mathcal{T}_r}}\left\{\left(\mathcal{P}_{\widetilde{K}_j}(\mathcal{L}_{B_{\max}}(t)),\ j \in \{v\} \cup \text{dsc}(v)\right)\right\}\right)\right\} = \mathcal{P}_W(\mathcal{L}_{B_{\max}}(t)), \tag{3.3}$$

for $t \geq 0$, where $\widetilde{K}_j$ is as defined in Appendix A.4, $B_{\max} \sim \text{Exp}(\max_{v\in\mathcal{V}}\beta_v)$ and $\boldsymbol{G}_v^{\mathcal{T}_r} = (G_{v,j}^{\mathcal{T}_r}, j \in \{v\} \cup \text{dsc}(v))$ is a vector of discrete random variables whose joint pgf is given by $\eta_v^{\mathcal{T}_r}(\boldsymbol{t}_{\text{vdsc}(v)}; \boldsymbol{\theta}_{\text{dsc}(v)}), \boldsymbol{t} \in [-1, 1]^d$, as in Definition 2.5. We recognize in (3.3) the LST of a mixed Erlang distribution.

Hence, to perform computations regarding S, one must simply compute the pmf of W, relying on (3.3) and Algorithm 1. This is at the core of Algorithm 2 in Supplement B, which computes the cumulative distribution function (cdf) of S under mixed Erlang claim amounts. With the distribution of S, one may compute the portfolio's required capital through different risk measures. This allows to complete our first risk management task regarding the quantification of the portfolio's risk.

4. Risk sharing

A subsequent risk management task involves the proper allocation of the portfolio's required capital to each component. This allocation can be performed *ex ante*; the allocation rule divides the overall portfolio's risk, quantified by a risk measure, into shares for each component of $\boldsymbol{X}$ based on their respective levels of risk. When dealing with positively homogeneous risk measures, Euler's principle can be utilized to determine the value of these shares. A well-known example of such risk measures is the TVaR. For a random variable Z, the TVaR at confidence level $\kappa \in [0, 1)$ is given by $\text{TVaR}_\kappa(Z) = \frac{1}{1-\kappa}\int_\kappa^1 \text{VaR}_u(Z)\,\mathrm{d}u$, where $\text{VaR}_u(Z) = \inf\{x \in \mathbb{R} : F_Z(x) \geq u\}$, and $u \in [0, 1)$. Let us recall that mixed Erlang distributions may approximate any continuous claim amount distributions; we showed in Section 3 that the pmf of S can be exactly computed in this case. The results from Cossette *et al*. (2012) are thus readily applicable for computing the exact contribution to the TVaR based on Euler's rule. If

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

claim amount distributions are discrete, additional manipulations are required to allocate risk. In such a case, the contribution of X_v, $v \in \mathcal{V}$, to the TVaR of S under Euler's principle is given by

$$\begin{aligned}\mathcal{C}_\kappa^{\mathrm{TVaR}}(X_v; S) &= \frac{1}{1-\kappa}\left(\mathbb{E}[X_v \mathbb{1}_{\{S>\mathrm{VaR}_\kappa(S)\}}] + \mathbb{E}[X_v | S = \mathrm{VaR}_\kappa(S)](F_S(\mathrm{VaR}_\kappa(S)) - \kappa)\right) \\ &= \frac{1}{1-\kappa}\left(\mathbb{E}[X_v] - \sum_{k=0}^{\mathrm{VaR}_\kappa(S)} \mathbb{E}[X_v \mathbb{1}_{\{S=k\}}] + \frac{F_S(\mathrm{VaR}_\kappa(S)) - \kappa}{p_S(\mathrm{VaR}_\kappa(S))} \mathbb{E}[X_v \mathbb{1}_{\{S=\mathrm{VaR}_\kappa(S)\}}]\right), \end{aligned} \tag{4.1}$$

for $\kappa \in [0, 1)$; see, for instance, Section 2 in Mausser and Romanko (2018).

A risk modeler may prefer the covariance-based allocation rule instead of $C_\kappa^{\mathrm{TVaR}}(X_v, S)$ for $v \in \mathcal{V}$. The contribution amount of risk X_v, denoted by $C_\kappa^{\mathrm{Cov}}(X_v, S)$, is given by

$$C_\kappa^{\mathrm{Cov}}(X_v, S) = \mathbb{E}[X_v] + \frac{\mathrm{Cov}(X_v, S)}{\mathrm{Var}(S)} \left(\mathrm{TVaR}_\kappa(S) - \mathbb{E}[S]\right), \quad v \in \mathcal{V}.$$

Both allocation rules ensure that the sum of the contributions equals $\mathrm{TVaR}_\kappa(S)$, the required capital for the tail risk of the portfolio. Moreover, both rules satisfy Euler's principle. For detailed discussions on these allocation principles, we refer the reader to Tasche (1999) and McNeil *et al*. (2015) and to Hesselager and Andersson (2002) for further information on the covariance-based allocation rule.

To allocate the aggregate risk *ex post*, one may choose a fair risk sharing rule. A *risk sharing rule* is a mapping that assigns to each participant a contribution $h_{v,d}(S)$ such that $\sum_{v=1}^d h_{v,d}(S) = S$. A rule is said to be *fair* if it also satisfies $\mathbb{E}[h_{v,d}(S)] = \mathbb{E}[X_v] = \mu_v$ for all v, ensuring participants pay their expected loss on average. In the context of peer-to-peer insurance, for instance, risk sharing rules serve to determine each participant's contribution to the pool (Denuit *et al*. 2022).

Linear fair rules take the form $h_{v,d}^{\mathrm{lin}}(S) = \mu_v + a_{v,d}(S - \mathbb{E}[S])$, where the coefficients $a_{v,d}$ satisfy $\sum_{v\in\mathcal{V}} a_{v,d} = 1$. Two notable examples include the proportional rule, where $a_{v,d} = \mu_v/\mathbb{E}[S]$, allocating risk in proportion to expected losses; the linear regression rule, where $a_{v,d} = \mathrm{Cov}(X_v, S)/\mathrm{Var}(S)$, allocating deviations according to volatility. This rule minimizes the mean squared error $\mathbb{E}[(X_v - h_{v,d}(S))^2]$ among all linear fair rules.

The (nonlinear) conditional-mean risk sharing rule (Denuit and Dhaene, 2012), defined by $h_{v,d}^\star(S) = \mathbb{E}[X_v|S]$, minimizes $\mathbb{E}[(X_v - h_{v,d}(S))^2]$ over all measurable functions $h_{v,d}$ with finite variance. This rule is Pareto-optimal under risk aversion and does not rely on individual preference inclusion, making it particularly suitable for peer-to-peer insurance frameworks (Denuit and Dhaene, 2012). For discrete distributions, we have $\mathbb{E}[X_v|S = k] = \mathbb{E}[X_v \mathbb{1}_{\{S=k\}}]/p_S(k)$, $v \in \mathcal{V}$, $k \in \mathrm{supp}(S)$.

4.1. Computation of risk allocations

A crucial component for calculating both $\mathcal{C}_\kappa^{\mathrm{TVaR}}(X_v; S)$ and $\mathbb{E}[X_v|S = k]$ is the expected allocation: $\mathbb{E}[X_v \mathbb{1}_{\{S=k\}}]$, for $k \in \mathrm{supp}(S)$. The significance of expected allocations in the context of capital allocation is thoroughly discussed in Blier-Wong *et al*. (2025). The authors introduce an ordinary generating function for expected allocations, which is defined as follows.

Definition 4.1 (OGFEA). *Consider a vector of discrete random variables $\mathbf{Z} = (Z_1, \ldots, Z_d)$ taking values in $\mathbb{N}^d$. The ordinary generating function of expected allocations (OGFEA) of Z_v, $v \in \{1, \ldots, d\}$, to the sum of components $\sum_{v=1}^d Z_v$ is given by $\mathcal{P}^{[v]}_{\sum_{v=1}^d Z_v}(t) = \sum_{k=0}^\infty \mathbb{E}[Z_v \mathbb{1}_{\{\sum_{v=1}^d Z_v = k\}}] t^k$, $t \in [-1, 1]$.*

The convenience of OGFEAs lies in the fact that information on expected allocations for all total outcomes is encapsulated within a single power series. We present the OGFEA for our model in the following theorem.

Theorem 4.2. *Consider the risk model in (1.1), where $\mathbf{N} = (N_v, v \in \mathcal{V}) \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T})$, for $(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \in \boldsymbol{\Lambda}$, and a tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$. The OGFEA for X_v to S is given by*

$$\mathcal{P}_S^{[v]}(t) = \lambda_v \, \mathbb{E}[B_v] \, \eta_v^{\mathcal{T}_v}\left(\boldsymbol{s}; \boldsymbol{\theta}^{\mathcal{T}_v}_{\mathrm{dsc}(v)}\right) \, \mathcal{P}_S(t), \quad t \in [-1, 1], \tag{4.2}$$

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

where $\boldsymbol{s} = (s_j, j \in \mathcal{V})$ *is a vector with* $s_v = t\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_{B_v}(t)/\mathbb{E}[B_v]$, *for* $j = v$, *and* $s_j = \mathcal{P}_{B_j}(t)$ *for* $j \in \mathcal{V}\backslash\{v\}$, $t \in [-1, 1]$, *and* $\boldsymbol{\theta}^{\mathcal{T}_v}_{\mathrm{dsc}(v)} = (\alpha_{(\mathrm{pa}(k),k)}\sqrt{\lambda_k/\lambda_{\mathrm{pa}(k)}}, k \in \mathrm{dsc}(v))$.

Let us highlight that $\eta^{\mathcal{T}_v}_v$ and $\theta^{\mathcal{T}_v}_v$ in (4.2) are predicated on the tree rooted at vertex v specifically. In addition to $\lambda_v\,\mathbb{E}[B_v]$, the other two factors in (4.2) are pgfs. Their product can be interpreted as the sum of two independent random variables. Therefore, the coefficients of the OGFEA can be expressed in terms of the pmf of that sum. This is illustrated in the following corollary, which offers a stochastic interpretation of expected allocations.

Corollary 4.3. *Consider the risk model in (1.1), where* $\boldsymbol{N} = (N_v,\ v \in \mathcal{V}) \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T})$, *for* $(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \in \boldsymbol{\Lambda}$ *and a tree* $\mathcal{T} = (\mathcal{V}, \mathcal{E})$. *Define* $\boldsymbol{G}^{\mathcal{T}_v} = (G^{\mathcal{T}_v}_w,\ w \in \mathcal{V})$ *as a vector of random variables with joint pgf given by* $\eta^{\mathcal{T}_v}_w(\boldsymbol{t}_{w\mathrm{dsc}(w)}; \boldsymbol{\theta}^{\mathcal{T}_v}_{\mathrm{dsc}(w)})$ *as in (2.3),* $\boldsymbol{t} \in [-1, 1]^d$. *Consider the random variable*

$$K^{(v)} = \sum_{i=1}^{G^{\mathcal{T}_v}_v} B^*_{v,i} + \sum_{j\in\mathrm{dsc}(v)}\sum_{i=1}^{G^{\mathcal{T}_v}_j} B_{j,i}, \tag{4.3}$$

where B^*_v *is the size-biased transform of* B_v, *that is,* $p_{B^*_v}(x) = \frac{x}{\mathbb{E}[B_v]}p_{B_v}(x)$, *for* $x \in \mathbb{N}_1$. *The expected allocation of* X_v *to* S *for a total outcome* $k \in \mathbb{N}$ *is*

$$\mathbb{E}\left[X_v \mathbb{1}_{\{S=k\}}\right] = \lambda_v\,\mathbb{E}[B_v]\,p_{K^{(v)}+S}(k), \tag{4.4}$$

with $K^{(v)}$ *and* S *mutually independent.*

Since $\sum_{k=0}^{\infty} p_{K^{(v)}+S}(k) = 1$, the summation of $\mathbb{E}[X_v\mathbb{1}_{\{S=k\}}]$ over $k \in \mathbb{N}$ is equal to $\lambda_v\mathbb{E}[B_v]$, as expected. Note that, in the case of independent compound Poisson, that is, $\boldsymbol{\alpha} = 0\,\mathbf{1}_{|\mathcal{E}|}$, we have $K^{(v)} = B^*_v$ and thus recover results of Section 3.2 of Denuit (2019). The result in Corollary 4.3 allows for an explicit expression of contributions to the TVaR under Euler's rule.

Corollary 4.4. *Consider the risk model in (1.1), where* $\boldsymbol{N} = (N_v,\ v \in \mathcal{V}) \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T})$, *for* $(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \in \boldsymbol{\Lambda}$, *and a tree* $\mathcal{T} = (\mathcal{V}, \mathcal{E})$. *For* $v \in \mathcal{V}$, *the contribution of* X_v *to the TVaR of* S *under Euler's rule at confidence level* $\kappa \in [0, 1)$ *is*

$$\mathcal{C}^{\mathrm{TVaR}}_\kappa(X_v; S) = \frac{\lambda_v\mathbb{E}[B_v]}{1-\kappa}\left(1 - F_{K^{(v)}+S}(\mathrm{VaR}_\kappa(S)) + \frac{F_S(\mathrm{VaR}_\kappa(S)) - \kappa}{p_S(\mathrm{VaR}_\kappa(S))}p_{K^{(v)}+S}(\mathrm{VaR}_\kappa(S))\right),$$

where the random variable $K^{(v)}$ *admits the stochastic representation given in (4.3).*

If $\boldsymbol{\lambda} = \lambda\,\mathbf{1}_d$, $\boldsymbol{\alpha} = \alpha\,\mathbf{1}_{|\mathcal{E}|}$ with $\lambda > 0$, and $\alpha \in [0, 1]$, and all B_v are identically distributed, the TVaR contributions in Corollary 4.4 follow the same ordering as in Proposition 1 of Côté *et al*. (2024), bearing connection to the theory of network centrality. See Algorithm 3 in Supplement C for a procedure on computing expected allocations. It relies on the efficiency of the FFT algorithm and scales well to high-dimensional computations.

4.2. Asymptotic results on linear risk sharing

In this section, we investigate the asymptotic behavior of linear risk sharing under the MPMRF risk model. To this aim, we first formalize, in the following proposition, how the covariance between vertices varies as a function of their distance in the graph.

Proposition 4.5. *Let* $\boldsymbol{X}$ *follow the risk model in (1.1), with* $\boldsymbol{N} \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T})$ *as in Theorem 2.2. For all* $v, w \in \mathcal{V}$ *and letting* $\prod_{e\in\emptyset}\alpha_e = 1$, *the covariance between* N_v *and* N_w *is given by*

$$\mathrm{Cov}(N_v, N_w) = \sqrt{\lambda_v\lambda_w}\prod_{e\in\mathrm{path}(v,w)}\alpha_e; \quad \textit{and thus,}\ \mathrm{Cov}(X_v, X_w) = \mathbb{E}[B_v]\mathbb{E}[B_w]\sqrt{\lambda_v\lambda_w}\prod_{e\in\mathrm{path}(v,w)}\alpha_e.$$

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

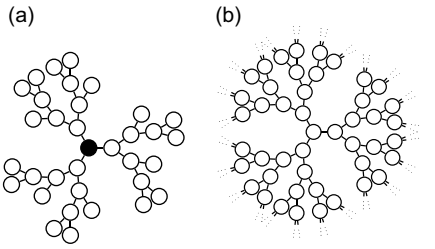


Figure 3. *Illustration of a Cayley tree and a Bethe lattice, both of degree 3.*

The dependence between risks X_v and X_w decays exponentially with the length of the path between v and w. Consequently, as trees grow in radius, correlations between distant vertices are naturally attenuated. We will show that this leads to the law of large numbers being applicable to the average risk S/d.

To have regularly growing trees of infinite vertices, we use Bethe lattices, the infinite analog of the Cayley tree. A Cayley tree $\mathcal{C}^{(\chi,\xi)}$ is a tree such that, with respect to some root $r \in \mathcal{V}$, each non-leaf vertex is connected to χ neighbors, $\chi \geq 2$, also referred to as such a tree's degree, and ξ denotes the length of the shortest path between a root and a leaf. Figure 3(a) provides an illustration of a Cayley tree $\mathcal{C}^{(3,3)}$. More precisely, a Bethe lattice $\mathcal{B}^{(\chi)}$ with degree χ is obtained by letting the length of the shortest path between a root and a leaf ξ tend to ∞. These trees offer a natural framework for studying asymptotic regimes on infinite and expanding tree structures (Ostilli 2012) and allow to derive tractable results. These structures are of particular interest in computational statistics; see Baxter (2016). Figure 3(b) shows a portion of a Bethe lattice with degree $\chi = 3$.

The convergence of S/d defined on a Bethe lattice is explicited in the following proposition. For all $v \in \mathcal{V}$, let $\xi_v = |\text{path}(r, v)|$ denote the distance between the root r and vertex v. For any $\xi \in \mathbb{N}$, we write $\mathcal{T}^{[\xi]} = (\mathcal{V}^{[\xi]}, \mathcal{E}^{[\xi]})$ for a subtree of $\mathcal{T}$ made of all $v \in \mathcal{V}$ such that $\xi_v \leq \xi$, for some root r. Let $d^{[\xi]} = |\mathcal{V}^{[\xi]}|$. Note that $\mathcal{V}^{[0]} = \{r\}$.

Proposition 4.6 (Weak law of large numbers). *Let* $\{\mathcal{B}^{(\chi)[\xi]},\ \xi \in \mathbb{N}\}$ *be a sequence of truncated Bethe lattices of degree* $\chi > 0$ *and* $\{\boldsymbol{X}^{[\xi]},\ \xi \in \mathbb{N}\}$ *be a sequence of portfolios, each defined as in (1.1), where* $\boldsymbol{N}^{[\xi]} \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{B}^{(\chi)[\xi]})$ *and* $\sup_{v\in\mathcal{V}} \mathbb{E}[B_v^2] < \infty$. *Also assume* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\lambda}$ *are uniformly upper bounded, meaning there exists* $\lambda_{\sup} := \sup_{v\in\mathcal{V}} \lambda_v < \infty$ *and* $\alpha_{\sup} := \sup_{e\in\mathcal{E}} \alpha_e \in [0, 1)$. *Let* $S^{[\xi]} = \sum_{v\in\mathcal{V}^{[\xi]}} X_v^{[\xi]}$, *for every* $\xi \in \mathbb{N}$. *The sequence of random variables* $\{W^{[\xi]},\ \xi \in \mathbb{N}\} = \{\frac{1}{d^{[\xi]}} S^{[\xi]},\ \xi \in \mathbb{N}\}$ *converges in probability to* $\mathbb{E}[W^{[\xi]}] < \infty$ *as* $\xi \to \infty$.

Proposition 4.6 implies that, regardless of the local dependence strength, the MPMRF compound distribution will seemingly lead, at the macroscopic level, to an average claim amount with the same behavior as in the independence case. The following theorem shows that, as the number of participants in a portfolio grows, linear risk sharing rules converge in probability to the pure premium.

Theorem 4.7. *Consider the setting of Proposition 4.6. If* $a_{v,d^{[\xi]}} = \mathcal{O}(1/d^{[\xi]})$, *then* $\lim_{\xi\to\infty} h^{lin}_{v,d^{[\xi]}}(S^{[\xi]}) = \mathbb{E}[X_v^{[\xi]}]$ *in probability, for every* $v \in \mathcal{V}$.

Encrypting the dependence structure on a Bethe lattice serves as a reference basis for the convergence results. The following corollary extends the results to general growing tree structures. Let $\deg(v)$ denote the degree of vertex v, that is, the number of neighbors of v in the graph.

Corollary 4.8. *Let* $\{\boldsymbol{X}^{[\xi]},\ \xi \in \mathbb{N}\}$ *be a sequence of portfolios, each defined as in (1.1), with* $\boldsymbol{N}^{[\xi]} \sim MPMRF(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T}^{[\xi]})$, *with* $\sup_{v\in\mathcal{V}^{[\xi]}} \deg(v) = m < \infty$ *for all* $\xi \in \mathbb{N}$. *Also assume* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\lambda}$ *are uniformly upper bounded, meaning there exists* $\lambda_{\sup} := \sup_{v\in\mathcal{V}} \lambda_v < \infty$ *and* $\alpha_{\sup} := \sup_{e\in\mathcal{E}} \alpha_e \in [0, 1)$. *As* $\xi \to \infty$,

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

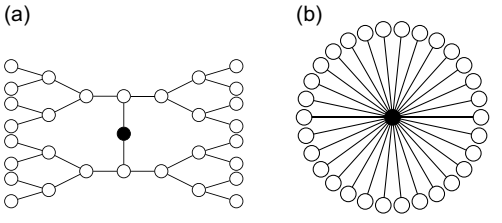


Figure 4. *Tree structures from Example 4.9.*

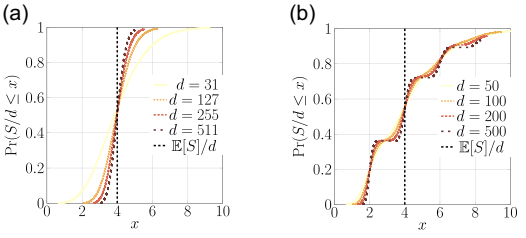


Figure 5. *Cdfs of S/d for Example 4.9.*

the variance of the average claim amount $S^{[\xi]}/d$ is asymptotically upper bounded by that of $S^{*[\xi]}/d$, which is defined on $\mathcal{B}^{(m)}$, with parameters λ_{sup} for all vertices and α_{sup} for all edges.

This result shows that, in large and regularly growing trees with bounded degree, local dependence does not generate clustering strong enough to alter the aggregate behavior of risks, and thus, the latter is akin to that of an independent system for risk sharing purposes. We illustrate an application of Corollary 4.8 in the following example.

Example 4.9 (Asymptotic behavior of S/d). *Consider $\mathcal{T}^{\text{binary}}$, a binary tree (as in Figure 5(a)), and a portfolio of risks $\boldsymbol{X}$ where $\boldsymbol{N} \sim \text{MPMRF}(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathcal{T}^{\text{binary}})$, with $\boldsymbol{\lambda} = \mathbf{1}_d$ and $\boldsymbol{\alpha} = 0.5\mathbf{1}_{d-1}$, and $B_v \sim NBinom(2, 1/3)$ such that $\mathbb{E}[B_v] = 4$. We let $\mathcal{T}^{\text{binary}}$ grow regularly by adding levels to the binary tree structure. By Corollary 4.8, the variance of S/d is asymptotically dominated by that of S^*/d, defined on $\mathcal{B}^{(3)}$. Thus, by Proposition 4.6, the law of large numbers apply. Figure 5(a) shows the cdfs of S/d for progressively larger versions of the trees: we see that the distribution of S/d becomes increasingly concentrated around its theoretical mean $\mathbb{E}[S]/d = 4$. Next, consider $\mathcal{T}^{\text{star}}$, a star tree (as in Figure 5(b)), and let it grow by adding vertices directly to the central node (the dark vertex in Figure 4(b)). Since the resulting sequence of trees does not have uniformly bounded degree, we cannot apply Corollary 4.8. Incidentally, the cdf of S/d, in Figure 5(b), with $\boldsymbol{X}$ defined using the same parameters as before, suggests that the law of large numbers is not applicable in this case.*

Example 4.9 illustrates that S/d may converge to a non-degenerate distribution as $d \to \infty$, which has nontrivial implications for risk pooling and diversification (see, e.g., Denuit and Robert, 2021).

5. Estimation procedure

We describe the estimation procedure for the MPMRF frequency model, which involves constructing a correlation-based maximum spanning tree (MST) to estimate the tree structure and obtaining the maximum likelihood (ML) estimates for $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ given the resulting tree.

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

5.1. Tree structure estimation

To estimate the tree structure, we construct a MST from the matrix of empirical correlations; one may use Kruskal's algorithm for this construction (Kruskal, 1956).

For Ising models with symmetric marginal distributions (i.e., with no external fields), Bresler and Karzand (2020) established that the Chow–Liu dependence tree (Chow and Liu, 1968), which maximizes likelihood, coincides with the tree maximizing empirical correlations. This equivalence reflects the fact that, in that setting, pairwise correlation is a strictly monotone function of the Ising underlying interaction parameter. Consequently, ordering edges by mutual information or by correlation yields the same dependence tree, and the corresponding approximation minimizes the Kullback–Leibler divergence to the true distribution.

An analogous simplification arises in our context. Although the mutual information of the bivariate Poisson distribution does not admit a closed-form expression, the approximation $I(X_1, X_2) \approx -\ln(1 - \rho_P(X_1, X_2))$, is strictly increasing in $\rho_P(X_1, X_2)$ and very accurate for $\rho < 0.7$ (Guerrero-Cusumano, 1995). Ranking edges by empirical mutual information is thus approximately equivalent to ranking them by empirical correlations. This monotonicity allows us to recover the Chow–Liu tree using correlations.

A further justification arises from the small sample sizes in our numerical illustration: a sensitivity analysis (Supplement D) shows that the true tree structure is recovered more reliably if empirical correlations are used as edge weights. This behaviour is consistent with the design of the MPMRF model: edge-wise dependence is parameterized through correlations. We theoretically investigate the sample size required to recover the Chow–Liu MST in a separate working paper.

5.2. MPMRF maximum likelihood estimators

We estimate the frequency model parameters $(\boldsymbol{\lambda}, \boldsymbol{\alpha})$ by maximizing the log-likelihood derived from Proposition 2.6(i), under the constraints of Theorem 2.2, using numerical optimization. The maximization may be carried out using base R `optim` quasi-Newton BFGS algorithm, with parameter bounds enforced through a scaled logit reparametrization and a logarithmic reparametrization for the dependence parameters and marginal mean parameters, respectively. The structure of $\boldsymbol{N}$ facilitates the ML estimation task: its stochastic construction separates marginal and dependence parameters, and the joint pmf (2.4) factorizes over cliques of the tree (pairs of connected vertices). As a result, ML estimation naturally aligns with the sequential procedure frequently applied in copula modeling, itself a specific instance of composite likelihood (Varin *et al*., 2011). Initialization exploits the separation between frequency parameters and dependence parameters in the likelihood. Since the marginal intensities can be estimated independently of the dependence structure, the frequency parameters $\boldsymbol{\lambda}$ are first obtained by maximizing the marginal likelihood. Conditional on these estimates, the dependence parameters $\boldsymbol{\alpha}$ are initialized through sequential bivariate likelihood maximization. These initial values can serve as starting points for the joint ML estimation of the full model. Since the likelihood function is invariant to the choice of root, the optimization results do not depend on a selected root vertex.

6. Data illustration

We examine the applicability and performance of the MPMRF model by analyzing yearly rainfall measurements (in millimeters) from extreme events collected at weather stations in Nova Scotia. The data were sourced from the archives of Environment and Climate Change Canada (ECCC) using the `weathercan` package (LaZerte and Albers, 2018). Our analysis is inspired by the work of Murphy and Schulz (2025), who proposed a multivariate Poisson distribution based on a triangular comonotonic shock construction (MPTCS) and applied it to annual extreme event count data from three weather stations. We compare the performance of the MPMRF frequency model, as proposed in Theorem 2.2, with

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

Table 2. *Vertex numbers, meteorological stations, and climate ID suffixes, and the datasets in which they are used.*

ID	Station	ID suffix[a]	Dataset
1	Baddeck	0300, 0301	3
2	Upper Stewiacke	6200	3
3	Mount Uniacke	3600	2,3
4	Salmon Hole	5000	2,3
5	St Margaret's Bay	4800	3
6	Greenwood A	2000	3
7	Shearwater A	5090	3
8	Springfield	5200	1,3
9	LiverPool Big Falls	3001, 3100	2,3
10	Yarmouth	6490, 6500	3
11	Annapolis Royal	0100	1
12	Kentville CDA	2800, 2810	1

[a]All full station IDs are of the form 820xxxx.

the MPTCS model. Both models feature marginal Poisson distributions and account for positive dependence. Additionally, we extend our analysis to jointly model the frequency and severity of rainfall events across a portfolio of 10 weather stations using the MPMRF risk model.

To conduct our analysis, we work with three datasets. The first dataset, sourced from Murphy and Schulz (2025), consists of 71 yearly trivariate observations of extreme events recorded at three stations covering the period from 1919 to 2000. We create the second and third datasets based on three other stations and 10 stations, respectively, with data spanning from 1949 to 2000. We provide in Table 2 the complete list of weather stations included in this study. Note that some stations were moved during these periods, explaining their two climate identification numbers. The third dataset also records the rainfall amounts associated with each extreme event.

6.1. Datasets 2 and 3 construction

We use the *peak-over-threshold* (POT) approach to model extreme events; see Embrechts *et al*. (1997), Chapter 6. By Theorem 7.1 of Coles (2001), counts of extreme events, over a sufficiently high threshold u, are Poisson distributed. The excesses $Y = X - u | X > u$ follow a generalized Pareto distribution (GPD) (Chavez-Demoulin and Davison, 2005), whose cdf is given by

$$F_Y(y) = \begin{cases} 1 - (1 + \xi y/\sigma)^{-1/\xi}, & \xi \neq 0, \\ 1 - \mathrm{e}^{-y/\sigma}, & \xi = 0, \end{cases}$$

where $\sigma > 0$, and $\xi \in \mathbb{R}$ denote the scale and shape parameters of the GPD.

The construction of Datasets 2 and 3 requires selecting a threshold such that exceedances' frequency and their corresponding excesses follow a Poisson and a GPD distribution, respectively, for each station. Following Thiombiano *et al*. (2017), we relied on the mean-excess plot and the stability of the estimated GPD shape parameter to guide threshold selection. A chi-squared goodness-of-fit test was then applied to the excesses above the chosen threshold. We grouped exceedance-count classes so that all expected frequencies were at least 5. For all stations, the null hypothesis of a GPD distribution could not be rejected at the 20% level. We also performed Anderson–Darling tests, which confirm these conclusions. Because the threshold determines both the number of POT events and the validity of modeling assumptions, we also verified the independence of exceedances. A one-day declustering rule was applied to daily precipitation data, after which the autocorrelation function and Ljung–Box test (p-value of 5%) confirmed negligible residual dependence. For the frequency component, independence of annual

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

Table 3. *Extreme precipitation events threshold u and distribution of cluster sizes (in days) after declustering.*

ID	u	1 day	2 days	3 days	Total
1	37.6	143	6	0	149
2	23.9	389	18	2	409
3	34	289	12	0	301
4	33	236	11	1	248
5	31	282	12	0	294
6	22.8	284	13	1	298
7	30.7	322	10	1	333
8	27.9	345	20	0	365
9	31.2	360	14	1	375
10	24.6	428	15	2	445

Table 4. *Event counts estimates for Datasets 1 and 2.*

Dataset	ID	Station	$\hat{\lambda}^E$	$\hat{\lambda}^A$	$\hat{\lambda}^B$
1	1	Annapolis Royal	7.37	7.37	7.36
	2	Springfield	13.41	13.41	13.37
	3	Kentville CDA	10.66	10.66	10.61
2	1	Liverpool Big Falls	8.72	8.72	8.72
	2	Mount Uniacke	7.00	7.00	7.01
	3	Salmon Hole	5.77	5.77	5.73

Empirical (E), MPMRF (A), Murphy and Schulz (2025) (B).

exceedance counts was examined using the same diagnostics, and a chi-squared goodness-of-fit test showed that the null hypothesis of a Poisson distribution could not be rejected at the 30% level. The selected thresholds, together with the distribution of cluster sizes for each station used in Dataset 2 and 3, are reported in Table 3. The complete dataset construction procedure, along with diagnostic plots and statistical tests are provided in Supplement E.

6.2. Frequency models comparison

We compare the performance of the MPMRF and MPTCS frequency models for the three-station datasets. The ML estimates of the frequency means and the estimated Pearson correlations for both datasets and models are reported in Tables 4 and 5. For both datasets, the means of the MPMRF marginal distributions are exactly the empirical means, but not for MPTCS; this is because MPMRF is able to dissociate marginal and dependence modeling. In terms of dependence, for Dataset 1, the MPMRF model yields correlations that are closer to the empirical values for the edges of its tree, given by (2)–(1)–(3). In contrast, the MPTCS model performs better for the pair (2,3). For Dataset 2, the inferred dependence structure forms the tree (1)–(2)–(3), with MPMRF outperforming MPTCS on the edge (1,2). MPTCS is slightly better on edge (2,3). However, MPTCS is better at capturing the correlation within (1,3). This highlights a structural limitation of the MPMRF model: correlations between non-adjacent vertices necessarily factorize along the unique path of the tree, which restricts the maximal dependence the model can reproduce. The tree structure identifies the most influential relationships within the complete network.

To compare the models formally, we use the Bayesian Information Criterion (BIC) and the corrected Akaike Information Criterion (AICc). The AICc adjusts the traditional AIC for small-sample bias, which

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

Table 5. *Pairwise Pearson's correlations for Datasets 1 and 2.*

Pair	$\hat{\rho}_P^E$	$\hat{\rho}_P^A$	$\hat{\rho}_P^B$
(1,2)	0.625	0.585	0.533
(1,3)	0.570	0.569	0.548
(2,3)	0.494	0.333	0.517
(1,2)	0.604	0.541	0.557
(1,3)	0.460	0.305	0.368
(2,3)	0.627	0.564	0.406

Empirical (E), MPMRF (A), Murphy and Schulz (2025) (B).

Table 6. *Model comparison using BIC and AICc criteria for Datasets 1 and 2.*

Dataset	Criterion	MPMRF model	Murphy and Schulz (2025)
1	BIC	1063.07	**1058.08**
	AICc	1052.68	**1045.82**
2	BIC	**622.51**	626.12
	AICc	**615.33**	617.89

is particularly relevant given our relatively small sample sizes (Hurvich and Tsai, 1989). It is defined as $\text{AICc} = \text{AIC} + 2k(k+1)/(n-k-1)$. Table 6 presents the BIC and AICc values for each model across both datasets. For Dataset 1, the MPTCS model shows lower values for both criteria, indicating a better statistical fit. In contrast, for Dataset 2, the MPMRF model achieves lower values. These findings suggest that neither model consistently outperforms the other across different trivariate datasets.

In terms of computational performance, the MPTCS model encounters challenges in parameter estimation as the dimension increases. As discussed in Section 4 of Murphy and Schulz (2025), non-convergence may occur, particularly in scenarios with smaller sample sizes. To improve stability, the authors use a parameter grid search to initialize the sequential likelihood estimation and recommend employing multiple parameter initializations. For the three-dimensional MPTCS estimation in Dataset 2, we followed this strategy by generating 1000 random initializations of the free parameters and selecting the configuration that attained the highest sequential likelihood value. The runtime was approximately 100 s. Starting from the selected initialization, the two-step ML procedure converges in roughly 18 s, whereas the full ML estimation requires about 32 s. While this grid-search approach is feasible in lower dimensions, it becomes increasingly complex in higher dimensions. For example, achieving estimation at a dimension of $d = 10$ is unfeasible on a personal computer.

In contrast, both the sequential likelihood estimation and the ML estimation of the MPMRF model on Dataset 2 require less than 1 second. The MPMRF frequency model scales well with dimension and retains interpretability in its dependence structure, as demonstrated in the subsequent risk model analysis.

6.3. MPMRF risk model on the 10-station dataset

To showcase the applicability of the MPMRF model for risk modeling in greater dimensions, we now turn to Dataset 3. We fit a risk model as in (1.1) for the rain excesses (not just the frequency of exceedances). For the frequency model, we compute the parameters' estimates using the procedure outlined in Section 5; severities are modeled separately, given their independence with the frequency. The total runtime is less than 2 seconds.

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

Table 7. *Parameter estimates for the frequency, severity, and dependence structure. Bootstrap standard deviations on 1000 samples are provided for* $\boldsymbol{\lambda}$ *and* $\boldsymbol{\alpha}$.

(a) Estimated frequency and severity parameters with characteristics.

	Frequency	Severity			
ID	$\hat{\lambda}$ (sd)	$\hat{\sigma}$	$\hat{\xi}$	$\widehat{E}[B_v]$	$\widehat{\mathrm{Var}}(B_v)$
1	3.47 (0.31)	12.85	0	**50.45**	165.12
2	9.51 (0.48)	11.16	0	35.06	124.55
3	7.00 (0.46)	13.51	0	47.51	182.52
4	5.77 (0.41)	11.79	0.19	**47.56**	**341.72**
5	6.84 (0.44)	13.18	0	44.18	173.71
6	6.93 (0.43)	10.44	0.12	37.26	185.19
7	7.67 (0.48)	13.94	−0.08	43.60	143.62
8	8.49 (0.50)	13.68	0	41.58	187.14
9	8.72 (0.44)	15.05	0	**46.25**	**226.50**
10	**10.35** (0.46)	10.84	0.18	37.82	**273.06**

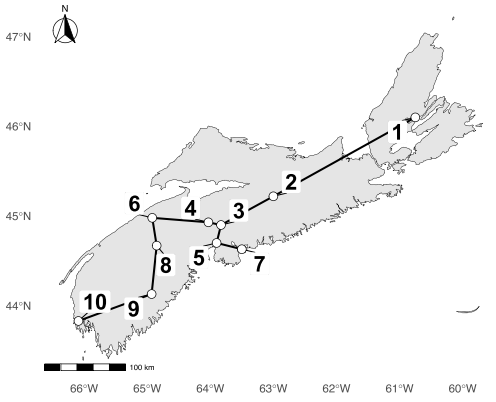
(b) Estimated dependence parameters.

(u, v)	$\hat{\alpha}_{(u,v)}$ (sd)	$\alpha^{\max}_{(u,v)}$	$\frac{\hat{\alpha}_{(u,v)}}{\alpha^{\max}_{(u,v)}}$
(8,9)	0.625 (0.08)	0.987	0.633
(2,3)	0.622 (0.08)	0.858	0.725
(4,6)	0.579 (0.09)	0.912	0.635
(3,5)	0.554 (0.08)	0.988	0.561
(3,4)	0.564 (0.08)	0.908	0.621
(6,8)	0.586 (0.10)	0.904	0.649
(5,7)	0.488 (0.08)	0.944	0.517
(9,10)	0.549 (0.08)	0.918	0.598
(1,2)	0.512 (0.08)	0.604	0.849

For the marginal distributions of $\boldsymbol{X}$, the values of the ML estimates are reported in Table 7(a). For the severity, ML estimation was produced using the `QRM` package of McNeil *et al*. (2015). For stations where the profile likelihood 95% confidence interval for the shape parameter ξ included zero, we set $\hat{\xi} = 0$ and fitted an exponential distribution, following the principle of parsimony. Salmon Hole (ID 4) and Liverpool Big Falls (ID 9) exhibit high severity means and variances, making them significant from a risk modeling perspective. For diagnostics on the marginal fits, see Supplement E.4.

Regarding dependence modeling, the MST constructed from the empirical correlations is diplayed in Figure 6. Dependence parameters are reported in Table 7(b) following the ordering induced by the MST. Column 4 of Table 7(b) reports, for each edge, the ratio of the estimated dependence parameter to its corresponding maximal admissible value, from the constraint $\alpha_{(u,v)} \in (0, \alpha^{\max}_{(u,v)}]$, with $\alpha^{\max}_{(u,v)} := \min\left(\sqrt{\lambda_u/\lambda_v}, \sqrt{\lambda_v/\lambda_u}\right)$, $(u, v) \in \mathcal{E}$. We note that, for most edge combinations, the maximal admissible value largely exceeds the ML estimate. In Figure 7(a), we present the correlation for the frequency, with empirical correlations (lower triangle) and model-induced correlations (upper triangle). Figure 7(b) shows the analogous comparison for annual excesses $\boldsymbol{X}$. Under the independence assumption for claim severities in the MPMRF risk model, the correlations ratio between $\boldsymbol{X}$ and $\boldsymbol{N}$ is $\rho_X(X_u, X_v)/\rho_N(N_u, N_v) = \mathbb{E}[B_u]\mathbb{E}[B_v]/\sqrt{\mathbb{E}[B_u^2]\mathbb{E}[B_v^2]}$ for vertices $u, v \in \mathcal{V}$, which ranges from 0.8541 to 0.9344. The comparison of theoretical (lower triangle) and empirical (upper triangle) pairwise correlations in Figure 7(a) illustrates the effect of path-based correlations on a tree structure. For both heatmaps, we observe clearly how the tree structure shows through the dependence captured by the model: close

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press



Vertex	Station Name
1	Baddeck
2	Upper Stewiacke
3	Mount Uniacke
4	Salmon Hole
5	St Margaret's Bay
6	Greenwood A
7	Shearwater A
8	Springfield
9	Liverpool Big Falls
10	Yarmouth

Figure 6. *Correlation-based maximum spanning tree of 10 meteorological stations in Nova Scotia.*

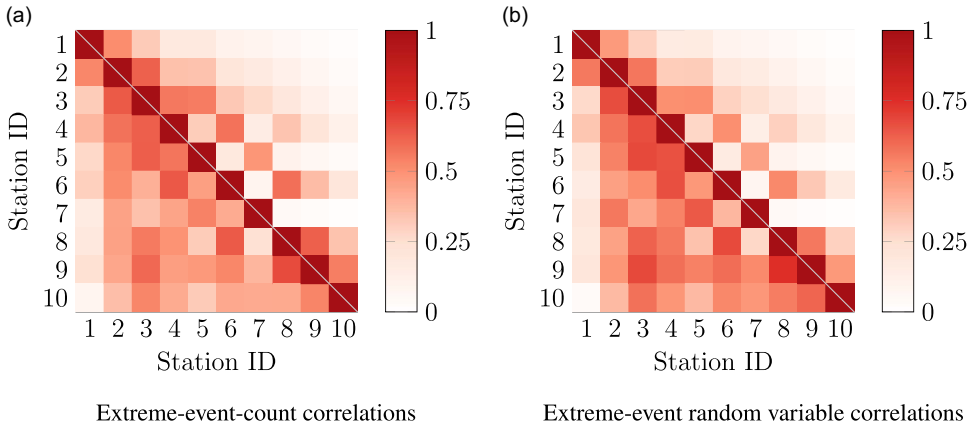


Figure 7. *Correlation heatmaps of yearly extreme precipitation events counts (left) and amounts (right) comparing empirical (lower triangle) and theoretical (upper triangle).*

to the diagonal, where most edges lie, the model's correlations are very similar to the empirical ones. However, as paths between components lengthens, the model cannot capture as much dependence due to the correlation decay feature of tree-structured MRFs: correlations between non-adjacent vertices are forced to factorize multiplicatively along the unique path of the tree.

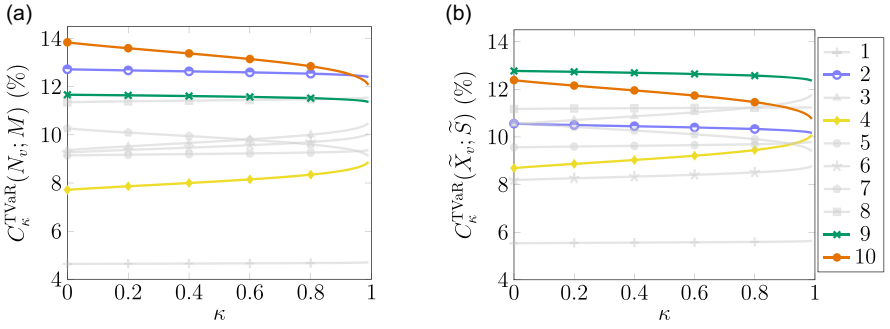
Despite the cost of reduced flexibility in capturing long-range or higher-order interactions, the tree structure confers substantial analytical tractability, enabling closed-form likelihood evaluation, scalable inference, and efficient computation of portfolio-level risk measures. It is also interpretable: Figure 6 confirms the nontrivial spatial-dependence channels through which yearly extreme events co-vary.

We next proceed to perform the two risk management tasks discussed in Sections 3 and 4 for the obtained MPMRF model. The first risk management task is to assess the distribution of S and measure the risk. We proceed using discretization methods. Let $\widetilde{B}$ be the discretized version of a continuous severity random variable B. If $B \sim \text{GPD}(\xi, \sigma; u)$, then $\widetilde{B}$ follows a discretized generalized Pareto distribution (DGPD) with parameters $(\xi, \sigma; u)$, with $p_{\widetilde{B}}(x) = \overline{F}_B(xh) - \overline{F}_B((x+1)h)$, for every $x \in \{u + h : h \in \mathbb{N}\}$ where $\overline{F}$ denotes the survival function and h is the discretization step. For the use of the DGPD in a practical context, see Prieto *et al*. (2014). We use this specification for the severity component in our portfolio analysis with $h = 0.1$, as the dataset's total rainfall is reported on a decimal scale. Using the FFT algorithm, we obtain the distribution of the discretized aggregate loss, denoted $\widetilde{S}$. Table 8 summarizes key

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

Table 8. *Characteristics and risk measures for $\widetilde{S}$, with comparison to the independent case, $\widetilde{S}^{\perp\!\!\!\perp}$. Values are rounded to the nearest whole number.*

	Characteristics			$\text{TVaR}_\kappa(Z)$			
	$\mathbb{E}[Z]$	$\text{Var}(Z)$	CV	$\kappa = 0.80$	$\kappa = 0.90$	$\kappa = 0.95$	$\kappa = 0.99$
$Z = \widetilde{S}^{\perp\!\!\!\perp}$	3155	149,798	0.12	3707	3854	3984	4243
$Z = \widetilde{S}$	3155	442,542	0. 21	4124	4396	4639	5133



Relative contributions under $\boldsymbol{N}$. Relative contributions under $\boldsymbol{X}$.

Figure 8. *Relative contributions to $TVaR_\kappa(Z)$, using the frequency vector $\boldsymbol{N}$ $(Z = M)$ and the risk model vector $\boldsymbol{X}$ $(Z = S)$.*

portfolio-level statistics and compares them with those obtained using the assumption of independent compound Poisson distributions. The marked differences illustrate the critical role of dependence in the frequency component for an accurate assessment of portfolios' risks.

The second risk management task is to analyze how each station individually contributes to the total risk. We use the TVaR allocation principle in (4.1). We compute the TVaR contributions for each vertex in the frequency and compound model using the method based on OGFEAs from Section 4. These contributions are denoted as $C_\kappa^{\text{TVaR}}(N_v; M)$ and $C_\kappa^{\text{TVaR}}(\widetilde{X}_v; \widetilde{S})$, respectively. Figure 8 illustrates the relative contributions to $\text{TVaR}_\kappa(M)$ and $\text{TVaR}_\kappa(\widetilde{S})$ from each station.

By examining Figure 8(a), we observe that stations Upper Stewiacke (ID 2) and Yarmouth (ID 10) have a more significant impact on the overall frequency distribution, which is consistent with their high-frequency mean parameters. For Figure 8(b), which incorporates both the frequency and severity components, Liverpool Big Falls (ID 9) emerges as the dominant contributor across all values of κ. This behavior is driven by the combination of its relatively high mean frequency parameter and its large severity mean and variance (compared to other stations), as reported in Table 7(a). Indeed, a comparison between Figure 8(a) and (b) reveals that, while the shape of the relative contribution curves remains consistent, the curves are shifted vertically depending on the combined frequence-severity distributions.

The relative contribution to aggregate risk of certain random variables decreases as the κ value increases. Locations connected to the tree through weaker (and fewer) edge correlations exhibit declining relative importance as κ increases, reflecting their limited dependence with other vertices. Conversely, locations connected by strong edges and occupying central positions in the MST show increasing contributions at higher κ values. This observation offers critical insights for modeling. By examining the correlation edges of each vertex within the dependence structure, risk modelers can predict which components of the portfolio will see increased relative contributions as they move further into the tail of the aggregate risk distribution. Most notably, Salmon Hole (ID 4) demonstrates the steepest upward trend across κ values. As a consequence of both high severity and its globally influential correlations within

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

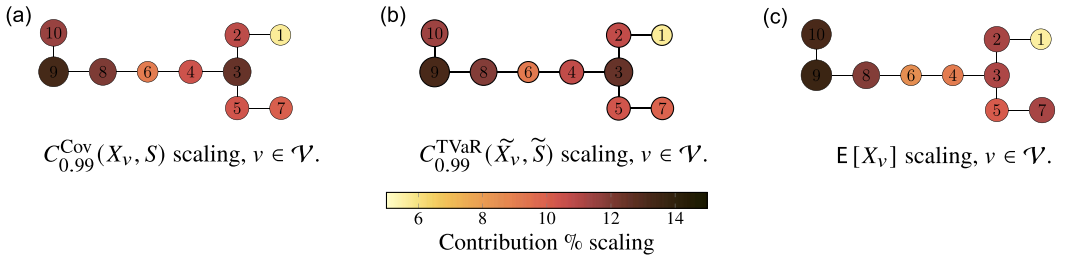


Figure 9. *$\mathcal{T}$ with vertex sizes scaled based on (a) $C_{0.99}^{\mathrm{Cov}}(X_v, S)$, (b) $C_{0.99}^{\mathrm{TVaR}}(\widetilde{X}_v, \widetilde{S})$, and (c) $\mathbb{E}[X_v]$.*

the tree (see Table 6), Salmon Hole's ranking in relative $\mathrm{TVaR}_{0.99}$ contribution rises from 9th (frequency alone) to 6th (frequency-severity).

To compare allocations to vertices using different principles, we evaluate the contribution of each risk under both covariance-based and TVaR-based allocation rules. This comparison is illustrated in Figure 9, which shows the contributions of each station by scaling vertex sizes based on: (a) the covariance-based allocation $C_{\kappa}^{\mathrm{Cov}}(X_v, S)$ and (b) the TVaR-based allocation $C_{\kappa}^{\mathrm{TVaR}}(\widetilde{X}_v, \widetilde{S})$, for each $v \in \mathcal{V}$ and $\kappa = 0.99$. In both cases, scaling is performed relative to the portfolio's TVaR. As a benchmark, panel (c) displays the relative contribution of each station's marginal mean to the portfolio expectation $\mathbb{E}[S]$. Note that we compute $C_{\kappa}^{\mathrm{Cov}}(X_v, S)$ exactly using Proposition 4.5. Supplement F complements Figure 9 by reporting the contributions and their differences, confirming that $C_{\kappa}^{\mathrm{TVaR}}(X_v, S)$ and $C_{\kappa}^{\mathrm{Cov}}(X_v, S)$ are close.

7. Conclusion

We have examined the risk model in (1.1) wherein we introduced dependence between the claim counts using distributions from the family $\mathbb{MPMRF}$. We established that $\mathbb{MPMRF}$ is a subset of $\mathbb{MPCS}$, but its specific parameterization enables more tractable analysis of high-dimensional models with Poisson marginals. We discussed two risk management tasks. First, we provided tools for analyzing the aggregate claim amount S of a portfolio. Second, we investigated risk allocation, providing analytical expression of expected allocations and methods for exact computations of risk allocations. We discussed linear risk sharing for infinite-size portfolios and showed convergence in probability of the average risk amount if the tree's radius becomes large. We described estimation procedures for the model's parameter. Finally, we illustrated our findings through a real data analysis.

Further research can be undertaken on this family of risk models. Tighter and more general asymptotic results for risk sharing rules, in the spirit of Denuit *et al*. (2022), could be established. Incorporating dependence among the severity random variables could further enhance the risk model's use. Focusing on the frequency component, the separation between marginal distributions and the dependence structure naturally lends itself to extensions within the framework of generalized linear models (e.g., Poisson regression with log-linear link). The framework could also be broadened to other distributional families.

A potential limitation of the proposed risk model is the underestimation of correlations between non-adjacent vertices in real-data applications. Analysis of the statistical frequency model indicated that, while the model performs well on edges, it tends to underestimate correlations for non-adjacent pairs. Non-adjacent correlations may be better captured by modeling residual dependence through an auxiliary tree.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/asb.2026.10087

Acknowledgements. We would like to thank three anonymous referees and the handling editor for precious comments that helped improve the quality of the paper. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (Cossette: RGPIN-2025-04077; Marceau: RGPIN-2020-05605; Côté: 581589859, Dubeau: BESC-M) and by the Chaire en actuariat de l'Université Laval (Marceau).

Data availability statement. The data and code will be made available upon request.

Competing interests. The authors have no potential competing or conflicting interests to declare.

References

Baxter, R.J. (2016) *Exactly Solved Models in Statistical Mechanics*. New York, US: Elsevier.

Blier-Wong, C., Cossette, H. and Marceau, E. (2025) Efficient evaluation of risk allocations. *Insurance: Mathematics and Economics*, **122**, 119–136.

Boucher, J.-P., Crainic, A., LeBlanc, A. and Masse, V. (2024) Modeling of fire contagion in farms insurance. *Variance*, **17**(1), 1–16.

Bresler, G. and Karzand, M. (2020) Learning a tree-structured Ising model in order to make predictions. *The Annals of Statistics*, **48**(2), 713–737.

Çekyay, B., Frenk, J. and Javadi, S. (2023) On computing the multivariate Poisson probability distribution. *Methodology and Computing in Applied Probability*, **25**(3), 70.

Chavez-Demoulin, V. and Davison, A.C. (2005) Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **54**(1), 207–222.

Chow, C. and Liu, C. (1968) Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, **14**(3), 462–467.

Coles, S. (2001) Classical extreme value theory and models. In *An Introduction to Statistical Modeling of Extreme Values*, pp. 45–73. London, UK: Springer.

Cossette, H., Mailhot, M. and Marceau, E. (2012) TVaR-based capital allocation for multivariate compound distributions with positive continuous claim amounts. *Insurance: Mathematics and Economics*, **50**(2), 247–256.

Côté, B., Cossette, H. and Marceau, E. (2024) Centrality and shape-related comparisons in a tree-structured Markov random field. arXiv preprint arXiv:2410.20240.

Côté, B., Cossette, H. and Marceau, E. (2025a) Tree-structured Ising models under mean parameterization. arXiv preprint arXiv:2507.18749.

Côté, B., Cossette, H. and Marceau, E. (2025b) Tree-structured Markov random fields with Poisson marginal distributions. *Journal of Multivariate Analysis*, **207**, 105418.

Cressie, N. and Wikle, C.K. (2015) *Statistics for Spatio-Temporal Data*. Hoboken, New Jersey: Wiley.

Cummins, J.D. and Wiltbank, L.J. (1983) Estimating the total claims distribution using multivariate frequency and severity distributions. *Journal of Risk and Insurance*, **50**(3), 377–403.

Denuit, M. (2019) Size-biased transform and conditional mean risk sharing, with application to P2P insurance and tontines. *ASTIN Bulletin*, **49**(3), 591–617.

Denuit, M. and Dhaene, J. (2012) Convex order and comonotonic conditional mean risk sharing. *Insurance: Mathematics and Economics*, **51**(2), 265–270.

Denuit, M., Dhaene, J. and Robert, C.Y. (2022) Risk-sharing rules and their properties, with applications to peer-to-peer insurance. *Journal of Risk and Insurance*, **89**(3), 615–667.

Denuit, M. and Robert, C.Y. (2021) From risk sharing to pure premium for a large number of heterogeneous losses. *Insurance: Mathematics and Economics*, **96**, 116–126.

Denuit, M. and Robert, C.Y. (2022) Conditional mean risk sharing in the individual model with graphical dependencies. *Annals of Actuarial Science*, **16**(1), 183–209.

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

Genest, C., Mesfioui, M. and Schulz, J. (2018) A new bivariate Poisson common shock model covering all possible degrees of dependence. *Statistics & Probability Letters*, **140**, 202–209.

Genest, C. and Nešlehová, J. (2007) A primer on copulas for count data. *ASTIN Bulletin*, **37**(2), 475–515.

Guerrero-Cusumano, J.-L. (1995) The entropy of the multivariate Poisson: An approximation. *Information Sciences*, **86**(1-3), 1–17.

Henn, L.L. (2022) Limitations and performance of three approaches to Bayesian inference for Gaussian copula regression models of discrete data. *Computational Statistics*, **37**(2), 909–946.

Hesselager, O. and Andersson, U. (2002) Risk sharing and capital allocation. Tryg Insurance.

Hurvich, C.M. and Tsai, C.-L. (1989) Regression and time series model selection in small samples. *Biometrika*, **76**(2), 297–307.

Inouye, D.I., Yang, E., Allen, G.I. and Ravikumar, P. (2017) A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, **9**(3), e1398.

Karlis, D. (2003) An EM algorithm for multivariate Poisson distribution and related models. *Journal of Applied Statistics*, **30**(1), 63–77.

Kim, J.H., Jang, J. and Pyun, C. (2019) Capital allocation for a sum of dependent compound mixed Poisson variables: A recursive algorithm approach. *North American Actuarial Journal*, **23**(1), 82–97.

Koller, D. and Friedman, N. (2009) *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, Massachusetts: MIT Press.

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

Krishnamoorthy, A. (1951) Multivariate binomial and Poisson distributions. *Sankhyā: The Indian Journal of Statistics*, 11, 117–124.
Kruskal, J.B. (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, **7**(1), 48–50.
Kzldemir, B. and Privault, N. (2017) Supermodular ordering of Poisson and binomial random vectors by tree-based correlations. *Probability and Mathematical Statistics*, **38**(2), 385–405.
LaZerte, S.E. and Albers, S.J. (2018) weathercan: Download and format weather data from Environment and Climate Change Canada. *Journal of Open Source Software*, **3**(22), 571.
Lindskog, F. and McNeil, A.J. (2003) Common Poisson shock models: Applications to insurance and credit risk modelling. *ASTIN Bulletin: The Journal of the IAA*, **33**(2), 209–238.
Liu, Z., Shi, F., Yao, J. and Yang, Y. (2024) On the maximum and minimum of a multivariate Poisson distribution. arXiv preprint arXiv:2412.13535.
Mausser, H. and Romanko, O. (2018) Long-only equal risk contribution portfolios for CVaR under discrete distributions. *Quantitative Finance*, **18**(11), 1927–1945.
McKenzie, E. (1985) Some simple models for discrete variate time series 1. *Journal of the American Water Resources Association*, **21**(4), 645–650.
McKenzie, E. (1988) Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, **20**(4), 822–835.
McNeil, A.J., Frey, R. and Embrechts, P. (2015) *Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition*. Princeton, New Jersey: Princeton University Press.
M'Kendrick, A. (1925) Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, **44**, 98–130.
Murphy, O.A. and Schulz, J. (2025) A multivariate Poisson model based on a triangular comonotonic shock construction. *Canadian Journal of Statistics*, **53**(3), e70010.
Oberoi, J.S., Pittea, A. and Tapadar, P. (2020) A graphical model approach to simulating economic variables over long horizons. *Annals of Actuarial Science*, **14**(1), 20–41.
Ostilli, M. (2012) Cayley Trees and Bethe Lattices: A concise analysis for mathematicians and physicists. *Physica A: Statistical Mechanics and its Applications*, **391**(12), 3417–3423.
Prieto, F., Gómez-Déniz, E. and Sarabia, J.M. (2014) Modelling road accident blackspots data with the discrete generalized Pareto distribution. *Accident Analysis & Prevention*, **71**, 38–49.
Saoub, K.R. (2021) *Graph Theory: An Introduction to Proofs, Algorithms, and Applications*. Boca Raton, Florida: CRC Press.
Schulz, J., Genest, C. and Mesfioui, M. (2021) A multivariate Poisson model based on comonotonic shocks. *International Statistical Review*, **89**(2), 323–348.
Steutel, F.W., Vervaat, W. and Wolfe, S.J. (1983) Integer-valued branching processes with immigration. *Advances in Applied Probability*, **15**(4), 713–725.
Sundt, B. and Vernic, R. (2009) *Recursions for Convolutions and Compound Distributions with Insurance Applications*. Berlin, Germany: Springer Science & Business Media.
Tasche, D. (1999) Risk contributions and performance measurement. *Report of the Lehrstuhl für Mathematische Statistik, TU München.*
Tasche, D. (2007) Capital allocation to business units and sub-portfolios: The Euler principle. arXiv preprint arXiv:0708.2542.
Teicher, H. (1954) On the multivariate Poisson distribution. *Scandinavian Actuarial Journal*, **1954**(1), 1–9.
Thiombiano, A.N., El Adlouni, S., St-Hilaire, A., Ouarda, T.B. and El-Jabi, N. (2017) Nonstationary frequency analysis of extreme daily precipitation amounts in Southeastern Canada using a peaks-over-threshold approach. *Theoretical and Applied Climatology*, **129**(1), 413–426.
Tijms, H.C. (1994) *Stochastic Models: An Algorithmic Approach*. Chichester, UK: Wiley.
Varin, C., Reid, N. and Firth, D. (2011) An overview of composite likelihood methods. *Statistica Sinica*, **21**(1), 5–42.
Wang, G. and Yuen, K.C. (2005) On a correlated aggregate claims model with thinning-dependence structure. *Insurance: Mathematics and Economics*, **36**(3), 456–468.
Wang, S.S. (1998) Aggregation of correlated risk portfolios: Models and algorithms. *Proceedings of the Casualty Actuarial Society*, **86**, 848–939.
Weiß, C.H. (2008) Thinning operations for modeling time series of counts—a survey. *Advances in Statistical Analysis*, **92**, 319–341.
Willmot, G.E. and Lin, X.S. (2011) Risk modelling with the mixed Erlang distribution. *Applied Stochastic Models in Business and Industry*, **27**(1), 2–16.
Yuen, K.C. and Wang, G. (2002) Comparing two models with dependent classes of business. *Proceedings of the 36th Actuarial Research Conference*.

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

Appendix

A. Proofs

A.1. Proof of Theorem 2.2

First, we argue that $\boldsymbol{N}$ is a MRF. The construction given in (2.2) is akin to the one presented in Theorem 1 of Côté *et al*. (2025b) about the stochastic dynamics at play. The arguments provided for the proof of that theorem remain relevant: the maximum information about a random variable N_v, $v \in \mathcal{V}$, is obtained by knowing the value of its neighbors. Thus, it satisfies the local Markov property, meaning $\boldsymbol{N}$ is a MRF.

We next prove by induction that $N_v \sim \text{Poisson}(\lambda_v)$ for all $v \in \mathcal{V}$, with the root r as the starting point; evidently, $N_r \sim \text{Poisson}(\lambda_r)$. We next suppose the statement holds true for $N_{\text{pa}(w)}$, $w \in \mathcal{V}\backslash\{r\}$, and prove $N_w \sim \text{Poisson}(\lambda_w)$. Following the construction in (2.2), L_w is independent of all L_v, $v \in \mathcal{V}\backslash\{w\}$ and of $N_{\text{pa}(w)}$, since $w \notin \text{path}(\text{pa}(w), r)$. It follows that the pgf of N_w is given by

$$\mathcal{P}_{N_w}(t) = \mathcal{P}_{\left(\alpha_{(\text{pa}(w),w)}\sqrt{\lambda_w/\lambda_{\text{pa}(w)}}\right)\circ N_{\text{pa}(w)}}(t)\mathcal{P}_{L_w}(t), \quad t \in [-1, 1].$$

From the properties of the binomial thinning operator (one may refer to Theorem 11(d) of Côté *et al*. 2025b), we have

$$\begin{aligned}\mathcal{P}_{N_w}(t) &= \mathcal{P}_{N_{\text{pa}(w)}}\left(1 - \alpha_{(\text{pa}(w),w)}\sqrt{\lambda_w/\lambda_{\text{pa}(w)}} + \alpha_{(\text{pa}(w),w)}\sqrt{\lambda_w/\lambda_{\text{pa}(w)}}t\right)\mathcal{P}_{L_w}(t)\\ &= \mathrm{e}^{\lambda_{\text{pa}(w)}\left(1+\alpha_{(\text{pa}(w),w)}\sqrt{\lambda_w/\lambda_{\text{pa}(w)}}(t-1)-1\right)}\mathrm{e}^{\left(\lambda_w-\alpha_{(\text{pa}(w),w)}\sqrt{\lambda_{\text{pa}(w)}\lambda_w}\right)(t-1)}, \quad t \in [-1, 1], \end{aligned} \tag{A1}$$

from the respective pgfs of L_w and $N_{\text{pa}(w)}$ given the induction hypothesis. Simplifying (A1) provides $\mathcal{P}_{N_w}(t) = \mathrm{e}^{\lambda_w(t-1)}$, $t \in [-1, 1]$; thus, $N_w \sim \text{Poisson}(\lambda_w)$. The assertion is validated for both the case of the root and the parent–child inductive case; we conclude $N_v \sim \text{Poisson}(\lambda_v)$ for every $v \in \mathcal{V}$.

Conditioning on $N_{\text{pa}(v)}$ and using (2.2), the joint pgf of two neighbors $(N_{\text{pa}(v)}, N_v)$ is given by

$$\mathcal{P}_{N_{\text{pa}(v)},N_v}(t_{\text{pa}(v)}, t_v) = \mathrm{e}^{\lambda_{\text{pa}(v)}(t_{\text{pa}(v)}-1)+\lambda_v(t_v-1)+\alpha_{(\text{pa}(v),v)}\sqrt{\lambda_{\text{pa}(v)}\lambda_v}(t_v-1)(t_{\text{pa}(v)}-1)}, \quad t_{\text{pa}(v)}, t_v \in [-1, 1].$$

Since $\mathcal{P}_{N_{\text{pa}(v)},N_v}(t_{\text{pa}(v)}, t_v)$ is symmetric regarding $N_{\text{pa}(v)}$ and N_v, the stochastic dynamics on an edge are reversible. Given the local Markov property, this reversibility extends to the stochastic dynamics on a path. Moreover, note that choosing a root $r' \neq r$ only affects the parent–child relationships of the vertices on path(r, r'). Other vertices remain children to their parent, with unchanged stochastic dynamics.

A.2. Proof of Proposition 2.6

The proof resembles those of Theorems 3 and 4 of Côté *et al*. (2025b) and is thus omitted.

A.3. Proof of Proposition 2.7

The choice of the root has no incidence on the distribution of $\boldsymbol{N}$ (Theorem 2.2). For all $v \in \mathcal{V}$, we have

$$\eta_v^{\mathcal{T}_r}(\boldsymbol{t}_{v\text{dsc}(v)}; \boldsymbol{\theta}_{\text{dsc}(v)}) = t_v \sum_{\mathcal{C} \in \mathscr{P}(\text{ch}(v))} \left(\prod_{j \in \text{ch}(v)\backslash\mathcal{W}} (1 - \theta_j)\right)\left(\prod_{i \in \mathcal{W}} \theta_i \eta_i^{\mathcal{T}_r}(\boldsymbol{t}_{i\text{dsc}(i)}; \boldsymbol{\theta}_{\text{dsc}(i)})\right). \tag{A2}$$

Let $\Xi_v = \{\mathcal{W} \cap v\text{dsc}(v) : \mathcal{W} \in \Theta_v\}$. Write $\text{ch}^i(v)$ to refer to the ith generation of the descendants of v, $i \in \mathbb{N}$; for instance, $\text{ch}^2(v)$ are the children of the children of v. Let $\Xi_v^{[i]} = \{\mathcal{W} \cap \text{ch}^i(v) : \mathcal{W} \in \Xi_v\}$. From

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

(A2), we use the result recursively to further develop $\eta_v^{\mathcal{T}_v}$, yielding after one iteration

$$\begin{aligned}
&\eta_v^{\mathcal{T}_r}(\boldsymbol{t}_{v\mathrm{dsc}(v)};\boldsymbol{\theta}_{\mathrm{dsc}(v)})\\
&= t_v \sum_{\mathcal{W}\in\Xi_v^{[2]}} \left(\prod_{j\in \mathrm{ch}(v)\setminus\mathcal{C}} (1-\theta_j)\right)\left(\prod_{i\in\mathcal{C}} \theta_i t_i \left(\prod_{k\in\mathrm{ch}(i)\setminus\mathcal{W}} (1-\theta_k)\right)\left(\prod_{\ell\in\mathrm{ch}(i)\cup\mathcal{W}} \theta_\ell \eta_i^{\mathcal{T}_r}(\boldsymbol{t}_{i\mathrm{dsc}(i)};\boldsymbol{\theta}_{\mathrm{dsc}(i)})\right)\right)\\
&= t_v \sum_{\mathcal{W}\in\Xi_v^{[2]}} \left(\prod_{(\mathrm{pa}(j),j)\in\mathcal{E}_{\mathcal{W}}^{\dagger}} (1-\theta_j)\right)\left(\prod_{(\mathrm{pa}(i),i)\in\mathcal{E}_{\mathcal{W}}} \theta_i t_i\right)\left(\prod_{\ell\in\mathrm{ch}^2(i)} \frac{1}{t_\ell}\eta_\ell^{\mathcal{T}_r}(\boldsymbol{t}_{\ell\mathrm{dsc}(\ell)};\boldsymbol{\theta}_{\mathrm{dsc}(\ell)})\right). \qquad \text{(A3)}
\end{aligned}$$

Continuing recursively in this fashion down to the bottom of the tree (say it takes $m\in\mathbb{N}$ iterations), we have $\Xi_v^{[m]}=\Xi_v$, and (A3) becomes

$$\eta_v^{\mathcal{T}_r}(\boldsymbol{t}_{v\mathrm{dsc}(v)};\boldsymbol{\theta}_{\mathrm{dsc}(v)}) = t_v \sum_{\mathcal{W}\in\Xi_v} \left(\prod_{(\mathrm{pa}(j),j)\in\mathcal{E}_{\mathcal{W}}^{\dagger}} (1-\theta_j)\right)\left(\prod_{(\mathrm{pa}(i),i)\in\mathcal{E}_{\mathcal{W}}} \theta_i t_i\right), \qquad \text{(A4)}$$

since $\eta_\ell^{\mathcal{T}_r}(\boldsymbol{t}_{\ell\mathrm{dsc}(\ell)};\boldsymbol{\theta}_{\mathrm{dsc}(\ell)})=t_\ell$ for every $\ell\in\xi_v^{[m]}$. Inserting (A4) into (2.5), $\mathcal{P}_N$ is thus given by

$$\mathcal{P}_N(\boldsymbol{t}) = \prod_{v\in\mathcal{V}}\prod_{\mathcal{W}\in\Xi_v} \mathrm{e}^{\zeta_{L_v}\left(\prod_{(\mathrm{pa}(j),j)\in\mathcal{E}_{\mathcal{W}}^{\dagger}}(1-\theta_j)\right)\left(\prod_{(\mathrm{pa}(i),i)\in\mathcal{E}_{\mathcal{W}}}\theta_i t_i\right)t_v-1}, \quad \boldsymbol{t}\in[-1,1]^d, \qquad \text{(A5)}$$

where $\{\zeta_{L_v},\ v\in\mathcal{V}\}$ and $\{\theta_v,\ v\in\mathcal{V}\}$ are defined as in Proposition 2.6. Note that from the construction of Ξ_v, no combination of t_v's is repeated across the product operations. One recognizes in (A5) the joint pgf of the sum of independent Poisson random vectors of the form $\boldsymbol{Y}_{\mathcal{W}}=(Y_{\mathcal{W}}\mathbb{1}_{v\in\mathcal{W}},\ v\in\mathcal{V})$, one for each element of $\bigcup_{v\in\mathcal{V}}\Xi_v=\bigcup_{v\in\mathcal{V}}\Theta_v$, where the frequency parameter associated with $Y_{\mathcal{W}}$ is

$$\gamma_{\mathcal{W}} = \zeta_{L_v}\left(\prod_{(\mathrm{pa}(j),j)\in\mathcal{E}_{\mathcal{W}}^{\dagger}} (1-\theta_j)\right)\left(\prod_{(\mathrm{pa}(i),i)\in\mathcal{E}_{\mathcal{W}}} \theta_i\right), \quad \mathcal{W}\in\bigcup_{v\in\mathcal{V}}\Theta_v, \qquad \text{(A6)}$$

where v is the utmost vertex in $\mathcal{W}$ according to the rooting in r. Equation (A6) is equivalent to (2.8) after substituting back to the original parameters.

A.4. Proof of Remark 3.1

The techniques for that matter are inspired from the work in Willmot and Lin (2011). The cdf and LST of B_v, $v\in\mathcal{V}$, are given respectively by

$$F_{B_v}(x) = \sum_{k=1}^{\infty} \pi_{v,k} H(x;k,\beta_v) \quad x\geq 0, \qquad \mathcal{L}_{B_v}(t) = \sum_{k=1}^{\infty} \pi_{v,k}\left(\frac{\beta_v}{\beta_v+t}\right)^k, \quad t>0,$$

where $H(x;k,\beta_v)=1-\mathrm{e}^{-\beta_v x}\sum_{l=0}^{k-1}\frac{(\beta_v x)^l}{l!}$, $x\geq 0$, is the cdf of the kth Erlang distribution with rate β_v.

To reformulate $\mathcal{L}_S$ in (3.2), we first express every $\mathcal{L}_{B_v}$ according to the maximum rate parameter $\beta_{\max}:=\max\{\beta_v : v\in\mathcal{V}\}$. Let $v_{\max}:=\mathrm{argmax}\{\beta_v : v\in\mathcal{V}\}$. For $v\in\mathcal{V}\setminus\{v_{\max}\}$, we can express $\mathcal{L}_{B_v}$ as

$$\mathcal{L}_{B_v}(t) = \sum_{k=1}^{\infty}\pi_{v,k}\left[\frac{q_v}{1-(1-q_v)\left(\frac{\beta_{\max}}{\beta_{\max}+t}\right)}\left(\frac{\beta_{\max}}{\beta_{\max}+t}\right)\right]^k = \sum_{k=1}^{\infty}\pi_{v,k}\mathcal{P}_{K_{v,k}}\left(\frac{\beta_{\max}}{\beta_{\max}+t}\right), \quad t\geq 0, \qquad \text{(A7)}$$

where $K_{v,k}$ follows a negative binomial distribution with number of successful trials k and success probability $q_v=\beta_v/\beta_{\max}$. For all $v\in\mathcal{V}$, we may write $\mathcal{L}_{B_v}(t)=\mathcal{P}_{\widetilde{K}_v}(\mathcal{L}_{B_{\max}}(t))$, corresponding to a compound distribution with primary distribution given by the random variable $\widetilde{K}_v$ with pmf $p_{\widetilde{K}_v}(x)=\sum_{k=1}^{\infty}\pi_{v,k}\,p_{K_{v,k}}(x)$, $x\in\mathbb{N}_1$, and secondary distribution $B_{\max}\sim\mathrm{Exp}(\beta_{\max})$. The result follows inserting (A7) in (3.2).

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

A.5. Proof of Theorem 4.2

From Theorem 3.4 of Blier-Wong *et al*. (2025) and (3.1), taking v as the root for the joint pgf of $\boldsymbol{X}$,

$$\begin{aligned}\mathcal{P}_S^{[v]}(t) &= \left[t_v \frac{\partial}{\partial t_v} \mathcal{P}_{\boldsymbol{X}}(\boldsymbol{t}) \right]\Bigg|_{\boldsymbol{t}=t\mathbf{1}_d} \\ &= t \left[\frac{\partial}{\partial t_v} \mathrm{e}^{\lambda_v \eta_v^{\mathcal{T}_v}\left(\boldsymbol{\mathcal{P}}_{B_v}(\boldsymbol{t}_{v\mathrm{dsc}(v)});\, \boldsymbol{\theta}^{\mathcal{T}_v}_{\mathrm{dsc}(v)}\right)} \prod_{j\in\mathcal{V}\setminus\{v\}} \mathrm{e}^{\zeta_{L_j} \eta_j^{\mathcal{T}_v}\left(\boldsymbol{\mathcal{P}}_{B_j}(\boldsymbol{t}_{j\mathrm{dsc}(j)});\, \boldsymbol{\theta}^{\mathcal{T}_v}_{\mathrm{dsc}(j)}\right)} \right]\Bigg|_{\boldsymbol{t}=t\mathbf{1}_d}, \end{aligned} \tag{A8}$$

for $t \in [-1, 1]$, where $\zeta_{L_j} = \lambda_j(1 - \alpha_{(\mathrm{pa}(j),j)}\sqrt{\lambda_{\mathrm{pa}(j)}/\lambda_j})$. We choose v as the root for $\mathcal{P}_X$ as it simplifies the differentiation and has no incidence on the result. All the multiplicands in (A8) are free of t_v since, if v is the root, $v \notin j\mathrm{dsc}(j)$ for every other $j \in \mathcal{V}\setminus\{v\}$. Performing the differentiation in (A8) yields

$$\begin{aligned}\mathcal{P}_S^{[v]}(t) &= \lambda_v t \left[\frac{\partial}{\partial t_v} \eta_v^{\mathcal{T}_v}(\boldsymbol{\mathcal{P}}_{B_v}(\boldsymbol{t}_{v\mathrm{dsc}(v)}); \boldsymbol{\theta}^{\mathcal{T}_v}_{\mathrm{dsc}(v)}) \right]\Bigg|_{t=t\mathbf{1}_d} \prod_{j\in\mathcal{V}} \mathrm{e}^{\zeta_{L_j} \eta_j^{\mathcal{T}_v}\left(\boldsymbol{\mathcal{P}}_{B_j}(t\,\mathbf{1}_{j\mathrm{dsc}(j)});\, \boldsymbol{\theta}^{\mathcal{T}_v}_{\mathrm{dsc}(j)}\right)} \\ &= \lambda_v t \left[\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{P}_{B_v}(t) \right] \frac{1}{\mathcal{P}_{B_v}(t)} \eta_v^{\mathcal{T}_v}(\boldsymbol{\mathcal{P}}_{B_v}(t\,\mathbf{1}_{v\mathrm{dsc}(v)}); \boldsymbol{\theta}^{\mathcal{T}_v}_{\mathrm{dsc}(v)}) \mathcal{P}_S(t), \quad t \in [-1, 1], \end{aligned}$$

from $\eta_v^{\mathcal{T}_v}$ given in (2.3), with the vector $\boldsymbol{\mathcal{P}}_{B_v}(\boldsymbol{t}_{v\mathrm{dsc}(v)}) = (\mathcal{P}_{B_{v,j}}(t_j), j \in \{v\} \cup \mathrm{dsc}(v))$.

A.6. Proof of Corollary 4.3

The pgf of B_v^* is given by $\mathcal{P}_{B_v^*}(t) = \frac{t}{\mathbb{E}[B_v]} \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{P}_{B_v}(t)$. Hence, the OGFEA in (4.2) is rewritten as

$$\mathcal{P}_S^{[v]}(t) = \lambda_v \mathbb{E}[B_v] \mathcal{P}_{K^{(v)}}(t) \mathcal{P}_S(t) = \lambda_v \mathbb{E}[B_v] \mathcal{P}_{K^{(v)}+S}(t) = \sum_{k=0}^{\infty} \left(\lambda_v \mathbb{E}[B_v] p_{K^{(v)}+S}(k)\right) t^k,$$

$t \in [-1, 1]$, with the second equality following from the independence of $K^{(v)}$ and S. From the OGFEA definition in (4.2), the expected allocations for $k \in \mathbb{N}$ are given by the polynomial's coefficients.

A.7. Proof of Corollary 4.4

The result follows directly by inserting (4.4) into (4.1).

A.8. Proof of Proposition 4.5

The proof is straightforward for $\mathrm{Cov}(N_v, N_w)$ if $v = w$. We now suppose $v = \mathrm{pa}(w)$; then, given (2.2),

$$\mathrm{Cov}(N_v, N_w) = \mathrm{Cov}\left(N_v, \left(\alpha_{(v,w)}\sqrt{\lambda_w/\lambda_v}\right) \circ N_v + L_w\right) \overset{\perp\!\!\!\perp}{=} \mathrm{Cov}\left(N_v, \left(\alpha_{(v,w)}\sqrt{\lambda_w/\lambda_v}\right) \circ N_v\right). \tag{A9}$$

From the properties of the binomial thinning operator, (A9) becomes

$$\mathrm{Cov}(N_v, N_w) = \alpha_{(v,w)}\sqrt{\lambda_w/\lambda_v}\mathrm{Var}(N_v) = \sqrt{\lambda_v \lambda_w}\alpha_{(v,w)}, \tag{A10}$$

which corresponds to our result for the case $v = \mathrm{pa}(w)$. The general result for every $v, w \in \mathcal{V}$ is then obtained by using (A10) and the same *modus operandi* as in the proof of Theorem 5 of Côté *et al*. (2025b) – that is, by iterative conditioning on every successive vertex on the path from v to w. Conditioning on both claim count random variables, given that $\{B_{v,j},\ j \in \mathbb{N}_1\}$ and $\{B_{w,j},\ j \in \mathbb{N}_1\}$ are independent sequences of independent and identically distributed random variables, yields the desired result.

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

A.9. Proof of Proposition 4.6

The number of vertices in $\mathcal{B}^{(\chi)[\xi]}$ is given by

$$d^{[\xi]} = |\mathcal{V}^{[\xi]}| = 1 + \sum_{i=0}^{\xi} \chi(\chi-1)^i = 1 + \chi\frac{(\chi-1)^{\xi-1}-1}{\chi-2}, \quad \xi \in \mathbb{N}. \tag{A11}$$

Let $\mathcal{V}_i$ be the set of vertices in the ith level of $\mathcal{B}^{(\chi)[\xi]}$, $i \in \{0, 1, \ldots, \xi\}$. For $u \in \mathcal{V}_0$, the vertex at the center of the Cayley tree, we obtain, from Proposition 4.5,

$$\mathrm{Cov}(X_u^{[\xi]}, S^{[\xi]}) = \sum_{i=0}^{\xi}\sum_{v\in\mathcal{V}_i} \mathrm{Cov}(X_v^{[i]}, X_u^{[i]}) = \lambda_u\mathbb{E}[B_u^2] + \sum_{i=1}^{\xi}\mathbb{E}[B_u]\mathbb{E}[B_v]\sqrt{\lambda_u\lambda_v}\prod_{e\in\mathrm{path}(u,v)}\alpha_e. \tag{A12}$$

Bounding (A12) using $\lambda_{\sup}$, $\alpha_{\sup}$ and a constant $C = \max_{u,v\in\mathcal{V}}(\mathbb{E}[B_u^2], \mathbb{E}[B_u]\mathbb{E}[B_v])$, we obtain

$$\begin{aligned}
\mathrm{Cov}(X_u^{[\xi]}, S^{[\xi]}) &\le C\left(\lambda_{\sup} + \sum_{i=1}^{\xi}\sum_{v\in\mathcal{V}_i}\sqrt{\lambda_{\sup}\lambda_{\sup}}\alpha_{\sup}^{|\mathrm{path}(u,v)|}\right)\\
&= C\left(\lambda_{\sup} + \lambda_{\sup}\sum_{i=1}^{\xi}\chi\alpha_{\sup}((\chi-1)\alpha_{\sup})^{i-1}\right)\\
&= C\lambda_{\sup}\left(1 + \chi\alpha_{\sup}\frac{((\chi-1)\alpha_{\sup})^{\xi-1}-1}{(\chi-1)\alpha_{\sup}-1}\right),
\end{aligned} \tag{A13}$$

for $\xi \in \mathbb{N}$. The topology in a Bethe lattice is such that the structure around every vertex remains identical, regardless of the chosen vertex. Hence, for every $v \in \mathcal{V}$,

$$\lim_{\xi\to\infty}\mathrm{Cov}(X_v^{[\xi]}, S^{[\xi]}) = \lim_{\xi\to\infty}\mathrm{Cov}(X_u^{[\xi]}, S^{[\xi]}),\ u \in \mathcal{V}_0.$$

Therefore, we have

$$\mathrm{Var}(W^{[\xi]}) = \lim_{\xi\to\infty}\frac{1}{(d^{[\xi]})^2}\mathrm{Var}(S^{[\xi]}) = \lim_{\xi\to\infty}\frac{1}{(d^{[\xi]})^2}\sum_{v\in\mathcal{V}}\mathrm{Cov}(X_v^{[\xi]}, S^{[\xi]}). \tag{A14}$$

Given (A11), (A13), and since $\alpha_{\sup} \in [0, 1)$, (A14) becomes

$$\mathrm{Var}(W^{[\xi]}) \le \lim_{\xi\to\infty} C\lambda_{\sup}\left(1 + \chi\alpha_{\sup}\frac{((\chi-1)\alpha_{\sup})^{\xi-1}-1}{(\chi-1)\alpha_{\sup}-1}\right)\left(1 + \chi\frac{(\chi-1)^{\xi-1}-1}{\chi-2}\right)^{-1} = 0.$$

By Chebyshev's inequality,

$$\Pr\left(\left|W^{[\xi]} - \mathbb{E}[W^{[\xi]}]\right| > \varepsilon\right) \le \varepsilon^{-2}\mathrm{Var}(W^{[\xi]}). \tag{A15}$$

We have $\lim_{\xi\to\infty}\mathrm{Var}(W^{[\xi]}) = 0$ since the sequence $\{\mathcal{B}^{(\chi)[\xi]}, \xi \in \mathbb{N}\}$ converges to a Bethe lattice as $\xi \to \infty$. The right-hand side of (A15) vanishes. This implies $W^{[\xi]} \to \mathbb{E}[W^{[\xi]}]$ in probability. Since $\mathbb{E}[W^{[\xi]}] \le \sup_{v\in\mathcal{V}}\mathbb{E}[B_v]\lambda_{\sup}$, the result holds.

A.10. Proof of Theorem 4.7

We have $h^{\mathrm{lin}}_{v,d^{[\xi]}}(S^{[\xi]}) = \mu_v + d^{[\xi]}a_{v,d^{[\xi]}}\frac{S^{[\xi]}-\mathbb{E}[S^{[\xi]}]}{d^{[\xi]}}$. Using $a_{v,d^{[\xi]}} = \mathcal{O}(1/d^{[\xi]})$ and Proposition 4.6, the result follows.

B. Factorization of the joint pmf

In the proof of Theorem 2.2, we argued from the thinning-based stochastic representation that any member of the family $\mathbb{MPMRF}$ satisfies the conditional independencies required to make it a tree-structured

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press

MRF as per Definition 2.1. From these conditional independencies, we have directly, by successive conditioning along the tree,

$$p_{\boldsymbol{N}}(\boldsymbol{x}) = \prod_{v\in\mathcal{V}} p_{N_v|\bigcap_{j=1}^{v-1} N_j=x_j}(x_v) = p_{N_r}(x_r) \prod_{v\in\mathcal{V}\setminus\{r\}} p_{N_v|N_{\mathrm{pa}(v)}=x_{\mathrm{pa}(v)}}(x_v), \quad \boldsymbol{x}\in\mathbb{N}^d, \tag{B1}$$

which we recognize as the joint pmf of a Gibbs distribution. One can find the following definition in Koller and Friedman (2009), adapted for discrete random variables.

Definition B.1 (Gibbs distribution). *Let $\mathcal{V}_1,\dots,\mathcal{V}_m$ be subsets of $\mathcal{V}$, with $m\le|\mathcal{V}|$, and define $\phi_1,\dots,\phi_m$ as some functions $\phi_i:\mathbb{R}^{|\mathcal{V}_i|}\to\mathbb{R}$, $i\in\{1,\dots,m\}$. The joint pmf of a vector of discrete random variables $X=(X_v,\ v\in\mathcal{V})$ following a Gibbs distribution admits the representation $p_{\boldsymbol{X}}(\boldsymbol{x})=\frac{1}{Z}\prod_{i=1}^m \phi_i\big((x_v,\ v\in\mathcal{V}_i)\big)$, $\boldsymbol{x}\in\mathbb{N}^d$, where Z is a normalizing constant. A Gibbs distribution factorizes on $\mathcal{G}=(\mathcal{V},\mathcal{E})$ if $\mathcal{V}_1,\dots,\mathcal{V}_m$ are all cliques of $\mathcal{G}$.*

On a tree $\mathcal{T}$, cliques are any vertex or pair of vertices connected by an edge. The joint pmf of $\boldsymbol{N}$ in (2.4) is that of a Gibbs distribution factorizing on $\mathcal{T}$. The Hammersley–Clifford theorem states that a random vector $\boldsymbol{X}=(X_v)_{v\in\mathcal{V}}$ with positive probabilities on $\prod_{v\in\mathcal{V}}\mathrm{supp}(X_v)$ is a MRF defined on $\mathcal{G}$ if and only if it follows a Gibbs distribution factorizing on $\mathcal{G}$. Therefore, to design the family $\mathbb{MPMRF}$, satisfying the conditional independencies of a MRF and Poisson marginals, one could have proceeded directly using the factorized joint pmf in (B1), and, to ensure undirectedness, revert the conditional probabilities to joint probabilities. We have

$$\begin{aligned} p_{\boldsymbol{N}}(\boldsymbol{x}) = p_{N_r}(x_r) \prod_{v\in\mathcal{V}\setminus\{r\}} \frac{p_{N_{\mathrm{pa}(v)},N_v}(x_{\mathrm{pa}(v)},x_v)}{p_{N_{\mathrm{pa}(v)}}(x_{\mathrm{pa}(v)})} &= \frac{p_{N_r}(x_r)}{p_{N_r}(x_r)^{\deg(r)}} \frac{\prod_{(u,v)\in\mathcal{E}} p_{N_u,N_v}(x_u,x_v)}{\prod_{v\in\mathcal{V}\setminus\{r\}} p_{N_v}(x_v)^{\deg(v)-1}} \\ &= \frac{\prod_{(u,v)\in\mathcal{E}} p_{N_u,N_v}(x_u,x_v)}{\prod_{v\in\mathcal{V}} p_{N_v}(x_v)^{\deg(v)-1}}, \quad \boldsymbol{x}\in\mathbb{N}^d, \end{aligned} \tag{B2}$$

where the second equality comes from the fact that every vertex is the parent of $\deg(v)-1$ other vertices, except the root, for which it is $\deg(r)$. The factorization as in (B2) is referred to as the node-edge factorization of a MRF. The model will be undirected and invariant under any rooting of the tree if all bivariate joint pmfs in the multiplicands in (B2) are symmetric in that respect. It is well known to be the case for Poisson random variables joined through binomial thinning: letting $\xi_{u,v}=\alpha_{(u,v)}\sqrt{\lambda_u\lambda_v}$,

$$p_{N_u,N_v}(x_u,x_v) = \mathrm{e}^{-(\lambda_u+\lambda_v-\xi_{u,v})} \sum_{i=1}^{\min(x_u,x_v)} \frac{(\lambda_u-\xi_{u,v})^{m-k}(\lambda_v-\xi_{u,v})^{n-k}\xi_{u,v}^k}{(x_u-k)!(x_v-k)!k!}, \quad x_u,x_v\in\mathbb{N},$$

see McKenzie (1988), which we incidently recognize as the joint pmf of the bivariate Poisson based on common shocks. Substituting the bivariate and marginal pmfs in (B2), one recovers (2.4).

https://doi.org/10.1017/asb.2026.10087 Published online by Cambridge University Press