



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Avanzi, B;Taylor, G;Wong, B;Yang, X

Title:

On the modelling of multivariate counts with Cox processes and dependent shot noise intensities

Date:

2021-04-03

Citation:

Avanzi, B., Taylor, G., Wong, B. & Yang, X. (2021). On the modelling of multivariate counts with Cox processes and dependent shot noise intensities. *INSURANCE MATHEMATICS & ECONOMICS*, 99, pp.9-24. <https://doi.org/10.1016/j.insmatheco.2021.01.002>.

Persistent Link:

<https://hdl.handle.net/11343/324124>

On the modelling of multivariate counts with Cox processes and dependent shot noise intensities

Benjamin Avanzi^a, Greg Taylor^b, Bernard Wong^b, Xinda Yang^{*,b}

^aCentre for Actuarial Studies, Department of Economics, University of Melbourne VIC 3010, Australia

^bSchool of Risk and Actuarial Studies, UNSW Australia Business School, UNSW Sydney NSW 2052, Australia

Abstract

In this paper, we develop a method to model and estimate several, *dependent* count processes, using granular data. Specifically, we develop a multivariate Cox process with shot noise intensities to jointly model the arrival process of counts (e.g. insurance claims). The dependency structure is introduced via multivariate shot noise *intensity* processes which are connected with the help of Lévy copulas. In aggregate, our approach allows for (i) over-dispersion and auto-correlation within each line of business; (ii) realistic features involving time-varying, known covariates; and (iii) parsimonious dependence between processes without requiring simultaneous primary (e.g. accidents) events.

The explicit incorporation of time-varying, known covariates can accommodate characteristics of real data and hence facilitate implementation in practice. In an insurance context, these could be changes in policy volumes over time, as well as seasonality patterns and trends, which may explain some of the relationship (dependence) between multiple claims processes, or at least help tease out those relationships.

Finally, we develop a filtering algorithm based on the reversible-jump Markov Chain Monte Carlo (RJ-MCMC) method to estimate the latent stochastic intensities and illustrate model calibration using real data from the AUSI data set.

Key words: Dependency modelling, Cox process, Shot noise, Insurance claims counts, Micro-level model, Markov chain Monte Carlo

JEL codes: C51, C53, C55, G22

MSC classes: 91G70, 91G60, 62P05, 62H12

1. Introduction

1.1. Background

This paper is concerned with the multivariate modelling of counts, which are important in many actuarial applications. For instance, the distribution of aggregate losses would not only depend on the marginal loss distributions, but also on their dependency structure. Dependencies can manifest themselves in the severity or frequency of claims. By nature, this paper focuses on the latter, since it considers counts only.

The actuarial literature on counts processes is vast. Beyond the landmark work of Lundberg (1909); Cramér (1930, 1955); Bühlmann (1970), early extensions of the classical Poisson process approach include Arjas (1989); Norberg (1993, 1999), who developed a marked point process approach. There, the arrival process of claims follows a continuous time point process and the claim development is captured by a mark. This framework was implemented in Antonio and Plat (2014) and Larsen (2007). Jewell (1989); Zhao, Zhou and Wong (2009); Zhao and Zhou (2010) also studied the use of Poisson process to model claim arrival.

*Corresponding author.

Email addresses: b.avanzi@unimelb.edu.au (Benjamin Avanzi), gregory.taylor@unsw.edu.au (Greg Taylor), bernard.wong@unsw.edu.au (Bernard Wong), xinda.yang@unsw.edu.au (Xinda Yang)

Avanzi, Wong and Yang (2016c) further extended the study to a marked Cox process framework, where a stochastic intensity introduces over-dispersion and an estimation algorithm is developed with the presence of reporting delays. Badescu, Lin and Tang (2016) and Badescu, Lin and Tang (2019) have also adopted a Cox process approach where the intensity follows a hidden Markov model (HMM) with Erlang state-dependent distributions. Avanzi, Taylor, Wong and Xian (2020) model and analyse claim counts of the AUSI dataset using a univariate Markov modulated non homogeneous Poisson process. Discrete time frameworks were also considered (see, for example Pigeon, Antonio and Denuit, 2013, 2014, where numbers of claims follow Poisson distributions). Furthermore, Matsui and Mikosch (2010) establish the framework of using a Poisson cluster model in modelling insurance claim process, which is extended to the case of multiple clusters by Matsui (2015) and to non-homogeneous observations by Matsui (2014). Jessen, Mikosch and Samorodnitsky (2011) adopt a Poisson cluster model for the payment counts and the total loss amounts of claims with real data illustration.

1.2. Motivation and contributions

The availability of granular data (also called “micro-level” data) is an important motivation for the developments in this paper. When aggregate (“macro-level”) data is used, the small(er) size of the data set may lead to high parameter uncertainty and hence reduce the predictive power of the model. In particular, one can reasonably question whether there is enough data to fit a dependency structure (when required) to a reasonable level of confidence. An aggregate dataset by nature may eliminate some information about useful factors (such as, in a reserving context, seasonality factors in annual triangles). Besides, aggregation does not allow for a separate treatment of different components of the claims processes (such as information on the claim arrival process and individual claim development process). Those components may have different effects on the quantity of interest, and may also change over time in different ways. In short, the forces that drive the quantities of interest are hidden behind a black box in a macro-level model, which is a problem when the environment changes.

Two recurrent features that can be commonly observed in insurance data are overdispersion and autocorrelation within loss processes; see, for example, de Jong and Heller (2008) and Denuit, Maréchal, Pitrebois and Walhin (2007). In fact, one could argue that this is likely to arise from frequency rather than severity. This motivates our focus, in this paper, on count processes. The traditional Poisson process approach cannot introduce dependence within marginal processes without a stochastic intensity. Such a process is Cox by definition. This is developed extensively in Avanzi, Wong and Yang (2016c, in a univariate framework).

The Cox approach has further strengths in modelling dependence between processes. In the (multivariate) Poisson approach, dependence between processes can only occur due to simultaneous primary events (e.g. accidents involving 2 cars, or a hailstorm affecting multiple vehicles). While this might be appropriate in very rare circumstances, in general such an assumption is too unrealistic and restrictive, and hence will be unable to capture dependence not involving directly observable primary events. The Cox approach relaxes this restriction by modelling dependence via stochastic *intensity* processes instead. In this paper, we further take advantage of the Lévy copula approach to introduce dependence across processes through that stochastic intensity. In so doing, we achieve dependence *within* and *between* processes in a parsimonious way (a Lévy copula might involve just one parameter). The challenge, however, is to fit such a model to a dataset, as the intensity is not observed (only counts). In this paper, we develop a complete methodology for fitting by developing an EM algorithm with a reversible jump Markov (“RJMCMC”) filter.

We illustrate our approach with real data. To better match the characteristics of real data, our approach also explicitly incorporates time-varying, known covariates. In an insurance context, these could be changes in policy volumes over time, as well as seasonality patterns and trends, which may explain some of the relationship between multiple claims processes (or at least help tease out those relationships). Such allowances are vital to enable practical implementation, as such covariates are readily observable in real data.

Remark 1.1. *A prominent example of “micro-level” approaches lies in the area of reserving, which may advantageously use the information of individual claims (Antonio and Plat, 2014; Avanzi, Wong and Yang, 2016c). Typical micro-level datasets include information such as inception and expiry dates of each policy,*

the arrival and reporting dates of each claim and the history of transaction and settlement (if the claim has been settled) of each claim. Such data sets are very granular and complex. They allow the observations of varying components, such as the arrival and reporting dates as well as the transaction records and settlement status of each claim. Therefore, one can try to build a dependency model with micro-level components to explain the dependencies of the reserving estimate. We will build on this example in the illustration part of this paper. Note, however, that while the reserving application is used as a prominent example in this paper, a complete micro-level reserving model framework would include not only projection of ultimate claim frequencies, but also reporting delay distributions, claim severities with claim development patterns. The methodology of introducing and estimating the reporting delays has been studied in Avanzi, Wong and Yang (2016c). Under the assumption of independence between cross Line of Business (“LoB”) reporting delays (which we believe to be a good assumption for attritional claims as modelled in this paper), the introduction of reporting delay patterns in the multivariate context is essentially the same as that in the univariate context, where an independent reporting delay distribution is chosen for each process. Thus the estimation would follow the same procedures as that in Avanzi, Wong and Yang (2016c).

1.3. Structure of the paper

In Section 2 we set out our multivariate counts process. Dependence is introduced with the help of a Lévy copula, which is applied on the shot noise intensity processes of the marginal counts processes. The estimation (and prediction) of the model parameters is addressed in Section 3, where we develop an EM algorithm with a reversible jump Markov (“RJMCMC”) filter. This is where the availability of granular data is critically useful. Section 4 presents a comprehensive illustration using a bivariate motor insurance sub-set of the the AUSI data set. We show how the data is detrended by accounting for exposure and further covariate adjustments. The multivariate claim arrival process is built in a bottom up fashion by “assembling” two marginal claims arrival processes with the help of a Clayton Lévy copula. The final model is used to perform predictions and results are thoroughly discussed. Section 5 concludes.

2. A multivariate claim arrival process

In this section, we develop our multivariate count model. Section 2.1 introduces a multivariate Cox process model with common shock intensities. Section 2.2 provides an alternative bottom-up construction of a common shock model with a Lévy copula approach.

2.1. A Cox process model with common shocks

Let $\mathbf{N} = \{(N_1(t), \dots, N_G(t)), t \geq 0\}$ denote the multivariate claim arrival process. The g^{th} marginal process, $\{N_g(t), t \geq 0\}$, is a counting process for the number of claims of the g^{th} ($1 \leq g \leq G$) counts process. We use the definition of the multivariate Cox process from Møller and Waagepetersen (2004), and assume that \mathbf{N} is a multivariate Cox process driven by a multivariate intensity $\tilde{\lambda} = \left\{ \left(\tilde{\lambda}_1(t), \dots, \tilde{\lambda}_G(t) \right), t \geq 0 \right\}$ (see Definition 2.1.1).

Definition 2.1.1 (A multivariate Cox process). (see, for example, Møller and Waagepetersen, 2004) Suppose that $\tilde{\lambda}_g = \{\tilde{\lambda}_g(\xi) : \xi \in S\}$, $g = 1, \dots, G$, are non-negative random fields so that for $g = 1, \dots, G$, $\xi \rightarrow \tilde{\lambda}_g(\xi)$ is a locally integrable function (that is, $\tilde{\lambda}_g(\xi)$ is integrable over the space of events) with probability one (where ξ and S refer to the event and the space of events). Conditional on $\tilde{\lambda} = \left\{ \left(\tilde{\lambda}_1(t), \dots, \tilde{\lambda}_G(t) \right), t \geq 0 \right\}$, suppose that $\{N_g(t), t \geq 0\}_{g=1, \dots, G}$ are independent Poisson processes with intensity functions $\{\tilde{\lambda}(t), t \geq 0\}_{g=1, \dots, G}$ respectively. Then $\mathbf{N} = \{(N_1(t), \dots, N_G(t)), t \geq 0\}$ is said to be a multivariate Cox process driven by $\tilde{\lambda}$.

The dependency structure of a multivariate Cox process results from the multivariate intensity process. Here the multivariate intensity is influenced by the joint movement of the underlying risk regimes of multiple counts processes. Furthermore, the conditional independence of the marginal processes assumption provides

a simple mathematical representation and is able to account for most of the dependency across general insurance claims.

As in Avanzi, Wong and Yang (2016c) (in a univariate framework), we model the marginal intensity processes of $\boldsymbol{\lambda}$ as shot noise processes. This allows us to introduce a dependency structure via common shocks (common shots) across multiple stochastic intensities.

Definition 2.1.2 (A multivariate shot noise process). (*Li, 2002*) A multivariate stochastic process, $\tilde{\boldsymbol{\lambda}} = \left\{ \left(\tilde{\lambda}_1(t), \dots, \tilde{\lambda}_G(t) \right), t \geq 0 \right\}$, is a multivariate shot noise process if

$$\tilde{\lambda}_g(t) = \tilde{\lambda}_g(0)e^{-\kappa_g t} + \sum_{j=1}^{J(t)} X_{g,j} e^{-\kappa_g(t-\tau_j)}, \quad t \geq 0, \quad (2.1)$$

where $\{J(t), t \geq 0\}$ is a homogeneous Poisson process of intensity ρ (with $\rho > 0$); τ_j represents the arrival time of the j^{th} event of $\{J(t), t \geq 0\}$, which triggers a jump over the shot noise process $\left\{ \tilde{\lambda}_g(t), t \geq 0 \right\}$. The size of the jump is denoted by $X_{g,j}$ (where the subscripts refer to the g^{th} marginal shot noise process and j^{th} shot), which follows an exponential distribution with an expected value of $1/\eta_g$ if $X_{g,j} > 0$. We assume that the sequence of the G -dimensional random variable $\{\mathbf{X}_j = (X_{1,j}, \dots, X_{G,j})\}$ are independent and identically distributed over $j = 1, \dots, J(t), t \geq 0$. Furthermore, $X_{g_1,j}$ and $X_{g_2,j}$ ($g_1, g_2 = 1, \dots, G$) can be dependent for all $j = 0, \dots, J(t)$. Denote by $f_{\mathbf{X}}$ the density function of \mathbf{X}_j . The speed of decay is measured by the parameter κ_g . Note $f_{\mathbf{X}}$ can consist of a mixture of continuous and discrete components.

Definition 2.1.2 introduces a common shock dependency structure across multiple shot noise processes. It is worth mentioning that one or more marginals of \mathbf{X}_j ($j = 1, \dots, J(t)$) can be 0. In light of this, one can decompose the arrival of shots of a marginal shot noise process into different subsets, where each subset corresponds to a unique combination of non-zero shots on this marginal stochastic intensity process. In a bivariate context, for example, there can be up to three subsets, where shots can affect either exclusively on one marginal intensity process or simultaneously on both marginal intensity processes. We have provided an example (see Example 1 below) to illustrate how such a decomposition can be achieved and presented the covariance between the marginal shot noise intensities at time t ($t > 0$).

Remark 2.1. Here $J(t)$ triggers a multivariate jump, and it is possible to have only one marginal being non-zero (in which case it is a unique jump) or more than one marginal being non-zero (hence a common jump). Hence we write J and not of J_g . This is also why $f_{\mathbf{X}}$ does not depend on g , since it is a G -dimensional random variable. Furthermore, $X_{g_1,j}$ and $X_{g_2,j}$ can be dependent for all g_1 and g_2 .

Remark 2.2. Interested readers can also refer to Selch and Scherer (2018) for an alternative way of defining a multivariate Cox process. In Selch and Scherer (2018), a common stochastic clock is shared by all marginal processes. This requires a different interpretation of the model as well as different estimation techniques.

Example 1 (A bivariate example). A bivariate shot noise process,

$$\tilde{\boldsymbol{\lambda}}(t) = (\tilde{\lambda}_1(t), \tilde{\lambda}_2(t)), \quad (2.2)$$

can be expressed as

$$\begin{aligned} \tilde{\lambda}_1(t) &= \tilde{\lambda}_1(0)e^{-\kappa_1 t} + \sum_{j=1}^{J_1^\perp(t)} X_{1,j}^\perp e^{-\kappa_1(t-\tau_{1,j}^\perp)} + \sum_{j=1}^{J_{12}^\parallel(t)} X_{1:12,j}^\parallel e^{-\kappa_1(t-\tau_{12,j}^\parallel)} \\ \tilde{\lambda}_2(t) &= \tilde{\lambda}_2(0)e^{-\kappa_2 t} + \sum_{j=1}^{J_2^\perp(t)} X_{2,j}^\perp e^{-\kappa_2(t-\tau_{2,j}^\perp)} + \sum_{j=1}^{J_{12}^\parallel(t)} X_{2:12,j}^\parallel e^{-\kappa_2(t-\tau_{12,j}^\parallel)} \end{aligned} \quad (2.3)$$

where $J_g^\perp(t)$ ($g = 1, \dots, G$) is a homogeneous Poisson process (of intensity ρ_g^\perp) that triggers shots of size $X_{g,j}^\perp$ (with distribution function F_g^\perp) and arrival time $\tau_{g,j}^\perp$, only on the g^{th} marginal process (for $g = 1, 2$); and

$J_{12}^{\parallel}(t)$ is a homogeneous Poisson process (of intensity ρ_{12}^{\parallel}) that affects both marginal processes simultaneously with a bivariate shot of size $(X_{1:12}^{\parallel}, X_{2:12}^{\parallel})$ (with a bivariate distribution function F_{12}^{\parallel} of marginals $F_{1:12}^{\parallel}$ and $F_{2:12}^{\parallel}$) and arrival time $\tau_{12,j}^{\parallel}$, for $j = 1, \dots, J_{12}^{\parallel}(t)$. Note that we have omitted the unique shots of $\tilde{\lambda}_2(t)$ in the expression of $\tilde{\lambda}_1(t)$ for the ease of notation and also the benefit that F_1 is not necessarily a mixed distribution. Equivalently, one can adopt the convention in Definition 2.1.2, in which case some shots can be 0 for a marginal process and hence the distribution consists of a mass at 0.

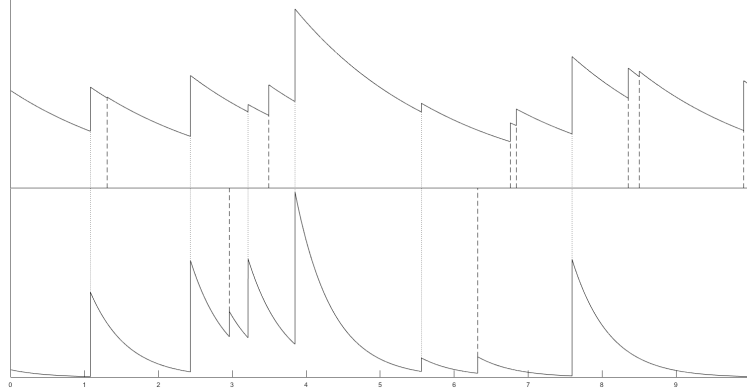


Figure 1: An illustration of a bivariate shot noise process: dotted lines - common shots, dashed lines - unique shots.

Figure 1 illustrates a realisation of the path of a bivariate shot noise process. The dotted lines indicate common shots which trigger jumps in both marginal shot noise processes. Moreover, the dashed lines indicate unique shots, where each shot only triggers jumps in one marginal process.

Furthermore,

$$\text{Cov}(\tilde{\lambda}_1(t), \tilde{\lambda}_2(t)) = \frac{\rho_{12}^{\parallel} \mathbb{E}[X_{1:12}^{\parallel} X_{2:12}^{\parallel}]}{\kappa_1 + \kappa_2} \quad (2.4)$$

where $(X_{1:12}^{\parallel}, X_{2:12}^{\parallel})$ follows a bivariate distribution function F_{12}^{\parallel} ; see Appendix A for a proof. Finally, the marginal process $\tilde{\lambda}_1(t)$ is a univariate shot noise process, that is,

$$\tilde{\lambda}_1(t) = \tilde{\lambda}_1(0)e^{-\kappa_1 t} + \sum_{j=1}^{J_1} X_{1,j} e^{-\kappa_1(t-\tau_{1,j})}, \quad (2.5)$$

where J_1 is a homogeneous Poisson process of intensity

$$\rho_1 = \rho_1^{\perp} + \rho_1^{\parallel}, \quad (2.6)$$

and where the size of shot $X_{1,j}$ follows the distribution function

$$F_1 = \frac{\rho_1^{\perp}}{\rho_1} F_1^{\perp} + \frac{\rho_1^{\parallel}}{\rho_1} F_{1:12}^{\parallel}. \quad (2.7)$$

One can further generalise Example 1 to higher dimensions in a similar manner. In general, there can be up to $2^G - 1$ subsets for a G -dimension shot noise process.

The frequencies of insurance claims are also driven by risk exposures. The exposure an insurance process faces depends on factors such as the volumes of business, and can sometimes be measured by the number of policies in force, for instance. Furthermore, there are also trends and seasonal patterns that affect the frequencies of claims. In a tropical region, for example, one would expect a higher number of claims in summer for a housing insurance product than winter. We incorporate risk exposure into the multivariate Cox process \mathbf{N} through $\{W_g(t)\}_{g=1,\dots,G}$, where $W_g(t)$ is the risk exposure of the g^{th} count process at time t ; see also Section 4.1.

Model assumption 2.1.3 (A non-stationary multivariate shot noise intensity). *We assume that there exists a stationary multivariate shot noise process $\tilde{\lambda}(\mathbf{t})$ such that for $g = 1, \dots, G$*

$$\lambda_g(t) = W_g(t)\tilde{\lambda}_g(t). \quad (2.8)$$

Assumption 2.1.3 allows for a multivariate intensity process for the arrivals. The common shock structure can be interpreted as common adverse events that affect multiple counts processes at different scales. Furthermore, we assume that the stochastic intensity of arrivals is proportional to the risk exposure of the corresponding counts process, which leads to a non-stationary shot noise process. Here a stationary shot noise process is a process that follows Definition 2.1.2, where the joint distribution of the process does not depend on t , and the only source of non-stationarity is introduced through the risk exposure.

Remark 2.3. *As the exposure process is non-stationary and the underlying intensity process is unobservable, the analysis of empirical moments (including autocorrelations) may be distorted. As such, direct evaluation of whether a Cox model is a good candidate is difficult. Instead, in this paper the choice of candidate is driven by theoretical properties and goodness-of-fit analysis of the model (as illustrated later in this paper with residual analysis). In our illustration, we investigated and were able to explain the non-stationary patterns of the claim intensity process resulting from exposure, seasonal patterns and trends.*

2.2. A bottom-up approach of common shot constructions with Lévy copulas

Definition 2.1.2 defines a general common shock model. This allows for dependency construction of higher dimensions, which is an extension of Example 1. However, such a common shock representation of the dependency structure faces several challenges. Firstly, the practical application of counts processes in actuarial contexts (more generally, compound loss processes) typically adopts a bottom-up approach in model construction and calibration. Such an approach involves the modelling and estimation of the marginal processes before a dependency structure is introduced (see page 221, McNeil, Frey and Embrechts, 2015, for the definition of a bottom-up approach). However, this is not directly available to a common shock model, where the specification of the dependency structure is *implied* in a multivariate set-up. Secondly, the direct construction of a common shock model may lead to complexity of model assumptions and a high number of parameters. This is particularly relevant at a high dimension where there will be a number of $2^G - 1$ arrival processes of shots. Furthermore, each common shock process requires a dependency model of the joint jump sizes, which effectively requires a number of $2^G - G - 1$ dependency structures (further discussion of the use of common shock models can be found in, e.g., Lindskog and McNeil, 2003). This is clearly not practical, and because the joint intensity process is unobservable, over-parametrisation may further lead to higher parameter errors and model errors.

Fortunately, one can achieve an identical outcome in parsimonious way thanks to Lévy copulas; refer to Tankov (2003) and Avanzi, Cassar and Wong (2011, in the actuarial literature) for the formal definition and discussion for the properties of Lévy copulas. A Lévy copula model typically includes only a small number of parameters (sometimes just one) and hence leads to a generally parsimonious model specification. Furthermore, a Lévy copula approach separates the model specification of the marginal processes and the dependency structure, which allows for a bottom-up approach in model construction. In Section 3, we will further explain how one can separate the calibration of the marginal components and the dependency component. For all those reasons we adopt the Lévy copula approach, which provides a practical and parsimonious approximation to the underlying joint intensities.

Under Definition 2.1.2, a large range of Lévy copulas can specify the dependency structure across the marked Poisson process (that is, arrival of shot *and* the associated jump sizes). Such a dependency relationship drives the dependency across the marginal Cox processes. A real world interpretation of this mechanism is that dependency of multiple insurance count processes do not arrive directly from common arrival of claims - rather, it is the dependency in the underlying risk generating regime (e.g. the intensity processes in this paper) that creates the dependency of claim counts. Continuing on Example 1, Example 2 introduces the bivariate Clayton Lévy copula, which we will use in Section 4.

Example 2. A bivariate Lévy copula \mathfrak{C} can be used to couple the marginal processes together via

$$U(x_1, x_2) = \mathfrak{C}(U_1(x_1), U_2(x_2)), \quad (2.9)$$

where $U_g(x_g) = \rho_g(1 - F_g(x_g))$ denotes the g^{th} tail integral of the marginal process ($g = 1, 2$) and $U(x_1, x_2) = \rho_{12}(1 + F_{12}^{\parallel}(x_1, x_2) - F_{12}^{\parallel}(x_1, \infty) - F_{12}^{\parallel}(\infty, x_2))$ denotes the bivariate tail integral of the joint bivariate process.

A possible choice for \mathfrak{C} is the bivariate Clayton Lévy copula (Cont and Tankov, 2004), which is defined as

$$\mathfrak{C}(u_1, u_2) = (u_1^{-\delta} + u_2^{-\delta})^{-1/\delta}, \quad \delta > 0. \quad (2.10)$$

It is worth emphasizing that, while there is a relationship between the copula of the common jumps induced by a Clayton Lévy copula, and the Clayton distributional copula, a Clayton Lévy copula is not a Clayton copula. A Clayton copula is a copula function for a multivariate random variables and a Clayton Lévy copula (which is not a copula function) aggregates multiple Lévy processes.

3. Parameter estimation and prediction

In this section, we explain how one can estimate the parameters of the multivariate shot noise model. This is particularly essential with a Cox process approach, where the joint intensity is unobservable and hence the model cannot be calibrated with a maximum likelihood approach. Furthermore, the estimation of the parameters and the underlying intensity are necessary for the projection of future claims counts (e.g., to help with the valuation of future claims liabilities). For the rest of this section, we use a trivariate Cox process as an example.

We start by introducing notation and likelihood in Section 3.1, before developing an EM algorithm with a reversible jump Markov chain Monte Carlo (“RJMCMC”) filter. We extend the estimation algorithm in Avanzi, Wong and Yang (2016c) to allow for a multivariate shot noise process. We explain our filtering algorithm with fixed parameters in Section 3.2 and develop a three-step EM algorithm to update the parameter estimates in Section 3.3. For simplicity, we illustrate the filtering and parameter estimation procedures in a trivariate context. These can be adapted in higher dimensions with a similar procedure.

3.1. Notation and likelihood

Suppose that the overall observation period (that is, the period for which which policy and claim data are extracted) is $[0, T]$, which is discretised into a number of L sub-periods of equal length Δ . We say that a claim arrives in the i^{th} accident period if the arrival time of the claim falls into $((i-1)\Delta, i\Delta]$. This set of observations is denoted

$$\mathbf{N}_{\mathbf{D}}^{\mathbf{G}} = \{N_{g,i}; g = 1, 2, 3, i = 1, 2, \dots, L\}. \quad (3.1)$$

Furthermore, we characterise the trajectory of the trivariate shot noise process by

$$\boldsymbol{\theta}_{\mathbf{G}} = \{(\tilde{\lambda}_1(0), \tilde{\lambda}_2(0), \tilde{\lambda}_3(0)), \tau_1, \dots, \tau_n, \mathbf{X}_1, \dots, \mathbf{X}_n\} \quad (3.2)$$

where τ_i and \mathbf{X}_i denote the arrival time and trivariate sizes of the i^{th} shot, respectively. As mentioned in the explanation for Definition 2.1.2, each \mathbf{X}_i is a trivariate random variable where one or more marginal can be 0. Note that we are modelling the *intensity* of the claim count process here (not the counts themselves), so that all components of $\boldsymbol{\theta}_{\mathbf{G}}$ are unobservable.

In the rest of this section, we will derive the log-likelihood functions for the trajectory of a trivariate shot noise process and the conditional observations of claim counts given the shot noise trajectory.

Denote by $\boldsymbol{\delta}$ the parameter vector of the Lévy copula. Furthermore, denote by ρ_j ($j = 1, 2, 3$) the intensity of the Poisson process of the non-zero shots of the j^{th} margin. Note that ρ_j ($j = 1, 2, 3$) is the

thinned Poisson process (conditional on positive shots) of the j^{th} margin of the trivariate Poisson arrival of shots with intensity ρ (see Definition 2.1.2). The log-likelihood of $\boldsymbol{\theta}_{\mathbf{G}}$ is

$$\begin{aligned} & \log p(\boldsymbol{\theta}_{\mathbf{G}}; \rho_1, \rho_2, \rho_3, \eta_1, \eta_2, \eta_3, \kappa_1, \kappa_2, \kappa_3, \boldsymbol{\delta}) \\ &= n \log \rho - \rho T + \log f_{\mathbf{X}}(\mathbf{X}_{\mathbf{n}}) \\ &+ \sum_{g=1}^3 \left[\left(\frac{\rho_g}{\kappa_g} - 1 \right) \log \tilde{\lambda}_g(0) - \eta_g \tilde{\lambda}_g(0) + \frac{\rho_g}{\kappa_g} \log \eta_g - \log \left(\Gamma \left(\frac{\rho_g}{\kappa_g} \right) \right) \right], \end{aligned} \quad (3.3)$$

where each term in the summation refers to the log-likelihood of the corresponding initial value of a marginal shot noise process, which is chosen to be the corresponding stationary distribution of each univariate process, that is, a Gamma random variable (see Centanni and Minozzo, 2006a,b), and where $\Gamma(\cdot)$ is the Gamma function. Conditional on the trajectory of the multivariate shot noise process, the conditional log-likelihood of our observations is

$$\begin{aligned} & \log L(\mathbf{N}_{\mathbf{D}}^{\mathbf{G}} | \boldsymbol{\theta}_{\mathbf{G}}; \kappa_1, \kappa_2, \kappa_3) \\ &= \sum_{g=1}^3 \sum_{i=1}^L \log A(N_{g,i}, M_{g,i}; \kappa_g) \\ &= \sum_{g=1}^3 \sum_{i=1}^L (-M_{g,i} + N_{g,i} \log M_{g,i}) + \text{constant} \end{aligned} \quad (3.4)$$

where

$$\begin{aligned} M_{g,i} &= \int_{(i-1)\Delta}^{i\Delta} \lambda_g(t) dt, \\ A(N_{g,i}, M_{g,i}; \kappa_g) &= \frac{e^{-M_{g,i}} M_{g,i}^{N_{g,i}}}{N_{g,i}!} \end{aligned} \quad (3.5)$$

Here the function $A(m, M_{g,i}; \kappa_g)$ calculates the probability of having m ultimate claims, which comes from a Poisson probability mass function of intensity $M_{g,i}$. This function depends on κ_g through the integral of $\lambda_g(t)$. Note that the format of the conditional likelihood function (3.4) results from the discretisation scheme introduced at the beginning of this section. Such a discretisation scheme accommodates the discrete nature of real data, which is an essential step in applying a continuous time Cox process (see also Avanzi, Wong and Yang, 2016c).

3.2. A RJMCMC filtering algorithm

The issue we face is to derive the conditional distribution of the unobservable component, $\boldsymbol{\theta}_{\mathbf{G}}$, given the knowledge of $\mathbf{N}_{\mathbf{D}}^{\mathbf{G}}$ (which is observable), which is a filtering problem (Centanni and Minozzo, 2006a,b). Since we can obtain the likelihood of the shot noise process itself—see Equation (3.3)—and the likelihood of the conditional distribution of $\mathbf{N}_{\mathbf{D}}^{\mathbf{G}}$ given $\boldsymbol{\theta}_{\mathbf{G}}$ —see Equation (3.4), we can adopt a MCMC algorithm. However, since the number of shots is unknown, the dimension of $\boldsymbol{\theta}_{\mathbf{G}}$ is also undetermined. We thus adopt a RJMCMC algorithm (Green, 1995) to allow for ‘moves’ with dimension changing. We will briefly outline the general steps of a RJMCMC simulation algorithm. One can refer to Gelman, Jones, Brooks and Meng (2011, Chapter 3) for more details about the RJMCMC simulation and Centanni and Minozzo (2006a,b) and Avanzi, Wong and Yang (2016c) regarding using a RJMCMC filter in the univariate case of a shot noise Cox process.

A RJMCMC simulation algorithm helps approximate the conditional distribution of the shot noise process given the observations. Firstly, one randomly chooses a move type with probability $p(r|n)$ (with $\sum_r p(r|n) = 1$), which depends on the existing number of shots. Given a chosen move type r and the existing shot noise process $\boldsymbol{\theta}_{\mathbf{G}}$, one proposes a new shot noise trajectory by generating a random component \mathbf{u} with a proposal

distribution $q(\mathbf{u}|r, n, \boldsymbol{\theta}_{\mathcal{G}})$. This proposal, once accepted, results in $\boldsymbol{\theta}'_{\mathcal{G}}$ with n' shots. The probability of accepting the proposal is $\min(1, \alpha[(n, \boldsymbol{\theta}_{\mathcal{G}}), (n', \boldsymbol{\theta}'_{\mathcal{G}})])$, where $\alpha[(n, \boldsymbol{\theta}_{\mathcal{G}}), (n', \boldsymbol{\theta}'_{\mathcal{G}})]$ is calculated according to (3.6) and can be further decomposed into the product of the likelihood ratio, prior ratio, proposal ratio, and Jacobian.

$$\alpha[(n, \boldsymbol{\theta}_{\mathcal{G}}), (n', \boldsymbol{\theta}'_{\mathcal{G}})] = \min \left\{ 1, \underbrace{\frac{L(\mathbf{N}_{\mathcal{D}}^{\mathcal{G}}|\boldsymbol{\theta}'_{\mathcal{G}})}{L(\mathbf{N}_{\mathcal{D}}^{\mathcal{G}}|\boldsymbol{\theta}_{\mathcal{G}})}}_{\text{likelihood ratio}} \times \underbrace{\frac{p(\boldsymbol{\theta}'_{\mathcal{G}})}{p(\boldsymbol{\theta}_{\mathcal{G}})}}_{\text{prior ratio}} \times \underbrace{\frac{p(r'|n') q(\mathbf{u}'|r', n', \boldsymbol{\theta}'_{\mathcal{G}})}{p(r|n) q(\mathbf{u}|r, n, \boldsymbol{\theta}_{\mathcal{G}})}}_{\text{proposal ratio}} \times \underbrace{\left| \frac{\partial f_{r,n}(\boldsymbol{\theta}_{\mathcal{G}}, \mathbf{u})}{\partial(\boldsymbol{\theta}_{\mathcal{G}}, \mathbf{u})} \right|}_{\text{Jacobian}} \right\}. \quad (3.6)$$

We introduce five move types, namely s, p, h, b and d. For simplicity, we choose $p(r|n) = 0.2$ for $r = s, p, h, b, d$ and $n > 1$. Furthermore, we assume that $p(s|0) = p(b|0) = 0.5$, that is, a move can only be either of type s or type b (with equal probabilities) if there is no existing shot. The algorithm is similar to that in Centanni and Minozzo (2006a,b) and Avanzi, Wong and Yang (2016c), except that we now look at the multivariate case. Furthermore, we adopt the dash symbol ('') for all variables related to the proposed shot noise trajectory (e.g. n' refers to the number of shots in the proposed state).

We start by introducing three moves that do not involve dimension changing, namely move s, p and h. In other words, we have $n = n'$.

Move s proposes to change the initial values of the multivariate shot noise process. For each of the g^{th} marginal shot noise process ($g = 1, 2, 3$), the new initial value, $(\tilde{\lambda}'_g(0))$, is drawn from the Gamma distribution with parameters $(\rho_g/\kappa_g, \eta_g)$. The prior ratio is calculated as

$$\prod_{g=1}^3 e^{-\eta_g(\tilde{\lambda}'_g(0) - \tilde{\lambda}_g(0))} \left(\frac{\tilde{\lambda}'_g(0)}{\tilde{\lambda}_g(0)} \right)^{\frac{\rho_g}{\kappa_g} - 1} \quad (3.7)$$

and the proposal ratio as

$$\prod_{g=1}^3 e^{-\eta_g(\tilde{\lambda}_g(0) - \tilde{\lambda}'_g(0))} \left(\frac{\tilde{\lambda}_g(0)}{\tilde{\lambda}'_g(0)} \right)^{\frac{\rho_g}{\kappa_g} - 1}. \quad (3.8)$$

Move p proposes to change the position of an existing shot. This involves choosing an existing shot by generating a value, denoted by n^* , from the discrete uniform distribution over $\{1, \dots, n\}$. Then the position of the n^{th} shot, τ_{n^*} , is proposed to be changed to τ'_{n^*} . The proposal follows a continuous uniform random distribution over $(\tau_{n^*-1}, \tau_{n^*+1})$ where $\tau_0 = 0$ and $\tau_{n+1} = T$. The prior ratio of this move is 1, since the location of the shot is irrelevant to the prior likelihood (see Equation (3.3)). Furthermore, one can also show that the proposal ratio is 1, which is due to the use of the uniform distribution in the proposal.

Move h proposes to change the size (height) of an existing shot. Similar to move p, move h requires selecting an existing shot n^* from the discrete uniform distribution over $\{1, \dots, n\}$. Then the size of the n^{th} shot, \mathbf{X}_{n^*} , is proposed to be changed to \mathbf{X}'_{n^*} . The proposal follows the mixed density function (2.7). The prior ratio is

$$\frac{f_{\mathbf{X}}(\mathbf{X}'_{n^*})}{f_{\mathbf{X}}(\mathbf{X}_{n^*})} \quad (3.9)$$

and the proposal ratio is

$$\frac{f_{\mathbf{X}}(\mathbf{X}_{n^*})}{f_{\mathbf{X}}(\mathbf{X}'_{n^*})}. \quad (3.10)$$

One may notice that the product of the proposal ratio and prior ratio is always 1 for moves s, p and h. This is because we choose the proposal distribution to follow the prior knowledge of the shot noise trajectory (given a fixed number of shots). Furthermore, the Jacobian is always 1 for these three moves since the dimension of $\boldsymbol{\theta}_{\mathcal{G}}$ does not change.

Now we introduce two moves that change the dimension of θ_G . Firstly, we have move b that gives birth to a new shot. This involves drawing a new position τ^* from a continuous uniform distribution over $[0, T]$. Denote by n^* the integer such that $\tau_{n^*-1} < \tau^* < \tau_{n^*+1}$ with $\tau_0 = 0$ and $\tau_{n+1} = T$. In this case, we have $n' = n + 1$. Furthermore, the size of this shot, \mathbf{X}'_{n^*} , is simulated from the (mixed) density function $f_{\mathbf{X}}$. Furthermore, we have a move type d that delete an existing shot. This includes drawing n^* from a discrete uniform distribution from $\{1, \dots, n\}$ and the n^{*th} shot is deleted. Then we have $n' = n - 1$.

For move b, the prior ratio is

$$\rho f_{\mathbf{X}}(\mathbf{X}'_{n^*}), \quad (3.11)$$

and the proposal ratio of move b is

$$\frac{p(d|n+1)}{p(b|n)} \frac{(1+n)^{-1}}{T^{-1} f_{\mathbf{X}}(\mathbf{X}'_{n^*})}. \quad (3.12)$$

For move d, the prior ratio is

$$(\rho f_{\mathbf{X}}(\mathbf{X}_{n^*}))^{-1}, \quad (3.13)$$

and the proposal ratio is

$$\frac{q(b|n-1)}{q(d|n)} \frac{T^{-1} f_{\mathbf{X}}(\mathbf{X}_{n^*})}{n^{-1}}. \quad (3.14)$$

One can show that the Jacobian for moves b and d is still 1. The likelihood ratio involved in (3.6) can be obtained from (3.4). It is worth mentioning that this is based on the observations of \mathbf{N}_D^G and hence involves the (non-stationary) risk exposure (as part of calculating M_l for $l = 1, \dots, L$).

We have summarised the various move types in Table 1.

move type ('r')	proposal of the next state	$p(r n)$
s	modifying the initial value of $\left(\left(\tilde{\lambda}'_1(0), \tilde{\lambda}'_2(0), \tilde{\lambda}'_3(0)\right)\right)$ by drawing $\left(\tilde{\lambda}'_g(0)\right)$ from the stationary distribution of the g^{th} marginal ($g = 1, 2, 3$)	0.5 ($n = 0$), 0.2 ($n > 0$)
b	generating a new shot by drawing a new position τ^* uniformly from $(0, t]$ and drawing a new jump height \mathbf{X}'_{n^*} from the shot size distribution $f_{\mathbf{X}}$	0.5 ($n = 0$), 0.2 ($n > 0$)
h	changing the height of a shot by drawing j from the discrete uniform distribution over $\{1, \dots, n\}$ and drawing \mathbf{X}'_j from the shot size distribution $f_{\mathbf{X}}$ to replace \mathbf{X}_j	0 ($n = 0$), 0.2 ($n > 0$)
p	changing the position of a shot by drawing j from the discrete uniform distribution over $\{1, \dots, n\}$ and drawing a new position, τ'_j , uniformly over (τ_{j-1}, τ_{j+1}) (where $\tau_0 = 0$ and $\tau_{n+1} = t$)	0 ($n = 0$), 0.2 ($n > 0$)
d	deleting a shot by drawing j from the discrete uniform distribution over $\{1, \dots, n\}$ and deleting the j^{th} shot	0 ($n = 0$), 0.2 ($n > 0$)

Table 1: Types of moves

3.3. Parameter estimation and prediction

An Expectation Maximisation (“EM”) algorithm will iteratively update the parameter estimates with the presence of incomplete observations (see Rydén, 1996, for more details). In particular, a Monte Carlo Expectation Maximisation (“MCEM”) algorithm is adopted where the conditional expectation in the E-step is approximated through simulations. Avanzi, Wong and Yang (2016c) has used an MCEM algorithm to estimate the shot noise parameters in a univariate scenario. In this case, we have further extended the algorithm to the multivariate case. In particular, we follow the idea of the Inference Functions for Margins (“IFM”) in the context of copula fitting (see Chapter 10 of Joe, 1997, for more details) and separate the estimation of the parameters of the univariate components and the dependency structure.

Firstly, one starts with calibrating the marginal Cox processes. For each marginal process, we follow the algorithm of Avanzi, Wong and Yang (2016c) to estimate the shot noise parameters. Secondly, we proceed to estimating the parameters of the Lévy copula with a MCEM algorithm while the parameters of the marginal processes are fixed. We will explain the full details of this MCEM algorithm in the rest of this section.

The initial estimates of the Lévy copula parameters γ , denoted by γ^0 , are estimated via moment matching based on Equation (2.4) (see Example 1). Such a moment matching estimation can be applied to a bivariate Lévy copula and more generally a higher-order nested Archimedean Lévy copula. With a higher-order nested Archimedean Lévy copula, one can calculate the covariances between claim frequencies, which can be used to calculate the parameters given the corresponding bivariate marginal Lévy copulas (see Avanzi, Tao, Wong and Yang, 2016a).

The initial estimate of the multivariate trajectory of the unobservable shot noise process is obtained based on the estimates of the marginal intensities. The latest estimate of each marginal shot noise is treated as a multivariate shot noise where the other marginals are 0. Therefore one can combine all the marginal estimates and create a multivariate shot noise where each shot is always unique. Furthermore, shots have been merged as long as the arrival times of two shots are less than a threshold (chosen as 0.01 of a day in this project), where the size of the resulting shot is simply the addition of the sizes two individual shots. Such a procedure creates a reasonable initial estimate that utilises the results of the univariate estimation. An illustration of the above procedure in a bivariate case is provided in Example 2.

Example 3. *In Table 2 we illustrate how initial estimates of a bivariate trajectory of the unobservable shot noise process can be obtained based on two marginal intensities. Here, the threshold used is 0.5.*

arrival time	size of shot
0	1
0.7	1.5
2.8	1.2

Marginal 1

arrival time	size of shot
0	2
1.5	1
3.2	1.9

Marginal 2

arrival time	size of shots	
	Marginal 1	Marginal 2
0	1	2
0.7	1.5	0
1.5	0	1
3.2	1.2	1.9

Initial estimate of the bivariate trajectory

Table 2: Example 3; illustration of the initial guess of a bivariate shot noise trajectory

Once the initial estimates are obtained, one can follow the following procedure:

- in the k^{th} iteration, generating a large number of RJMCMC iterations given γ (where γ^0 refer to the initial estimates) based on the algorithm developed in Section 3.2;
- approximating the conditional expectation

$$Q(\gamma, \gamma^k) = \mathbb{E}_{\gamma^k}[\log \mathcal{L}(N_D^G, \theta_G; \gamma) | N_D]. \quad (3.15)$$

as an average of the conditional likelihoods given the RJMCMC simulations;

- deriving the parameter estimates γ^{k+1} such that γ^{k+1} maximises the conditional expectation $Q(\gamma, \gamma^k)$.

4. Illustrative case study - a bivariate motor insurance dataset

In this section, we illustrate our model and estimation procedure with a real insurance data set. The data set, which is part of the AUSI data set (see Avanzi, Taylor and Wong, 2016b, for the details on the data set), consists of observations from a Motor insurance portfolio of a major Australian general insurer. We segment the Motor insurance portfolio by states and illustrate how our methodology can be used to understand the dependency of the claim counts between the states of New South Wales (“NSW”) and Victoria (“VIC”). We choose NSW and VIC because they are two adjacent large states in Australia (in terms of exposure), and contribute to more than half of the national population. Here we use observations of insurance claims and policy information from 01/January/2006 to 31/December/2010. We performed the following manual adjustments to the data set:

- We excluded claims resulting from catastrophe events. Although catastrophe events explicitly contribute to the dependency of insurance claims in the neighbourhood states, this source of dependency can be observed and hence explicitly modelled. Indeed, the modelling of catastrophe events benefits from knowledge that is beyond the actuarial field, and is typically conducted with separate catastrophe models in practice (see e.g. Grossi and Kunreuther, 2005). In the AUSI data set, catastrophe claims are identified with non-empty catastrophe flags;
- We excluded policy records of non-positive gross premium, total sum insured or total excess;
- We excluded invalid policy records (for example, where inception dates are after expiry dates and/or inception dates are no earlier than year 9999). This only corresponds to a negligible portion of the data set.

In Section 4.1, we explain how we allow for covariates in modelling claims counts. Sections 4.2 and 4.3 illustrate the procedures and provide the results of the univariate and bivariate fitting.

4.1. Exposure and covariates adjustment

Figures 2 and 3 present the time series of daily claim counts, numbers of policy holders, autocorrelation of daily claim counts (standardised by numbers of daily policyholders) as well as autocorrelation of weekly claim counts (standardised by numbers of weekly policyholders) for both states. This is to account for the non-constant risk exposure of total claims counts and to hence reveal a clearer picture of potential trends, weekly and seasonal patterns of claim frequencies.

There are a few key observations. Firstly, Figures 2a and 3a present the reported claim count per accident days along with the numbers of policies. The daily autocorrelation plots (see Figures 2b and 3b) reveal strong weekly cycles for both states. Furthermore, the weekly autocorrelation plots (see Figures 2c and 3c) suggest significant annual cycles for both states.

The existence of strong empirical autocorrelations (see, e.g., Figures 2b and 3b) means one should investigate the seasonality patterns from the data. Furthermore, we have also noticed the presence of potential trends in the data. We have carried an empirical investigation where we compare the average claim count per policy across (1) different days of a week, (2) different months and (3) during major Australian holidays. Please see the bar charts in Figures 2f and 2g the empirical monthly and weekly patterns of (standardised) claim count per policy for NSW, and the bar charts in Figures 3f and 3g for VIC.

Let us further specify the risk exposure, $W_g(t)$ (where $g = \text{NSW, VIC}$) as

$$\log W_g(t) = \log(\text{number of policies in force at time } t) + f_g(t), \quad (4.1)$$

where a deterministic function of time, $f_g(t)$, is adopted to capture time covariates (e.g. seasonality, trends, etc.). The particular form of $f_g(t)$ depends on the features of data.

In this project, we adopt an additive structure of $f_g(t)$ such that

$$f_g(t) = \mathbf{x}_g(t)\mathbf{a}_g \quad (4.2)$$

where $\mathbf{a}_g(t)$ is a vector of covariate coefficients and $\mathbf{x}_g(t)$ is a matrix of the covariates (with each row representing each accident day and each column representing each covariate). Note we assume that both

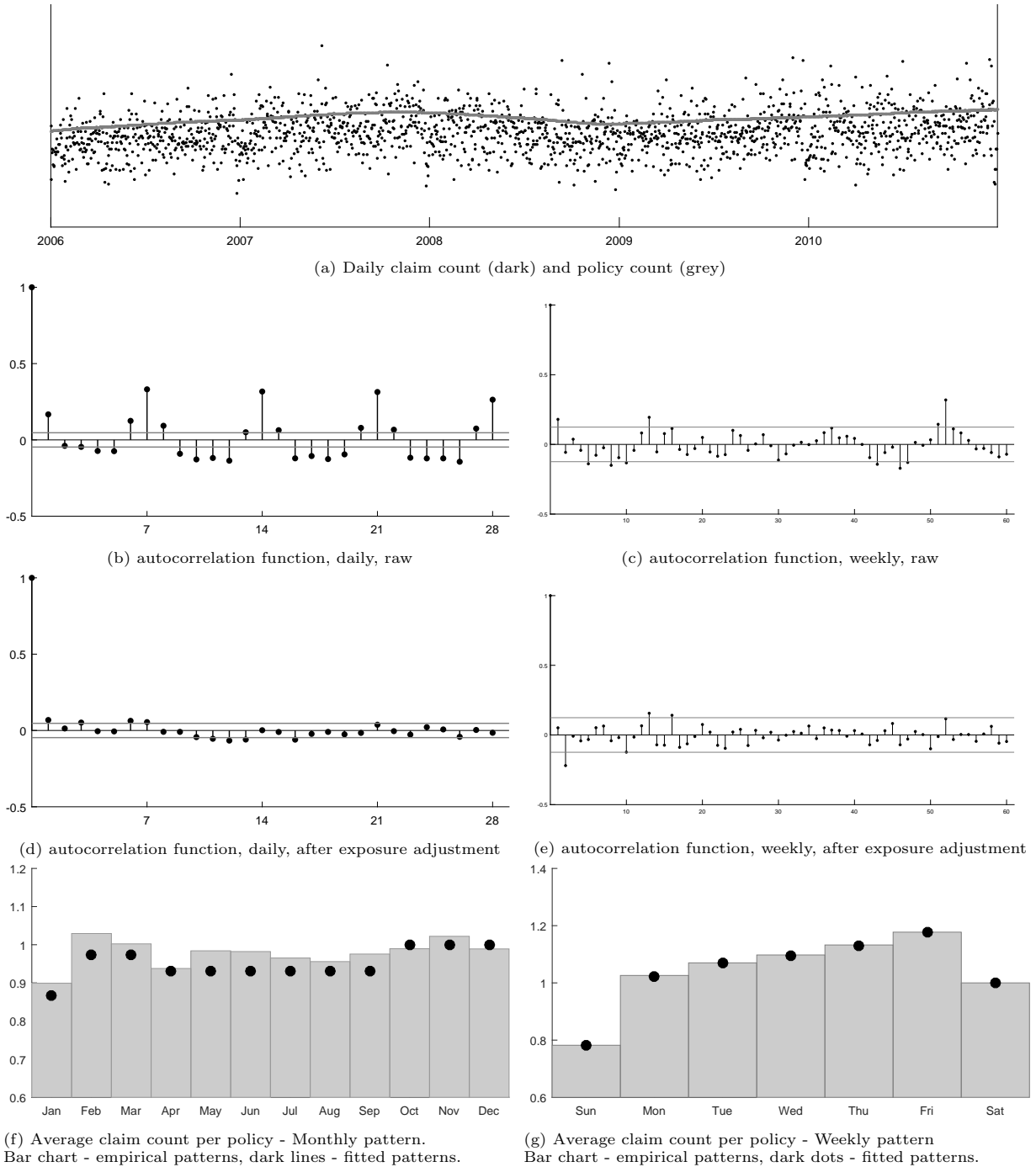


Figure 2: Summary of the claim arrival process of the Motor LoB in NSW

risk exposure $W_g(t)$ and covariates $\mathbf{x}_g(t)$ are piece-wise constant over daily intervals. This is a natural assumption given that daily observations are the most granular level of information one can possibly obtain in practice.

Joint estimation of the seasonality component of function f_g and the unobservable shot noise component

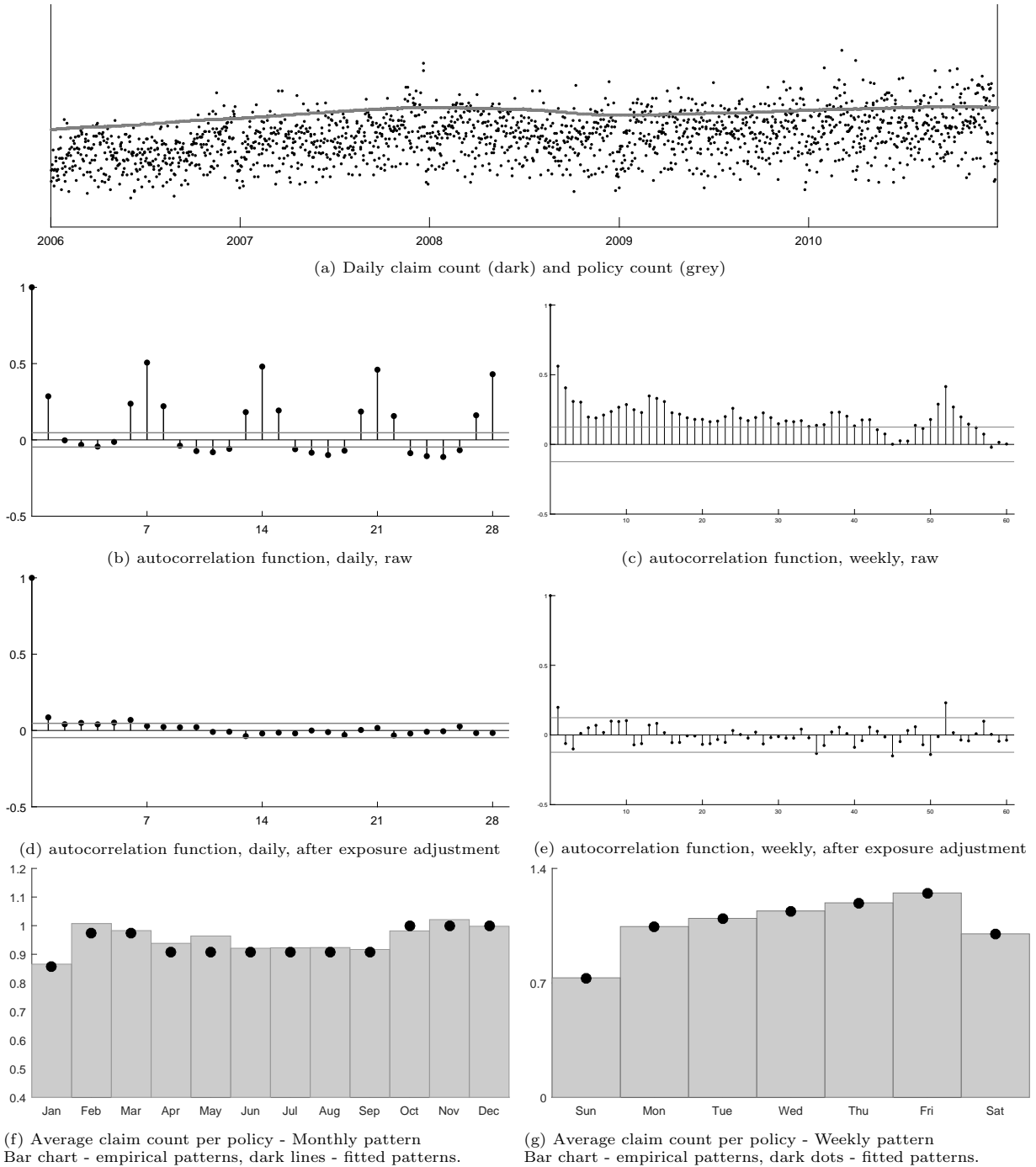


Figure 3: Summary of the claim arrival process of the Motor LoB in VIC

of $\tilde{\lambda}_g(t)$ ($g = \text{NSW, VIC}$) involves a large number of parameters. We propose to estimate the seasonality parameters prior to that of the shot noise parameters. This is achieved by fitting a Poisson GLM model to the daily claim count. This effectively approximates the Cox model by replacing the shot noise component

$\tilde{\lambda}_g(t)$ with a constant c_g along with the following discretisation:

$$\log(\tilde{\lambda}_g(\lfloor t \rfloor + 1)) = \log(W_g(\lfloor t \rfloor + 1)) + \mathbf{x}_g(\lfloor t \rfloor + \mathbf{1})\mathbf{a}_g + c_g. \quad (4.3)$$

where $\lfloor t \rfloor$ is the integer part of t (with unit of day). Such a method effectively fits a Poisson GLM model with a log-link function to the data of daily claim counts.

There can be potentially a large number of parameters in the specification of the seasonality components — in particular, for the monthly patterns. We have attempted to reduce the number of monthly covariates in two different ways. The first method applies a linear spline model which is aimed at capturing the change of monthly behaviour of claim frequencies with a simple and continuous parametric form. The second method involves grouping months (hence requires a benchmark month). Both methods were used and we selected the one that performed better based on the AIC criterion. It turns out that the method of grouping works better for both the claim processes of NSW and VIC. We have also attempted to reduce the number of weekly covariates by grouping days in a week, however, it turns out that having 6 parameters for the weekly pattern outperforms grouping (in terms of the AIC criteria) for both states. The results of fitting the monthly and weekly patterns are presented in the dark dots in Figures 2f and 3f, as well as in dark dots in Figures 2g and 3g.

We adjust the empirical autocorrelation plots by examining daily claim count per exposure. The empirical autocorrelations after adjustment are plotted in Figures 2d, 2e, 3d and 3e. Compared to those before adjustments, both states display significantly reduced levels of autocorrelation. This suggests that we have modelled away most of what could be explained with our covariates, as would be expected for any appropriate model, and we are now ready to move to the next stage and fit a dependence structure (see Avanzi et al., 2016b, for a detailed discussions of the benefits and requirements of modelling trends before applying a dependence structure).

4.2. Univariate claim arrival analysis

The fitting of Lévy copulas to ‘assemble’ processes separate from the marginal processes. Hence, we start by modelling our margins. The calibration of each univariate Cox process follows a similar procedure as Avanzi, Wong and Yang (2016c), which includes the following steps.

The initial estimates for the shot noise parameters are obtained by matching of moments, followed by jointly updating both the reporting delay and shot noise parameters through MCEM algorithms. This involves 150 MCEM iterations with 20,000 RJMCMC simulations; and we select 100 simulations from the second half of each iteration in evaluating the M-step of the MCEM algorithm.

The final parameter estimates are summarised in Table 3. Here parameters ρ , η and k fully specify the marginal shot noise processes (see Theorem 2.1.2), and the implied moments are presented in Table 4. Figure 4 presents the relative change of parameter estimates (via EM iterations), which shows that the EM estimates are quite stable for all the parameters.

States	ρ	η	κ
NSW	33.77	0.17	2.37
VIC	18.74	0.18	1.28

Table 3: Parameter estimates of the univariate shot noise Cox processes

States	mean	variance	autocorrelation at 1 day lag
NSW	85.22	350.78	0.21
VIC	80.74	384.33	0.37

Table 4: Implied moments of claim count per accident day per person based on estimated parameters

Figures 5a and 5b present the assessment of goodness-of-fit via analysing the standardised residuals of estimating daily claim counts (assuming that a claim count is Poisson distributed given the filtered

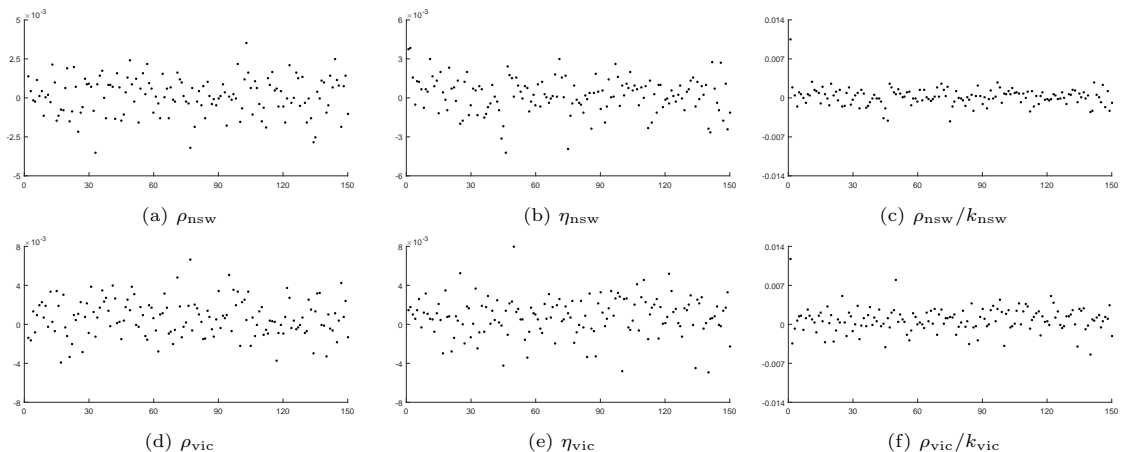


Figure 4: Relative changes of parameter estimates through EM iterations. Top - NSW, Bottom - VIC

intensity). For both states, the standard deviations of residuals are close to 1 and the means are close to 0, which indicates a satisfactory level of goodness-of-fit. Furthermore, the autocorrelations of the residuals are also close to 0 for both states, which shows that the shot noise Cox model is able to capture the serial dependency of claims counts.

4.3. Multivariate claim arrival analysis

The final step is to fit a dependence structure between the marginal processes using a Lévy copula. In this illustration, we use a Clayton Lévy copula. The Lévy copula parameter is updated via 150 MCEM iterations while the marginal parameters of both NSW and VIC are fixed. In each iteration, there are 20,000 RJMCMC simulations and we select 100 simulations from the second half of each iteration in evaluating the M-step of the MCEM algorithm. The final estimate is 0.4214 and the relative change of estimate in each EM iteration is shown in Figure 6.

Similar to the case of univariate fitting, the goodness-of-fit of the bivariate Cox model is examined by studying the residuals. Here the residuals are defined as the difference between the observed claim counts and the expected claim counts standardised by the standard deviations for all accident days. The empirical distributions and autocorrelations of residuals are presented in Figure 7.

Figure 8 displays the empirical residual plots of claims of the states, NSW and VIC, from the bivariate fitting. There is no visually significant dependency structure, which suggests good estimation of the underlying intensities and that the shot noise Cox assumptions are appropriate. In particular, Figure 8a shows that the residuals of univariate fitting are rather independent. This suggests that assuming no dependency across the Poisson processes, *given* the intensities, is reasonable. Therefore, despite the fact that the univariate fitting ignores the dependency structure, it filters out the marginal intensities and produces i.i.d. errors. Furthermore, the bivariate filtering decomposes each marginal intensity into a sum of unique and common shot noise.

Figure 9 displays the empirical copula of the integrated stochastic intensities (over accident days, free of risk exposure) of both states. It is aimed at presenting the dependency between the integrated intensities (that is, the frequencies of daily claims), and therefore the results are filtered intensity processes (which are not observable from data). This is because realised intensities are not directly observable (only claim counts are), so they must be ‘inferred’ from the claim counts through the lens of a specific model through filtering. Furthermore, a data point on Figure 9 refers to the number of daily integrated intensities that fall into each cell, while the integrated intensities are standardised to $[0,1]$. In particular, Figures 9a and 9b display the copulas of integrated intensities from the unique jumps and common jumps respectively. This illustrates that the Lévy copula structure further decomposes the marginal intensities into two components, where the

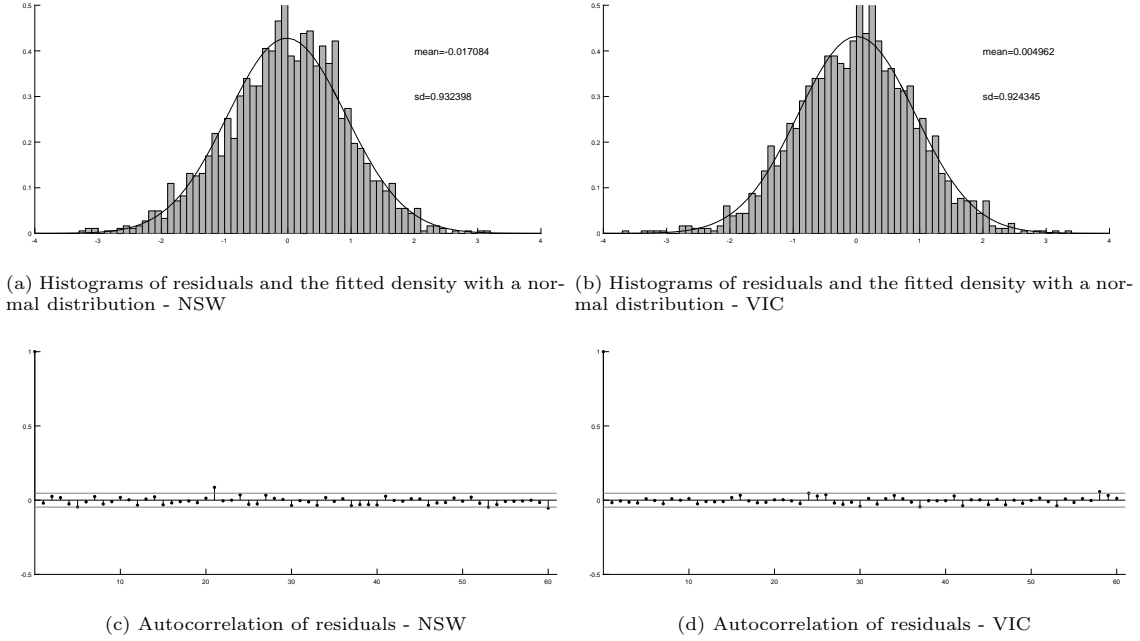


Figure 5: Residual analysis of the independent shot noise Cox process fitting

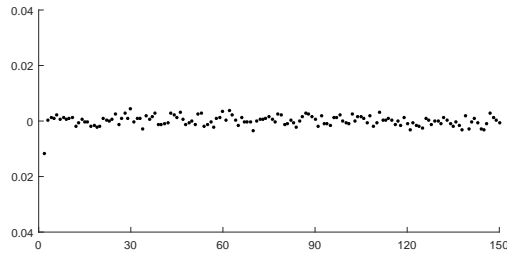


Figure 6: Relative changes of the estimate of the Lévy copula parameter

unique components are independent and there is a significant positive dependency structure between the common components.

Remark 4.1. *As discussed earlier in the paper, it is empirically difficult to judge whether a Cox model is a good candidate directly from the data. However, the model developed in this paper has properties displayed by the data (non-stationary exposure, autocorrelation, ...), and it seems that, once fitted, it can explain most of those convincingly. Specifically, while the exposure adjustment removed much of the (very significant) autocorrelations within both processes (compare the pairs b–d and c–e in both Figures 2 and 3), the residual autocorrelations after de-trending are still significantly different from 0 at reasonable confidence level. Furthermore, when comparing Figures 2e and 3e with their counterparts after application of the model in Figures 7c and 7d, it appears that those were almost entirely captured by the model. While we cannot guarantee that our model is the best, we are convinced it is of reasonable quality for this data set.*

4.4. Prediction results and discussion

Given the filtered multivariate intensity, one can investigate further how the unique and common shots contribute to the integrated stochastic intensities. The results in Table 5 indicate substantial weights from common shocks in both marginal intensities. For VIC, the common events contribution to 49.30% of the stochastic intensities of the claim arrival process. Here 49.30% does not refer to the number of events;

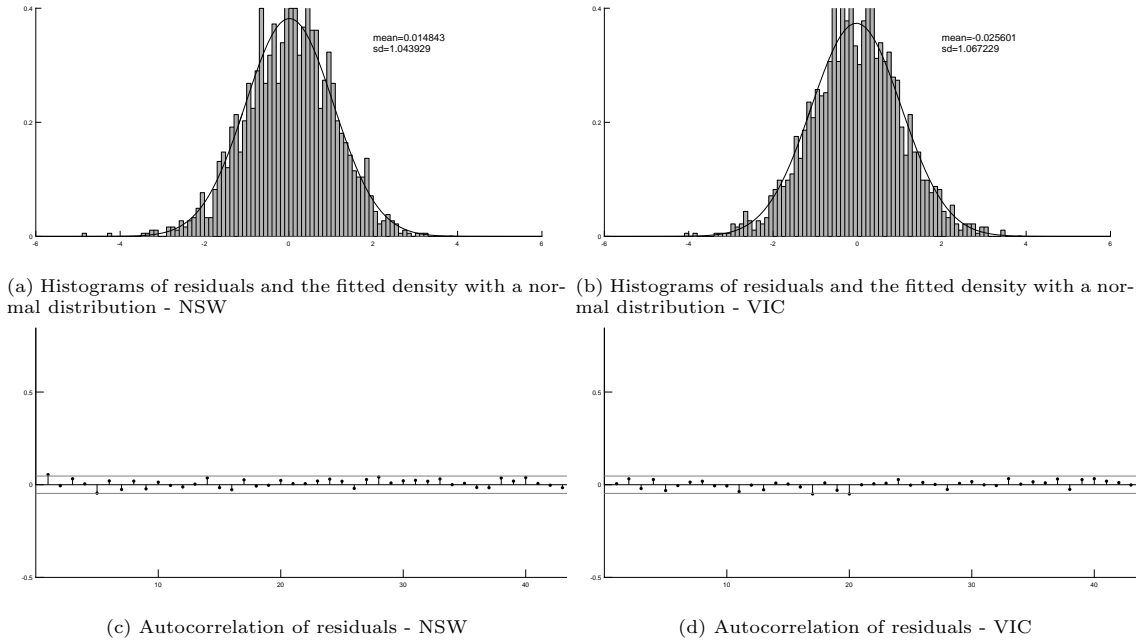


Figure 7: Residual analysis of the bivariate shot noise Cox process fitting

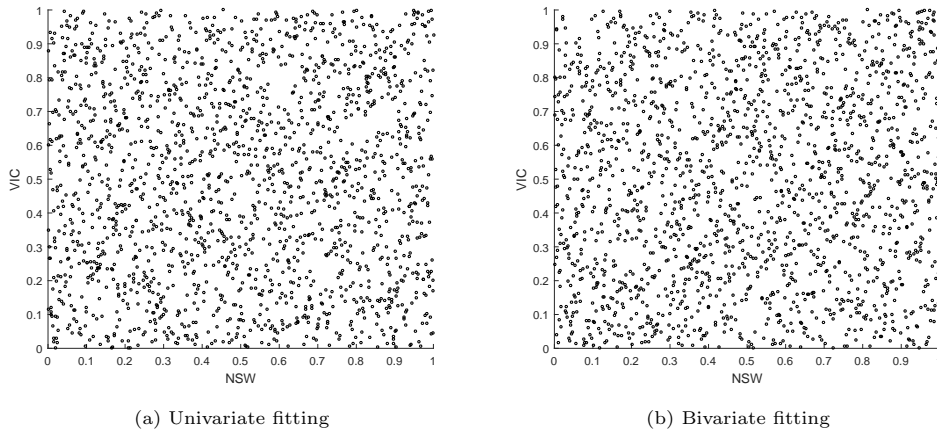


Figure 8: Empirical copulas of *residuals* from fitting the claim arrival components to the NSW and VIC. *Absence of any noticeable dependence structure suggests a good fit.*

instead, it refers to the combination effects of frequency, severity and decay of the shot events on the stochastic intensity of claim arrivals in the state of VIC.

We predicted the distributions of total future claim counts over the next one year (assuming constant risk exposures) through 100,000 simulations. Note that the initial value of the shot noise trajectory is chosen as the filtered shot noise intensity at the end of the observation period, hence the simulations of future scenarios are implicitly based on the latest status of the shot noise development. Figure 10 visualises the bivariate histogram of the copulas between claim counts in the two cases. Here the colour intensity in each cell refers to the number of simulations where the bivariate claim count fall into a specific joint quantile range. It shows that there is significant evidence of dependency, which is consistent with the results in Table 5. This is further confirmed with the numerical dependency measures in Table 6, where all the measures

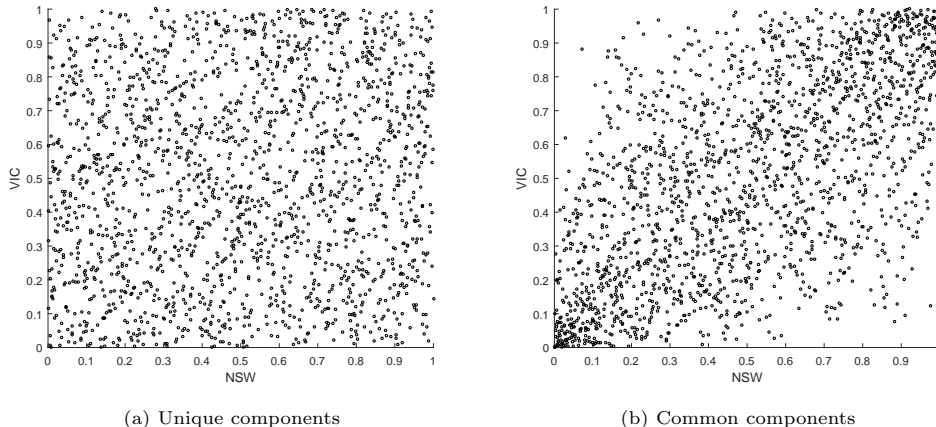


Figure 9: Empirical copula of claim counts (integrated shot noise intensities over each accident day) of the NSW and VIC states - bivariate fitting. This illustrates the dependence that is present in the data, as filtered and interpreted by our model.

	NSW	VIC
Contribution from unique shot noise	67.54%	50.70%
Contribution from common shot noise	32.46%	49.30%

Table 5: Decomposition of the integrated intensities

are material and statistically significant (with all p -values being almost 0). Note that the values in Table 6 should not be compared directly to dependency measures of real data. This is due to the non-constant risk exposures and also serial dependency of claim counts over time, which means the i.i.d. assumptions of claim counts in constructing the dependency measures is not valid.

The dependency measures in Table 6 refer to the dependency structure of the next year, based on a given set of controlled factors. Our methodology provides a statistical sound way of making assumptions about the dependency structures. The implied future dependency measures can then be used in more traditional (and straightforward) methodologies, for example, when a correlation matrix is required to aggregate individual portfolios to a company level. This improves the existing practice of multivariate reserving where dependency structures are usually based on expert knowledge and industry statistics. In particular, one common approach to estimate risk margin at a company level is by aggregating individual risk margins of various portfolios with the help of correlations. But correlations typically cannot be inferred from data (as 10 years of data would yield only 10 observations, which is insufficient to estimate yearly correlations), and are hence generally chosen judgmentally (see also Avanzi, Taylor and Wong, 2016b, for further discussion of this). Our approach provides a promising first step towards estimating dependency measures through a more rigorous and objective approach. Indeed, only a few years of data are required to fit a model that can subsequently yield implied dependence measures for any time horizon.

Pearson's correlation	Kendall's tau	Spearman's rho
0.3633	0.2355	0.3469

Table 6: Dependency measurement across the bivariate prediction of claims counts

Remark 4.2. *We have also attempted to provide comparison between our model and the Chain Ladder model, however we realise that such a comparison can be challenging and may provide misleading results. There are fundamental differences in the theoretical properties and how the models are implemented between the Chain Ladder model (and most of the existing reserving models) and our approach. These facts make it very difficult to conduct a meaningful comparison. That being said, the multivariate Cox model can be*

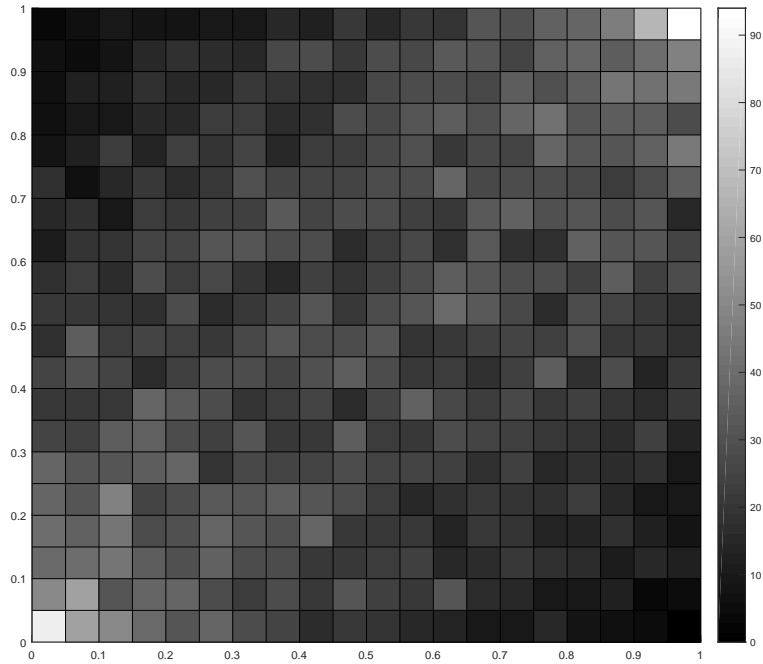


Figure 10: Heat maps of the histograms of the empirical copulas between the simulated claims counts

complementary to the Chain Ladder model. While the Chain Ladder model is commonly used to estimate the risk margins of individual LoBs, our multivariate Cox model can provide insights into the dependence structures at play, and hence help modellers make informed and educated choice for the correlations that would be used in traditional modelling approaches to aggregate results.

5. Conclusion

In this paper, we have developed a *multivariate* Cox shot noise process approach to model (claim) counts. Common shocks were introduced on the intensities of claim arrivals across multiple counts processes with the help of Lévy copulas. We also developed a filtering algorithm to estimate the underlying joint intensity.

The independent increment property of a Lévy process means the dependency structure of a multivariate compound Poisson process can only be introduced via a common shock model. However, there is a bijective relationship between a Lévy copula model and a common shock model (see Sklar’s theorem for Lévy copulas in Tankov, 2003, for example). Hence, our proposed approach is actually not different from a common shock approach, but it is significantly more tractable. It also has the advantage of introducing dependence *within* each process, a common feature observed in actuarial data sets.

The model construction and calibration procedures are illustrated with a real bivariate data set. In particular, we account for time covariates (including weekly patterns, annual patterns, and trends). After allowing for these covariates, the results show that the dependency between the claim arrival processes of the Motor LoB in both NSW and VIC states is still significant. This is expected, due to random phenomena that may affect both states at the same time (e.g., weather events). The Cox model presented in this paper proved successful in capturing the residual correlation structures that were left in the data after detrending.

Our work focused on the particular issue of dependency modelling in claim frequency. Such a framework may help in understanding the dependency structure between the claim processes of multiple LoBs, and

in particular, its impact on the quantiles of the aggregated loss. For example, one implementation of the multivariate Cox process is to derive an implied dependence structure (see e.g. Figure 10), which can be utilised in aggregating individual reserve estimates to a company level.

The model implementation procedures developed in this paper can be easily extended from our bivariate illustration. This will require an appropriate choice of a higher dimension Lévy copula. Should the dependency structure across multiple counts processes be non-exchangeable, appropriate Lévy copulas can be constructed (see, for example Avanzi, Tao, Wong and Yang, 2016a).

Acknowledgements

Authors are very grateful for comments from two referees, which led to significant improvements of the paper. Earlier versions of this paper were presented at the 20th International Congress on Insurance: Mathematics and Economics (Atlanta, U.S.A.), at the 3rd European Actuarial Journal Conference (Lyon, France), at the 22nd International Congress on Insurance: Mathematics and Economics (Sydney, Australia), and at the Australian National University Research School of Finance, Actuarial Studies and Statistics 2018 Summer Camp. The authors are grateful for constructive comments received from colleagues who attended those events.

This research was supported under Australian Research Council's Linkage (LP130100723, with funding partners Allianz Australia Insurance Ltd, Insurance Australia Group Ltd, and Suncorp Metway Ltd) and Discovery (DP200101859) Projects funding schemes. Furthermore, Avanzi acknowledges support from a grant of the Natural Science and Engineering Research Council of Canada (project number RGPIN-2015-04975). Finally, Yang acknowledges financial support from an Australian Postgraduate Award and supplementary scholarships provided by the UNSW Business School. The views expressed herein are those of the authors and are not necessarily those of the supporting organisations.

References

- Antonio, K., Plat, R., 2014. Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal* 2014.
- Arjas, E., 1989. The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bulletin* 19, 139–152.
- Avanzi, B., Cassar, L.C., Wong, B., 2011. Modelling dependence in insurance claims processes with Lévy copulas. *ASTIN Bulletin* 41, 575–609.
- Avanzi, B., Tao, J., Wong, B., Yang, X., 2016a. Capturing non-exchangeable dependence in multivariate loss processes with nested lévy copulas. *Annals of Actuarial Science* 10, 87–117.
- Avanzi, B., Taylor, G.C., Wong, B., 2016b. Correlations between insurance lines of business: An illusion or a real phenomenon? Some methodological considerations. *ASTIN Bulletin* 46, 225–263.
- Avanzi, B., Taylor, G.C., Wong, B., Xian, A., 2020. Modelling and understanding count processes through a markov-modulated non-homogeneous poisson process framework. *European Journal of Operational Research* in press.
- Avanzi, B., Wong, B., Yang, X., 2016c. A micro-level claim count model with overdispersion and reporting delays. *Insurance: Mathematics and Economics* 71, 1–14.
- Badescu, A., Lin, X., Tang, D., 2016. A marked cox model for the number of ibnr claims: Theory. *Insurance Mathematics Economics* 69, 29–37.
- Badescu, A.L., Lin, X.S., Tang, D., 2019. A marked cox model for the number of ibnr claims: Estimation and application. *ASTIN Bulletin* 49, 709–739.
- Bühlmann, H., 1970. *Mathematical Methods in Risk Theory*. Grundlehren der mathematischen Wissenschaften, Springer-Verlag, Berlin, Heidelberg, New York.
- Centanni, S., Minozzo, M., 2006a. Estimation and filtering by reversible jump MCMC for a doubly stochastic poisson model for ultra-high-frequency financial data. *Statistical Modelling* 6, 97–118.
- Centanni, S., Minozzo, M., 2006b. A monte carlo approach to filtering for a class of marked doubly stochastic poisson processes. *Journal of the American Statistical Association* 101, 1582–1597.
- Cont, R., Tankov, P., 2004. *Financial Modelling With Jump Processes*. Chapman & Hall/CRC, London.
- Cramér, H., 1930. On the mathematical theory of risk. *Skand. Jubilee Volume*, Stockholm.
- Cramér, H., 1955. *Collective Risk Theory: A Survey of the Theory from the Point of View of the Theory of Stochastic Processes*. Ab Nordiska Bokhandeln, Stockholm.
- Denuit, M., Maréchal, X., Pitrebois, S., Walhin, J., 2007. *Actuarial modelling of claim counts*. Wiley Online Library.
- Gelman, A., Jones, G., Brooks, S., Meng, X.L. (Eds.), 2011. *Handbook of Markov Chain Monte Carlo*. Boca Raton : CRC Press.
- Green, P., 1995. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* 82, 711–732.

- Grossi, P., Kunreuther, H., 2005. Catastrophe modeling: a new approach to managing risk. huebner international series on risk, insurance and economic security.
- Jessen, A.H., Mikosch, T., Samorodnitsky, G., 2011. Prediction of outstanding payments in a poisson cluster model. Scandinavian Actuarial Journal 2011.
- Jewell, W.S., 1989. Predicting ibnr events and delays. ASTIN Bulletin 19.
- Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.
- de Jong, P., Heller, G.Z., 2008. Generalized linear models for insurance data. ambridge, UK ; New York : Cambridge University Press.
- Larsen, C.R., 2007. An individual claims reserving model. ASTIN Bulletin 37.
- Li, H., 2002. On the Dependence Structure of a System of Components with a Multivariate Shot-Noise Hazard Rate Process. Nova Publishers. Advances in Mathematics Research.
- Lindskog, F., McNeil, A.J., 2003. Common poisson shock models: Applications to insurance and credit risk modelling. ASTIN Bulletin 33, 209–238.
- Lundberg, F., 1909. Über die Theorie der Rückversicherung. Transactions of the VIth International Congress of Actuaries 1, 877–948.
- Matsui, M., 2014. Prediction in a non-homogeneous poisson cluster model. Insurance: Mathematics and Economics 55.
- Matsui, M., 2015. Prediction in a poisson cluster model with multiple cluster processes. Scandinavian Actuarial Journal 2015.
- Matsui, M., Mikosch, T., 2010. Prediction in a poisson cluster model. Journal of Applied Probability 47.
- McNeil, A.J., Frey, R., Embrechts, P., 2015. Quantitative risk management: Concepts, techniques and tools. Princeton, USA: Princeton university press.
- Møller, J., Waagepetersen, R.P., 2004. Statistical inference and simulation for spatial point processes. Chapman & Hall/CRC.
- Norberg, R., 1993. Prediction of Outstanding Liabilities in Non-Life Insurance. ASTIN Bulletin 23, 95–115.
- Norberg, R., 1999. Prediction of Outstanding Liabilities - II Model Variations and Extensions. ASTIN Bulletin 29, 5–27.
- Pigeon, M., Antonio, K., Denuit, M., 2013. Individual loss reserving with the multivariate skew normal framework. ASTIN Bulletin 43.
- Pigeon, M., Antonio, K., Denuit, M., 2014. Individual loss reserving using paid-incurred data. Insurance: Mathematics and Economics 58.
- Ross, S.M., 2014. Introduction to probability models. Academic press.
- Rydén, T., 1996. An EM algorithm for estimation in markov-modulated poisson processes. Computational Statistics & Data Analysis 21.
- Selch, D.A., Scherer, M., 2018. A Multivariate Claim Count Model for Applications in Insurance. Springer.
- Tankov, P., 2003. Dependence structure of spectrally positive multidimensional Lévy processes.
- Zhao, X., Zhou, X., 2010. Applying copula models to individual claim loss reserving methods. Insurance: Mathematics and Economics 46, 290–299.
- Zhao, X.B., Zhou, X., Wong, J.L., 2009. Semiparametric model for prediction of individual claim loss reserving. Insurance: Mathematics and Economics 45, 1–8.

A. Proof for Example 1

Proof. Consider the process of the common shots, we have

$$\begin{aligned}\tilde{\lambda}_1^\parallel(t) &= \tilde{\lambda}_1^\parallel(0)e^{-t\kappa_1} + \sum_{i=1}^{J_{12}^\parallel(t)} X_{i,1:12}^\parallel e^{-(t-\tau_{i,12}^\parallel)\kappa_1}, \\ \tilde{\lambda}_2^\parallel(t) &= \tilde{\lambda}_2^\parallel(0)e^{-t\kappa_2} + \sum_{i=1}^{J_{12}^\parallel(t)} X_{i,2:12}^\parallel e^{-(t-\tau_{i,12}^\parallel)\kappa_2}.\end{aligned}\tag{A.1}$$

We characterise the trajectory of the bivariate shot noise process by using a random vector, $\boldsymbol{\theta}_{12}^\parallel(\mathbf{t})$, which is the collection of the number of jumps, the location and joint severities of each shot up to time t , that is

$$\boldsymbol{\theta}_{12}^\parallel(\mathbf{t}) = \{J_{12}^\parallel(t), \tau_{1,12}^\parallel, \dots, \tau_{J_{12}^\parallel(t),12}^\parallel, X_{1,12}^\parallel, \dots, X_{J_{12}^\parallel(t),12}^\parallel\}.\tag{A.2}$$

Based on the Conditional Covariance Formula (see, for example Ross, 2014, Chapter 7), we have

$$\text{Cov}(\tilde{\lambda}_1^\parallel(t), \tilde{\lambda}_2^\parallel(t)) = \mathbb{E} \left[\text{Cov}(\tilde{\lambda}_1^\parallel(t), \tilde{\lambda}_2^\parallel(t) \mid J_{12}^\parallel(t)) \right] + \text{Cov} \left(\mathbb{E}[\tilde{\lambda}_1^\parallel(t) \mid J_{12}^\parallel(t)], \mathbb{E}[\tilde{\lambda}_2^\parallel(t) \mid J_{12}^\parallel(t)] \right).\tag{A.3}$$

Furthermore, conditional on $J_{12}^{\parallel}(t)$ (which is a Poisson random variable itself with intensity ρ_{12}^{\parallel}), the locations of shots (that is, $\{\tau_{i,12}^{\parallel} | J_{12}^{\parallel}(t)\}_{i=1,\dots,J_{12}^{\parallel}(t)}$) are independent uniform random variables.

Firstly, we start from deriving $\mathbb{E} \left[\text{Cov} \left(\tilde{\lambda}_1^{\parallel}(t), \tilde{\lambda}_2^{\parallel}(t) | J_{12}^{\parallel}(t) \right) \right]$. We have:

$$\begin{aligned}
& \mathbb{E} \left[\text{Cov} \left(\tilde{\lambda}_1^{\parallel}(t), \tilde{\lambda}_2^{\parallel}(t) | J_{12}^{\parallel}(t) \right) \right] \\
&= \mathbb{E} \left[\text{Cov} \left(\tilde{\lambda}_1^{\parallel}(0)e^{-t\kappa_1} + \sum_{i=1}^{J_{12}^{\parallel}(t)} X_{i,1:12}^{\parallel} e^{-(t-\tau_{i,12}^{\parallel})\kappa_1}, \tilde{\lambda}_2^{\parallel}(0)e^{-t\kappa_2} + \sum_{i=1}^{J_{12}^{\parallel}(t)} X_{i,2:12}^{\parallel} e^{-(t-\tau_{i,12}^{\parallel})\kappa_2} | J_{12}^{\parallel}(t) \right) \right] \quad (\text{A.4}) \\
&= \mathbb{E} \left[e^{-t(\kappa_1+\kappa_2)} \text{Cov} \left(\tilde{\lambda}_1^{\parallel}(0), \tilde{\lambda}_2^{\parallel}(0) \right) + J_{12}^{\parallel}(t) \text{Cov} \left(X_{1:12}^{\parallel} e^{-(t-\tau_{12}^{\parallel})\kappa_1}, X_{2:12}^{\parallel} e^{-(t-\tau_{12}^{\parallel})\kappa_1} | J_{12}^{\parallel}(t) \right) \right] \\
&= \mathbb{E} \left[e^{-t(\kappa_1+\kappa_2)} \text{Cov} \left(\tilde{\lambda}_1^{\parallel}(0), \tilde{\lambda}_2^{\parallel}(0) \right) + J_{12}^{\parallel}(t) e^{-t(\kappa_1+\kappa_2)} \text{Cov} \left(X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_1}, X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_2} | J_{12}^{\parallel}(t) \right) \right]
\end{aligned}$$

and

$$\begin{aligned}
& \text{Cov} \left(X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_1}, X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_2} | J_{12}^{\parallel}(t) \right) \\
&= \mathbb{E} \left[X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_1} X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_2} | J_{12}^{\parallel}(t) \right] - \mathbb{E} \left[X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_1} | J_{12}^{\parallel}(t) \right] \mathbb{E} \left[X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_2} | J_{12}^{\parallel}(t) \right] \quad (\text{A.5}) \\
&= \mathbb{E} \left[X_{1:12}^{\parallel} X_{2:12}^{\parallel} \right] \mathbb{E} \left[e^{\tau_{12}^{\parallel}(\kappa_1+\kappa_2)} | J_{12}^{\parallel}(12) \right] - \mathbb{E} \left[X_{1:12}^{\parallel} \right] \mathbb{E} \left[X_{2:12}^{\parallel} \right] \mathbb{E} \left[e^{\tau_{12}^{\parallel}\kappa_1} | J_{12}^{\parallel}(t) \right] \mathbb{E} \left[e^{\tau_{12}^{\parallel}\kappa_2} | J_{12}^{\parallel}(t) \right].
\end{aligned}$$

The location of the shot given the total number of shot is uniformly distributed over $[0, t]$, therefore one can obtain the first two moments by:

$$\mathbb{E} \left[e^{\tau_{12}^{\parallel}a} | J_{12}^{\parallel}(t) \right] = \int_0^t e^{ax} \frac{1}{t} dx = \frac{e^{at} - 1}{at}, \quad (\text{A.6})$$

hence

$$\begin{aligned}
& \text{Cov} \left(X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_1}, X_{1:12}^{\parallel} e^{\tau_{12}^{\parallel}\kappa_2} | J_{12}^{\parallel}(t) \right) \\
&= \mathbb{E} \left[X_{1:12}^{\parallel} X_{2:12}^{\parallel} \right] \frac{e^{(\kappa_1+\kappa_2)t} - 1}{(\kappa_1 + \kappa_2)t} - \mathbb{E} \left[X_{1:12}^{\parallel} \right] \mathbb{E} \left[X_{2:12}^{\parallel} \right] \frac{e^{\kappa_1 t} - 1}{\kappa_1 t} \frac{e^{\kappa_2 t} - 1}{\kappa_2 t}. \quad (\text{A.7})
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathbb{E} \left[\text{Cov} \left(\tilde{\lambda}_1^{\parallel}(t), \tilde{\lambda}_2^{\parallel}(t) | J_{12}^{\parallel}(t) \right) \right] \\
&= e^{-t(\kappa_1+\kappa_2)} \text{Cov} \left(\tilde{\lambda}_1^{\parallel}(0), \tilde{\lambda}_2^{\parallel}(0) \right) \quad (\text{A.8}) \\
&+ \rho_{12}^{\parallel} t e^{-t(\kappa_1+\kappa_2)} \left(\mathbb{E} \left[X_{1:12}^{\parallel} X_{2:12}^{\parallel} \right] \frac{e^{(\kappa_1+\kappa_2)t} - 1}{(\kappa_1 + \kappa_2)t} - \mathbb{E} \left[X_{1:12}^{\parallel} \right] \mathbb{E} \left[X_{2:12}^{\parallel} \right] \frac{e^{\kappa_1 t} - 1}{\kappa_1 t} \frac{e^{\kappa_2 t} - 1}{\kappa_2 t} \right).
\end{aligned}$$

Secondly, we derive $\text{Cov} \left(\mathbb{E} \left[\tilde{\lambda}_1^{\parallel}(t) | J_{12}^{\parallel}(t) \right], \mathbb{E} \left[\tilde{\lambda}_2^{\parallel}(t) | J_{12}^{\parallel}(t) \right] \right)$. We start from deriving the conditional expectation terms:

$$\begin{aligned}
\mathbb{E} \left[\tilde{\lambda}_1^{\parallel}(t) | J_{12}^{\parallel}(t) \right] &= \mathbb{E} \left[\tilde{\lambda}_1^{\parallel}(0)e^{-t\kappa_1} + \sum_{i=1}^{J_{12}^{\parallel}(t)} X_{i,1:12}^{\parallel} e^{-(t-\tau_{i,12}^{\parallel})\kappa_1} | J_{12}^{\parallel}(t) \right] \\
&= e^{-t\kappa_1} \left(\mathbb{E} \left[\tilde{\lambda}_1^{\parallel}(0) \right] + J_{12}^{\parallel}(t) \mathbb{E} \left[X_{1:12}^{\parallel} \right] \mathbb{E} \left[e^{\tau_{12}^{\parallel}\kappa_1} \right] \right) \\
&= e^{-t\kappa_1} \left(\mathbb{E} \left[\tilde{\lambda}_1^{\parallel}(0) \right] + J_{12}^{\parallel}(t) \mathbb{E} \left[X_{1:12}^{\parallel} \right] \frac{e^{\kappa_1 t} - 1}{\kappa_1 t} \right). \quad (\text{A.9})
\end{aligned}$$

$$\mathbb{E} \left[\tilde{\lambda}_2^\parallel(t) \mid J_{12}^\parallel t \right] = e^{-t\kappa_2} \left(\mathbb{E} \left[\tilde{\lambda}_2^\parallel(0) \right] + J_{12}^\parallel(t) \mathbb{E} \left[X_{2:12}^\parallel \right] \frac{e^{\kappa_2 t} - 1}{\kappa_2 t} \right) \quad (\text{A.10})$$

Therefore we arrive at

$$\begin{aligned} & \text{Cov} \left(\mathbb{E} \left[\tilde{\lambda}_1^\parallel(t) \mid J_{12}^\parallel t \right], \mathbb{E} \left[\tilde{\lambda}_2^\parallel(t) \mid J_{12}^\parallel t \right] \right) \\ &= e^{-t(\kappa_1 + \kappa_2)} \frac{e^{\kappa_1 t} - 1}{\kappa_1 t} \frac{e^{\kappa_2 t} - 1}{\kappa_2 t} \mathbb{E} \left[X_{1:12}^\parallel \right] \mathbb{E} \left[X_{2:12}^\parallel \right] \text{Var} \left(J_{12}^\parallel(t) \right) \\ &= e^{-t(\kappa_1 + \kappa_2)} \frac{e^{\kappa_1 t} - 1}{\kappa_1 t} \frac{e^{\kappa_2 t} - 1}{\kappa_2 t} \rho_{12}^\parallel t \mathbb{E} \left[X_{1:12}^\parallel \right] \mathbb{E} \left[X_{2:12}^\parallel \right] \end{aligned} \quad (\text{A.11})$$

and

$$\begin{aligned} & \text{Cov} \left(\tilde{\lambda}_{1:12}^\parallel(t), \tilde{\lambda}_{2:12}^\parallel(t) \right) \\ &= e^{-t(\kappa_1 + \kappa_2)} \text{Cov} \left(\tilde{\lambda}_1^\parallel(0), \tilde{\lambda}_2^\parallel(0) \right) \\ &+ \rho_{12}^\parallel t e^{-t(\kappa_1 + \kappa_2)} \left(\mathbb{E} \left[X_{1:12}^\parallel X_{2:12}^\parallel \right] \frac{e^{(\kappa_1 + \kappa_2)t} - 1}{(\kappa_1 + \kappa_2)t} - \mathbb{E} \left[X_{1:12}^\parallel \right] \mathbb{E} \left[X_{2:12}^\parallel \right] \frac{e^{\kappa_1 t} - 1}{\kappa_1 t} \frac{e^{\kappa_2 t} - 1}{\kappa_2 t} \right) \\ &+ e^{-t(\kappa_1 + \kappa_2)} \frac{e^{\kappa_1 t} - 1}{\kappa_1 t} \frac{e^{\kappa_2 t} - 1}{\kappa_2 t} \rho_{12}^\parallel t \mathbb{E} \left[X_{1:12}^\parallel \right] \mathbb{E} \left[X_{2:12}^\parallel \right] \\ &= e^{-t(\kappa_1 + \kappa_2)} \text{Cov} \left(\tilde{\lambda}_1^\parallel(0), \tilde{\lambda}_2^\parallel(0) \right) + \rho_{12}^\parallel t e^{-t(\kappa_1 + \kappa_2)} \left(\mathbb{E} \left[X_{1:12}^\parallel X_{2:12}^\parallel \right] \frac{e^{(\kappa_1 + \kappa_2)t} - 1}{(\kappa_1 + \kappa_2)t} \right). \end{aligned} \quad (\text{A.12})$$

Given that this is a stationary bivariate process, we have:

$$\text{Cov} \left(\tilde{\lambda}_{1:12}^\parallel(t), \tilde{\lambda}_{2:12}^\parallel(t) \right) = \text{Cov} \left(\tilde{\lambda}_{1:12}^\parallel(0), \tilde{\lambda}_{2:12}^\parallel(0) \right), \quad (\text{A.13})$$

hence

$$\begin{aligned} & \left(1 - e^{-t(\kappa_1 + \kappa_2)} \right) \text{Cov} \left(\tilde{\lambda}_{1:12}^\parallel(t), \tilde{\lambda}_{2:12}^\parallel(t) \right) \\ &= \rho_{12}^\parallel t e^{-t(\kappa_1 + \kappa_2)} \mathbb{E} \left[X_{1:12}^\parallel X_{2:12}^\parallel \right] \frac{e^{(\kappa_1 + \kappa_2)t} - 1}{(\kappa_1 + \kappa_2)t}. \end{aligned} \quad (\text{A.14})$$

Therefore

$$\text{Cov} \left(\tilde{\lambda}_{1:12}^\parallel(t), \tilde{\lambda}_{2:12}^\parallel(t) \right) = \frac{\rho_{12}^\parallel \mathbb{E} \left[X_{1:12}^\parallel X_{2:12}^\parallel \right]}{\kappa_1 + \kappa_2}. \quad (\text{A.15})$$

Since the two unique shot processes are independent with the bivariate common shot process, therefore

$$\text{Cov} \left(\tilde{\lambda}_1(t), \tilde{\lambda}_2(t) \right) = \text{Cov} \left(\tilde{\lambda}_{1:12}^\parallel(t) + \tilde{\lambda}_1^\perp(t), \tilde{\lambda}_{2:12}^\parallel(t) + \tilde{\lambda}_2^\perp(t) \right) = \text{Cov} \left(\tilde{\lambda}_{1:12}^\parallel(t), \tilde{\lambda}_{2:12}^\parallel(t) \right), \quad (\text{A.16})$$

which completes the proof. \square