A tractable method for modelling unobservable or complex drivers

Benjamin Avanzi, Greg Taylor, Bernard Wong and Alan Xian

School of Risk and Actuarial Studies, University of New South Wales

2019 IAA SECTION COLLOQUIUM
Non-Life (ASTIN) Session 5
4 April 2019

UNSW | AGSM
Business School

This presentation is based on the work of two papers:

1. *How to proxy the unmodellable: Analysing granular insurance claims data in the presence of unobservable or complex drivers*
   - Awarded the Taylor Fry Silver Prize at the 2018 Actuaries Institute General Insurance Seminar
2. *Inference and prediction of counts using Markov-modulated non-homogeneous Poisson processes*
   - Which is submitted for publication
   - Available on SSRN: `http://ssrn.com/abstract=3354342`

UNSW | AGSM
Business School

UNSW | AGSM
Business School

Macro claims models

Key step: Claims are aggregated and discretised into triangles (for example, by accident year i and development year j)

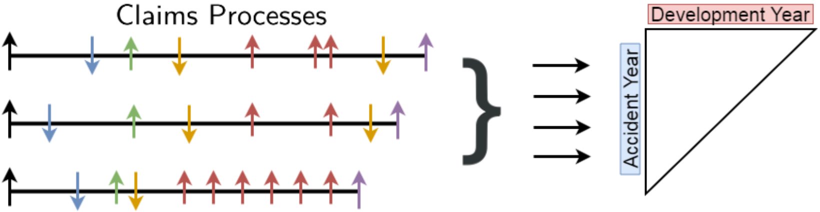


Figure: Aggregate Claims Modelling: Key Step

Why is this approach so prevalent? Flexible, interpretable, computationally cheap, historical reasons, ...

UNSW | AGSM Business School

Issues with Macro approaches

However, there are issues with such approaches documented in the literature:

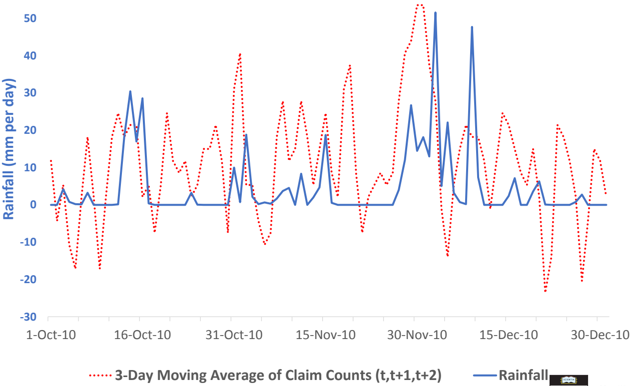
- Problems resulting from small sample sizes (Renshaw [1994], Verdonck, Wouwe, and Dhaene [2009])
- Problems with underlying processes and assumptions (Halliwell [2007], Taylor [2011], Taylor and McGuire [2004])
- Problems with practical implementation (Kunkler [2004], Liu and Verrall [2009], Parodi [2014])

An additional concern is that *potentially material information* could be unnecessarily discarded due "excessive" aggregation.

But how much more information and complexity do we want to bring? What is material?

Motor claims and rainfall in ACT

The Pearson correlation between the two time series shown below is a **statistically significant at 32%.**



UNSW | AGSM
Business School

What if element is intractable / annoying to model?

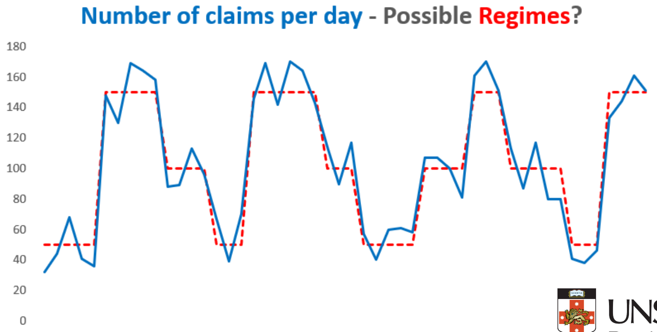
In our rainfall example, the impact on the granular daily claim frequencies was overdispersion and persistence.

Rainfall seems material, but explicit modelling of it would be tough. For instance, how to convincingly match geographical distribution of weather stations with exposure. It was OK for ACT, but not for NSW.

How about we find a way to ‘proxy’ the effect in a simple way?

overdispersion and persistence...

- Traditional claim count model is Poisson
- Only way to achieve overdispersion *and* persistence is through a stochastic intensity
- What are possible choices for a stochastic Poisson intensity?



UNSW | AGSM
Business School

Markov-modulated Poisson processes

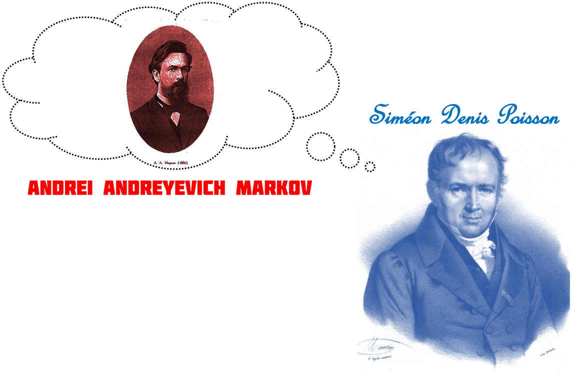


Figure: A Markov-modulated Poisson process (MMPP)

Or perhaps more formally...

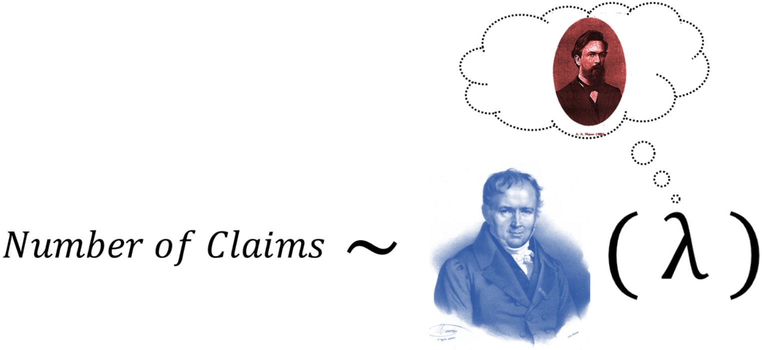


Figure: The “Formal” Definition for a MMPP

UNSW | AGSM Business School

Where are MMPP processes used?

These processes are commonly used to fit clustered or "bursty" processes where there are hidden components/drivers: earthquakes, rainfall, populations, traffic, data, demand, ...
Fields that use these models include:

1. **Natural sciences** (Lu [2012], Thayakaran and Ramesh [2013a], Thayakaran and Ramesh [2013b], Langrock, Borchers, and Skaug [2013])
2. **Signals and telecommunications** (Scott and Smyth [2003], Pan, Rao, Agarwal, and Gelfand [2016])
3. **Finance and economics** (Nasr and Maddah [2015])
4. **Queueing, inventory and reliability theory** (Arts [2017], Landon, Özekici, and Soyer [2013])

And in the actuarial literature?

MMPPs in actuarial literature - relatively undeveloped

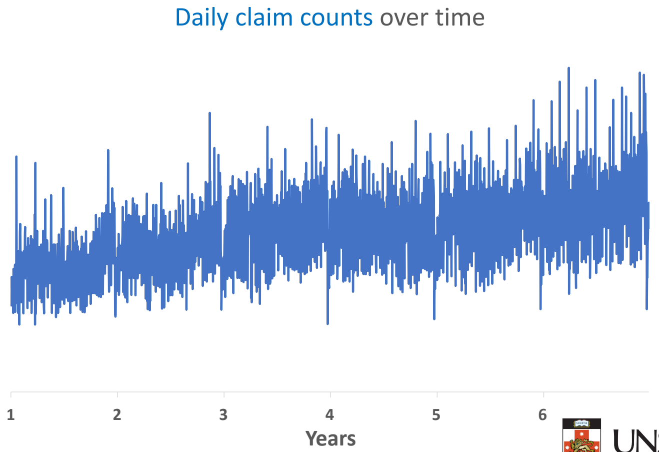
There are few papers applying MMPPs to insurance claim analysis, (particularly recently):

1. Elliott, Siu, and Yang [2007]
2. Guillou, Loisel, and Stupfler [2013]
3. Guillou, Loisel, and Stupfler [2015]

Why might this literature be underdeveloped?
(Note however that there is a substantial literature of such latent models in the more general family of HMM—see paper for details)

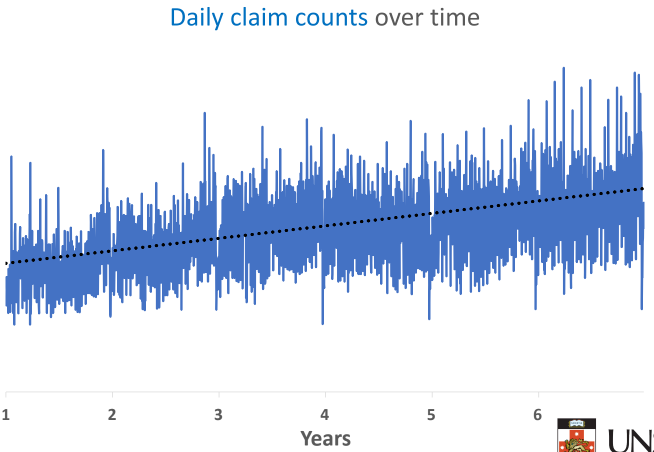
There are a number of barriers to practical implementation that need to be addressed first...

Domestic motor claim numbers over time

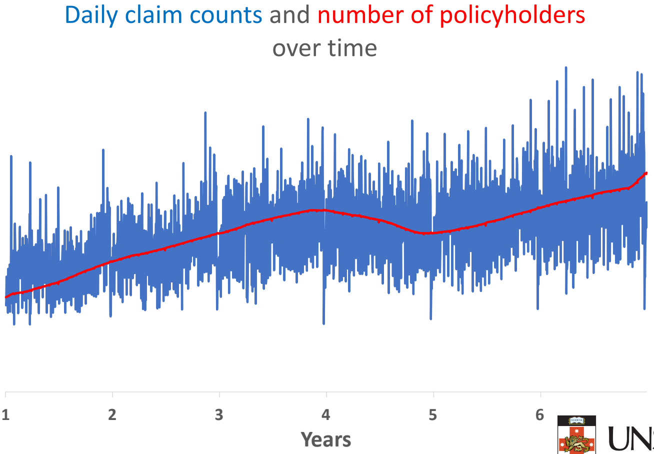


UNSW | AGSM Business School

Domestic motor claim numbers over time



. . . with number of policyholders over time



Barriers to practical implementation

What extensions are required for the MMPP micro-level model to be realistic and practicably useful?

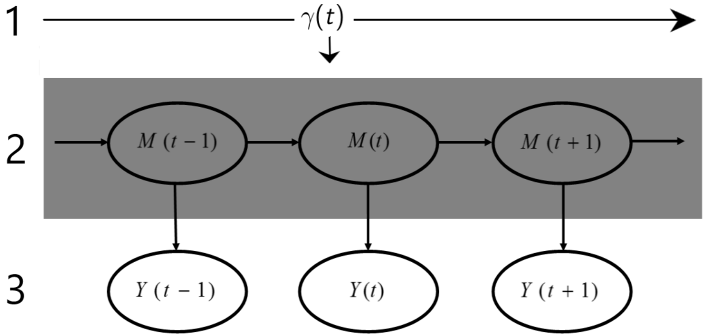
1. Flexible risk exposure/frequency perturbation measure
   1. Periodic (seasonality) ← contribution of Guillou et al. [2015]
   2. Non-periodic (number of policyholders, structural changes)
      ↑ our contribution
2. Numerical stability during the calibration process
3. Reasonable computation times for large insurance data sets

These are essentially our contributions in this paper.

Markov-modulated **non-homogeneous** Poisson process!

Figure: A Markov-modulated non-homogeneous Poisson process

Model Notation



1. $\gamma(t)$ is a general exposure measure ← what we model explicity
2. $M(t)$ is the Hidden Markov chain ← where the remainder goes
3. $Y(t)$ is the conditional Poisson process for claim arrivals

Again, a bit more formally...

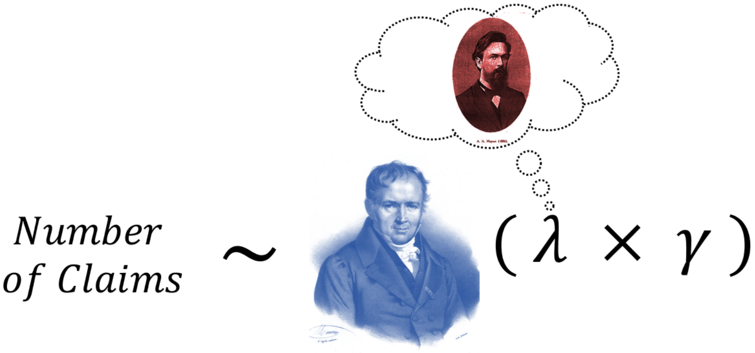
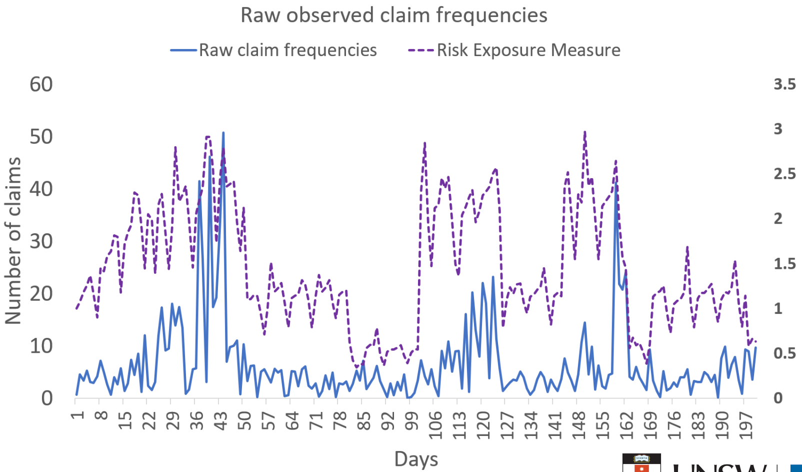
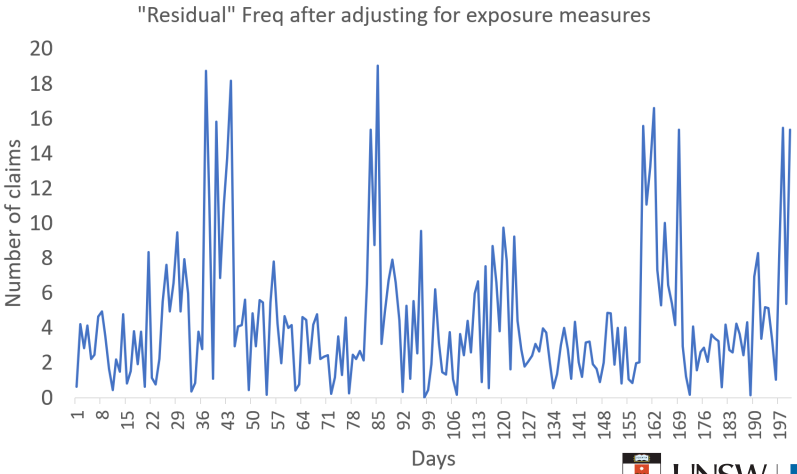


Figure: The “Formal” Definition for a MMNPP

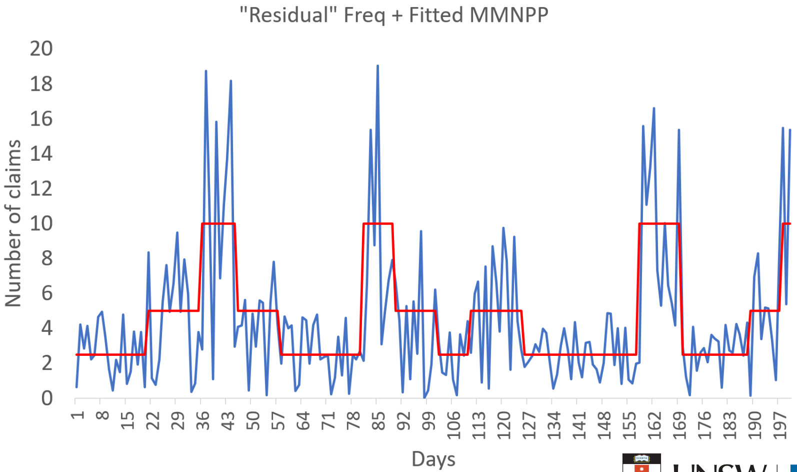
Toy illustration - Raw observed claim frequencies



Toy example - After accounting for risk exposure relativities

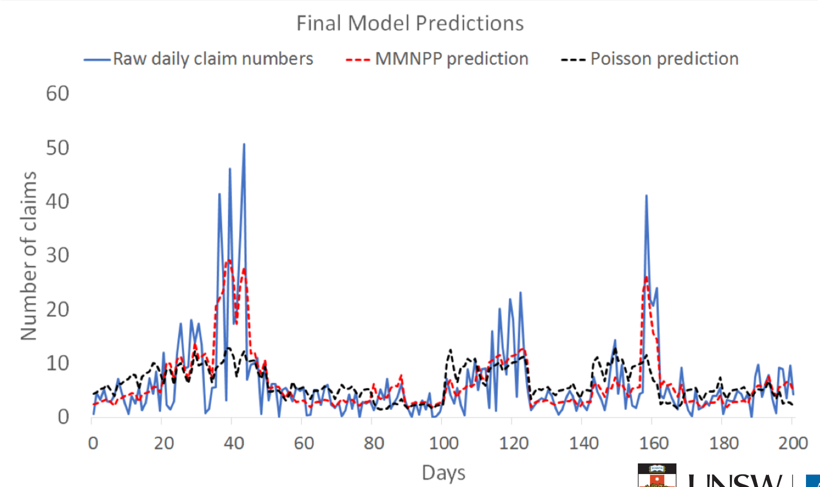


Toy example - After fitting the MMNPP model



UNSW | AGSM Business School

Toy example - Final Model Outputs



Theoretical contributions - A new EM algorithm

We derived a new EM algorithm, adapting results from Rydén [1996] and Roberts, Ephraim, and Dieguez [2006], in particular to the non-homogenous case. This algorithm

1. Resolved issues of **numerical stability** of large data sets, allowing the model to be implemented in standard software such as R or MATLAB.
2. Implemented several computational improvements/shortcuts that drastically **reduced calibration times**. (e.g. 4-5 hours for 700,000 claims)
3. Allowed for easy extraction of several **statistical quantities of interest** (more on this later...)

A simulation case study

Each period was set to be 100 time units.

Period	Raw Intensity	Exposure	"Base" λ	"True State"
1	0.5	1	0.5	1
2	1	1	1	2
3	1	2	0.5	1
4	2	1	2	3
5	10	10	1	2
6	10	20	0.5	1
7	10	5	2	3
8	0.75	1.5	0.5	1
9	0.1	0.1	1	2
10	0.1	0.05	2	3

Table: Simulated data set parameters

UNSW | AGSM Business School

Simulation calibration results

Regime	True Poisson Intensity	Calibrated Poisson Intensity
1	0.5	0.482
2	1	1.013
3	2	1.965

Table: True intensities versus calibrated intensities

We can also extract the probability of being in each state/regime at each claim arrival time.

State/Regime filtering

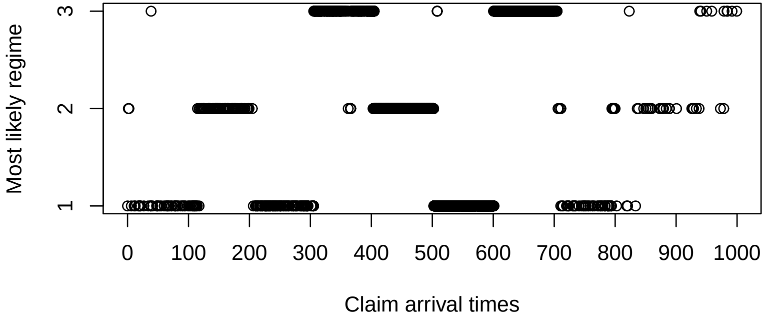


Figure: Plot of the most likely regime at each claim arrival time

The AUSI data set

The Allianz, University of New South Wales, Suncorp and Insurance Australia Group (AUSI) dataset is supported by the Australian Research Council and several major industry partners. It consists of

- 4 lines of business
  - Building and Contents
  - Domestic Motor
  - Public Liability
  - Compulsory Third Party
- Daily records
- Policy and Claim Characteristics

In the following, we will be looking at the Domestic Motor LoB.

What exposure measures should we adjust for?

An over-dispersed Poisson GLM was fit to daily claims data over a period of 6 years. Covariates were tested for statistical significance and where appropriate, bucketed for parsimony. The final model used the following features:

1. Number of policies in force
2. Various forms of seasonality:
    1. Weekday/Weekend
    2. Public Holiday
    3. Month
    4. Day of Month
3. Days since start of the period of investigation

UNSW | AGSM Business School

Do states/regimes actually exist in the data?

Some statistical results seem to indicate that even after consideration of the factors on the previous slide, regimes still exist.

- The dispersion parameter for the ODP GLM was 3.16, indicating **large over-dispersion**.
- A runs test was applied to the adjusted frequencies, indicating **persistence**. ← so there are at least 2 regimes
- An order selection methodology based on a recursive white-noise residual testing (in paper, Barlett B) was applied, and the resulting optimal order was **4**.

UNSW | AGSM
Business School

Final results - estimates and E-step estimators

The calibrated regime transition and claim intensities are

$$Q = \begin{bmatrix} -0.39 & 0.09 & 0.28 & 0.02 \\ 0.00 & -0.04 & 0.04 & 0.00 \\ 0.06 & 0.40 & -0.46 & 0.00 \\ 1.00 & 0.00 & 0.00 & -1.00 \end{bmatrix}, \quad \lambda = (135, 177, 204, 518) .$$

We can also extract some other information from the model

Quantity	Estimator
Number of changes to each state	$\begin{bmatrix} -20 & 7 & 12 & 1 \\ 4 & -79 & 75 & 0 \\ 14 & 73 & -87 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$
Proportion of time in each state	(2.3%, 88.9%, 8.7%, 0.0%)
Proportion of claims in each state	(1.7%, 88.5%, 9.8%, 0.0%)

Final results - Regime filtering

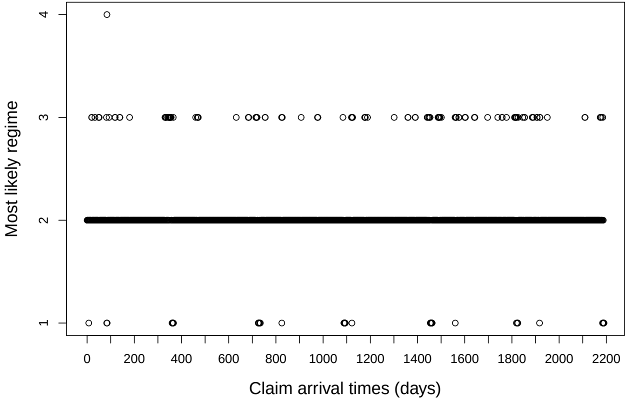


Figure: Plot of the regime with the highest probability per day

Summary

- Procedure is an easy extra step for practitioners, who are already used to model exposure γ
- MMNPP provides a simple modelling tool to bring persistence and overdispersion
- One way of identifying and dealing with outliers
- The calibration process informs the modelling of γ in an objective way (e.g. 27/12–31/12 insight)
- How much to put in γ depends on actuarial judgement

- Whether the MMNPP is used for predictions or not, its fitting is informative

References I

Joachim Arts. A multi-item approach to repairable stocking and expediting in a fluctuating demand environment. *European Journal of Operational Research*, 256(1):102–115, 2017.

Robert J Elliott, Tak Kuen Siu, and Hailiang Yang. Insurance claims modulated by a hidden marked point process. In *American Control Conference, 2007. ACC'07*, pages 390–395. IEEE, 2007.

A. Guillou, S. Loisel, and G. Stupfler. Estimating the parameters of a seasonal markov-modulated poisson process. *Statistical Methodology*, 26(0):103 – 123, 2015.

Armelle Guillou, Stéphane Loisel, and Gilles Stupfler. Estimation of the parameters of a markov-modulated loss process in insurance. *Insurance: Mathematics and Economics*, 53(2):388–404, 2013.

L. J. Halliwell. Chain-ladder bias: Its reason and meaning. *Casualty Actuarial Society*, 1(2):214–247, 2007.

References II

M. Kunkler. Modelling zeros in stochastic reserving models. *Insurance: Mathematics and Economics*, 34(1):23–35, 2004.

Joshua Landon, Süleyman Özekici, and Refik Soyer. A markov modulated poisson model for software reliability. *European Journal of Operational Research*, 229(2):404–410, 2013.

Roland Langrock, David L Borchers, and Hans J Skaug. Markov-modulated nonhomogeneous poisson processes for modeling detections in surveys of marine mammal abundance. *Journal of the American Statistical Association*, 108(503):840–851, 2013.

Huijuan Liu and Richard Verrall. Predictive distributions for reserves which separate true ibnr and ibner claims. *ASTIN Bulletin*, 39:35–60, 2009.

Shaochuan Lu. Markov modulated poisson process associated with state-dependent marks and its applications to the deep earthquakes. *Annals of the Institute of Statistical Mathematics*, 64(1):87–106, 2012.

References III

Walid W Nasr and Bacel Maddah. Continuous (s, s) policy with mmpp correlated demand. *European Journal of Operational Research*, 246(3): 874–885, 2015.

Jiangwei Pan, COM Vinayak Rao, EDU Pankaj K Agarwal, and Alan E Gelfand. Markov-modulated marked poisson processes for check-in data. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2244–2253, 2016.

P. Parodi. Triangle-free reserving. *British Actuarial Journal*, 19(01):168–218, 2014.

A. E Renshaw. Modelling the claims process in the presence of covariates. *ASTIN Bulletin*, 24(2):265–299, 1994.

William J.J. Roberts, Y. Ephraim, and E. Dieguez. On rydén's em algorithm for estimating mmpps. *Signal Processing Letters, IEEE*, 13(6):373–376, 2006.

Tobias Rydén. An em algorithm for estimation in markov-modulated poisson processes. *Computional Statistics & Data Analysis*, 21(4):431–447, 1996.

References IV

S.L. Scott and P. Smyth. The markov modulated poisson process and markov poisson cascade with applications to web traffic modelling. *Bayesian Statistics*, 7, 2003.

Greg Taylor. Maximum likelihood and estimation efficiency of the chain ladder. *Astin Bulletin*, 41(01):131–155, 2011.

Greg Taylor and Gráinne McGuire. Loss reserving with glms: A case study. *Casualty Actuarial Society Discussion Paper Program, Applying and Evaluating Generalised Linear Models*, 2004.

R. Thayakaran and N. I. Ramesh. Multivariate models for rainfall based on markov modulated poisson processes. *Hydrology Research*, 44(4):631–643, 2013a. ISSN 0029-1277. doi: 10.2166/nh.2013.180. URL http://hr.iwaponline.com/content/44/4/631.

R Thayakaran and NI Ramesh. Markov modulated poisson process models incorporating covariates for rainfall intensity. *Water Science and Technology*, 67(8):1786–1792, 2013b.

References V

T. Verdonck, M. V. Wouwe, and J. Dhaene. A robustification of the chain-ladder method. *North American Actuarial Journal*, 13(2):280–298, 2009.