

Joint estimation of insurance loss development factors using Bayesian hidden Markov models

June 28, 2024

¹ Loss development modelling is the actuarial practice of predicting the total *ultimate*
² losses incurred on a set of policies once all claims are reported and settled. This
³ poses a challenging prediction task as losses frequently take years to fully emerge
⁴ from reported claims, and not all claims might yet be reported. Loss development
⁵ models frequently estimate a set of *link ratios* from insurance loss triangles, which
⁶ are multiplicative factors transforming losses at one time point to ultimate. However,
⁷ link ratios estimated using classical methods typically underestimate ultimate losses
⁸ and cannot be extrapolated outside the domains of the triangle, requiring extension
⁹ by *tail factors* from another model. Although flexible, this two-step process relies
¹⁰ on subjective decision points that might bias inference. Methods that jointly esti-
¹¹ mate ‘body’ link ratios and smooth tail factors offer an attractive alternative. This
¹² paper proposes a novel application of Bayesian hidden Markov models to loss devel-
¹³ opment modelling, where discrete, latent states representing body and tail processes
¹⁴ are automatically learned from the data. The hidden Markov development model is
¹⁵ found to perform comparably to, and frequently better than, the two-step approach
¹⁶ on numerical examples and industry datasets.

¹⁷ *Keywords:* actuarial science, loss reserving, mixture modelling, time series

¹⁸ **1 Introduction**

¹⁹ Loss development modelling in actuarial science is the practice of predicting the total
²⁰ losses incurred from all reported and settled claims on a set of insurance policies. At
²¹ any moment in time, these so-called *ultimate* losses, less the losses paid on already
²² reported claims, constitute the overall loss reserves, generally referred to as *incurred*
²³ *but not reported* (IBNR). The IBNR amount includes future payments on claims
²⁴ already reported ('incurred by not enough reported'), as well as payments for claims
²⁵ not yet reported ('incurred by not yet reported'). In general or non-life insurance,
²⁶ IBNR can be substantial as it may take years for all claims on a set of policies to be
²⁷ reported and settled. Therefore, the accurate estimation of IBNR is key to evaluating
²⁸ an insurance company's performance and solvency (Beard, 1960; Bornhuetter and
²⁹ Ferguson, 1972; Friedland, 2010; England and Verrall, 2002; Wüthrich and Merz,
³⁰ 2008). Transparent assessment of the uncertainty in IBNR is also critical to the
³¹ legal responsibilities of insurers and reinsurers, such as for the Solvency II directive
³² (England et al., 2019; Fröhlich and Weng, 2018; Munroe et al., 2018).

³³ Estimation of loss reserves is a challenging prediction problem that has received con-
³⁴ siderable attention from actuaries, econometricians, and statisticians for decades (e.g.
³⁵ for some key developments, see Bornhuetter and Ferguson, 1972; Clarke and Harland,
³⁶ 1974; Taylor, 1977; Taylor and Ashe, 1983; Mack, 1993; Barnett and Zehnwirth, 2000;
³⁷ England and Verrall, 2001, 2002; Taylor et al., 2003; Wüthrich and Merz, 2008). From
³⁸ early deterministic and algebraic approaches (e.g. Scurfield, 1968; Bornhuetter and
³⁹ Ferguson, 1972; Clarke and Harland, 1974; Taylor, 1977), loss development modelling
⁴⁰ has advanced to utilise a wide, and growing, variety of statistical methods, including
⁴¹ parametric and non-parametric regressions (Mack, 1994; England and Verrall, 2001,
⁴² 2002; Lally and Hartman, 2018), Bayesian estimation (England and Verrall, 2002;
⁴³ De Alba, 2002; Zhang et al., 2012; Meyers, 2015), differential equations (Gesmann
⁴⁴ and Morris, 2020), and neural networks and machine learning methods (Kunce and

⁴⁵ Chatterjee, 2017; Kuo, 2019; Al-Mudafer et al., 2022). Despite their differences, the
⁴⁶ majority of these approaches are applied to aggregated insurance risks, typically dis-
⁴⁷ played as a triangular matrix of experience periods and development periods or lags
⁴⁸ called a *loss triangle* (see Figure 1). Each experience period represents losses on a
⁴⁹ distinct set of policies, commonly losses from all policies with accidents occurring in a
⁵⁰ specific time period, and each development lag records the cumulative or incremental
⁵¹ development of those losses through time since reporting.

⁵² A key inferential quantity from any loss development model is the set of multiplicative
⁵³ factors transforming the losses at development period one to the ultimate losses at
⁵⁴ development period ∞ , known as the *loss development factors* or *link ratios*. While
⁵⁵ many models include link ratios as an explicit parameter to be inferred, notably the
⁵⁶ family of chain ladder methods (e.g. Mack, 1993; England and Verrall, 2002), others
⁵⁷ derive link ratios as a generated quantity (England and Verrall, 2001; Meyers, 2015).
⁵⁸ Ideally, link ratios will smoothly decline over time towards unity, but in practice
⁵⁹ often include periods of volatility, particularly for early development periods, and may
⁶⁰ further encode systematic and non-systematic effects across experience periods or date
⁶¹ of development evaluation (e.g. the influence of the Covid-19 pandemic). Depending
⁶² on the degree of volatility and the amount of data available, link ratios estimated
⁶³ from a triangle of finite risks will be insufficient to infer the ultimate losses for each
⁶⁴ experience period, because losses may still be emerging at the latest development
⁶⁵ period available in the data. Thus, estimation of ultimate losses will require extending
⁶⁶ the link ratios to outside the domains of the focal triangle to include *tail factors*. Like
⁶⁷ general loss development modelling, tail factor estimation has had its own expansive
⁶⁸ history of deterministic and stochastic methods (CAS Tail Factor Working Party,
⁶⁹ 2013), and is of particular importance in ‘long-tailed’ lines of business, such as workers’
⁷⁰ compensation or general liability, where experience periods might display continued
⁷¹ loss cost growth and volatility at relatively late development lags.

72 Of the various approaches to calculating tail factors, many use a second model fit
73 to a portion of the focal triangle that conveys how the triangle may behave in the
74 tail ([CAS Tail Factor Working Party, 2013](#)), or to the set of link ratios directly from
75 the first model. These models typically infer a parametric, monotonically increasing
76 growth curve of losses from the training data, such as various forms of inverse power
77 curves (e.g. [Sherman, 1984](#); [Evans, 2015](#); [Clark, 2017](#)). The link ratios derived from
78 this tail model are then appended to the link ratios estimated from the primary
79 loss development model to produce predictions an arbitrary number of development
80 lags into the future. For clarity, the primary link ratios will be referred to in this
81 paper as the ‘body’ link ratios to distinguish them from link ratios estimated in
82 the tail. Crucially, this two-step process includes a number of subjective decisions.
83 The body-to-tail switch-over development lag is frequently chosen to reflect when the
84 development process settles to a reasonable plateau, while the training windows for
85 both models will be chosen based on which sections of the triangle best match each
86 model’s assumptions. For instance, including periods of non-monotonic growth into
87 tail models might bias tail factors unreasonably high. These decisions are difficult
88 to reproduce and opens analysts to many ‘researchers degrees of freedom’ ([Simmons
89 et al., 2011](#)) – selecting one approach among many possible alternatives that might
90 have non-trivial impacts on predicting ultimate loss.

91 Methods that simultaneously estimate body and tail link ratios present an attrac-
92 tive alternative to traditional two-step processes. However, only a few solutions have
93 been proposed. [England and Verrall \(2001\)](#) presented a generalised additive model
94 for smooth estimation of loss development factors that can be extrapolated to points
95 further than the existing data. Although flexible and able to integrate different func-
96 tional forms and covariates of loss development processes, this approach still neces-
97 sitates careful selection of training data to include in the model so that regions of
98 volatility do not bias loss development curves. Generalised additive models further
99 require selecting the family of splines and number of knots to apply, which may lead

100 to another set of decisions analysts must act on. [Zhang et al. \(2012\)](#), alternatively,
101 implemented a hierarchical Bayesian logistic growth curve model to cumulative loss
102 data. However, fitting a single parametric curve that assumes monotonicity in ex-
103 pectation might under-estimate systematic volatility in portions of the triangle that
104 analysts do not want to label as residual noise. Finally, [Verrall and Wüthrich \(2012\)](#)
105 use reversible jump Markov chain Monte Carlo to combine a Bayesian chain ladder
106 model, applied before some tail cut-off point, with an exponential decay process after
107 the cut-off, and allow the model to infer where the cut-off should occur. Additionally,
108 [Verrall and Wüthrich \(2015\)](#) demonstrate how the same model can be estimated in
109 a Bayesian model averaging context. Although this approach is arguably the most
110 flexible, it requires either bespoke sampling algorithms (i.e. reversible jump Monte
111 Carlo) or multiple model fits (for model averaging purposes), which may dissuade
112 analysts and researchers.

113 This paper proposes the use of hidden Markov models to simultaneously estimate
114 body and tail link ratios in a single model from a variety of loss triangles. Hidden
115 Markov models are primarily discrete mixture models postulating an unknown latent
116 state underlying and generating patterns of observed data, and have found a multi-
117 tude of applications, from speech recognition (e.g. [Rabiner, 1989](#)) to animal behaviour
118 (e.g. [Leos-Barajas et al., 2017](#)). Indeed, Markov processes have been previously used
119 in micro-level claim modelling (e.g. [Hesselager, 1994](#)), but have not been applied to
120 aggregate insurance loss triangles. As described in this paper, hidden Markov devel-
121 opment models decompose a loss triangle into a sequence of body and tail processes.
122 Hidden Markov models are easily fit in existing, open-source software, and can cater
123 for complex data-generating assumptions, such as understanding the impact of covari-
124 ates or including non-parametric patterns of loss development ([England and Verrall,](#)
125 [2001](#)), Moreover, the approach presented here is flexible enough to integrate any of
126 the parametric and non-parametric models that have been presented in the actuarial
127 literature for modelling body and tail processes.

128 Below, the hidden Markov model is formulated and validated with simulated ex-
129 amples, and subsequently compared to a traditional two-step process using a num-
130 ber of data sets. The hidden Markov models are fit using Bayesian estimation
131 in Stan ([Carpenter et al., 2017](#)) following a modern Bayesian workflow ([Gelman](#)
132 [et al., 2020](#)). All Python and Stan code, and all datasets, to reproduce the results
133 are accessible at the Github repository <https://github.com/LedgerInvesting/>
134 [hidden-markov-development-2024](#).

135 2 Materials and methods

136 2.1 Hidden Markov development model

137 Consider the typical loss development context, where a large set of homogeneous
138 insurance risks (e.g. private car insurance policies) are aggregated into a loss devel-
139 opment triangle with cumulative loss amounts denoted \mathcal{Y} , defined by

$$\mathcal{Y} = \{y_{ij} : i = 1, \dots, N; j = 1, \dots, N - i + 1\} \quad (1)$$

140 where $i = (1, \dots, N)$ denotes experience periods, most commonly indexing all claims
141 occurring during period i , and $j = (1, \dots, M)$ denotes development periods or lags. For
142 a particular point in time, development information for period i is only known up to
143 lag $j = N - i + 1$, and therefore \mathcal{Y} represents the left upper diagonal of the loss triangle.
144 The complementary, lower diagonal triangle, denoted $\tilde{\mathcal{Y}}$, with $(N + M)^{\frac{N+M-1}{2}}$ data
145 points, is unknown and the goal of prediction.

146 The generative process considered here (Figure 1) is that cumulative losses in \mathcal{Y}
147 develop according to i) a period that is characterised by largely, but not strictly,

148 monotonically increasing losses (the body), followed by *ii*) a period of smooth growth
 149 to ultimate (the tail). Traditional methods treat estimation of body and tail as a
 150 two-step process. By contrast, the hidden Markov development model introduces a
 151 latent, discrete state $\mathbf{z} = (z_{11}, z_{12}, \dots, z_{ij}, \dots, z_{NM}) \in \{1, K\}$ with $K = 2$ that takes
 152 value $z_{ij} = 1$ if the body process generated the losses in the i th accident period and
 153 j th development period, and $z_{ij} = 2$ if the tail process generated the losses. The
 154 latent state at one time point, z_{ij} , is dependent on the state at the previous time
 155 point, z_{ij-1} , and subsequent states are connected via a state transition matrix, Θ_{ij} .
 156 Depending on the latent state and any other parameters ϕ , the marginal likelihood of
 157 the data, $p(y_{ij} | z_{ij}, \phi)$, is given by suitable *emission* distributions. Emission distribu-
 158 tions are the observation data distributions, which in loss development modelling are
 159 typically positive-bound, continuous probability density functions, such as lognormal
 160 or Gamma. In this paper, all models use lognormal distributions as the likelihood
 161 distribution.

162 The hidden Markov development model can be written generally as:

$$\begin{aligned}
 y_{ij} &\sim \begin{cases} \text{Lognormal}(\mu_1, \sigma_{ij}) & z_{ij} = 1 \\ \text{Lognormal}(\mu_2, \sigma_{ij}) & z_{ij} = 2 \end{cases} \\
 z_{ij} &\sim \text{Categorical}(\Theta_{ij}^{z_{ij-1}}) \\
 \Theta_{ij} &= \begin{pmatrix} \pi & 1 - \pi \\ \nu & 1 - \nu \end{pmatrix} \\
 \log \frac{\pi}{1 - \pi} &\sim \text{Normal}(0, 1) \\
 \log \frac{\nu}{1 - \nu} &\sim \text{Normal}(0, 1)
 \end{aligned} \tag{2}$$

163 The cumulative losses at accident period i and development period j are assumed
 164 positive-bound, lognormally-distributed random variates, with log-scale location μ_1

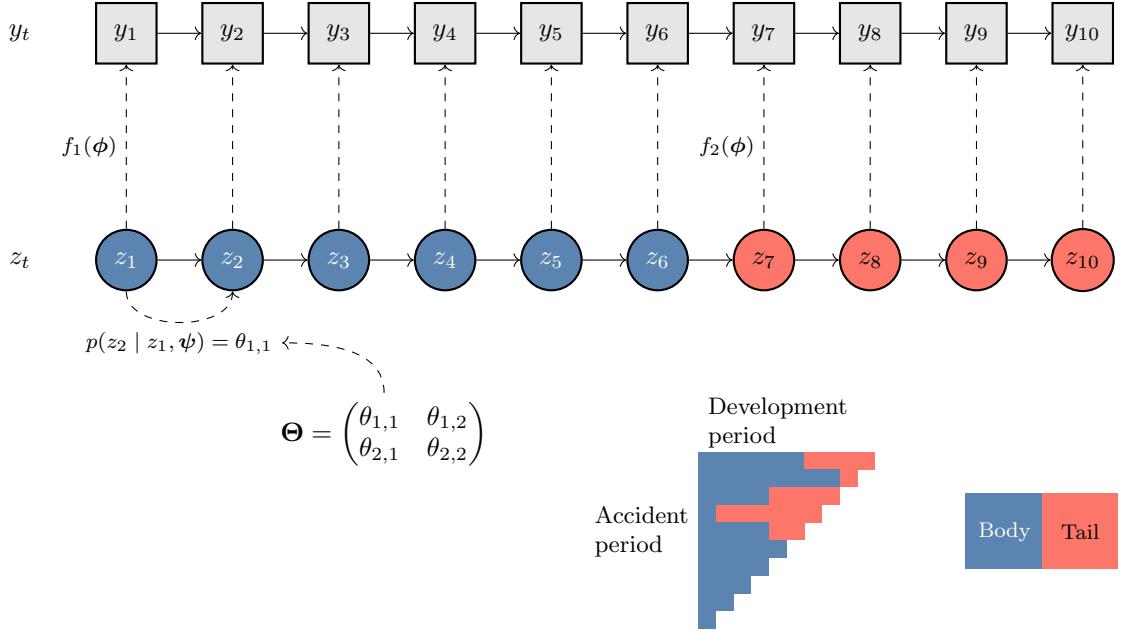


Figure 1: A schematic of the hidden Markov model process. A loss triangle of observed data is shown with 10 accident periods, and each development period generated from two processes: the body (blue) and the tail (orange). The dynamics of the body and tail states vary over accident periods. For a single accident period, observed losses at development lag t , y_t , are shown as grey squares, and are assumed generated from the latent discrete random variable z_t (circles), which transition according to the probabilities in matrix Θ .

if the process is in the body or μ_2 if the process is in the tail, and with scale σ_{ij} .
 The latent state z_{ij} determines the body or tail state at each time point, and follows a categorical distribution with unit simplex probabilities determined by row z_{ij-1} of the state transition matrix Θ_{ij} . As discussed below, the state transition matrix may be time-homogeneous or -inhomogeneous, depending on the context. In Θ , π denotes $p(z_{ij} = 1 | z_{ij-1} = 1)$, the probability of staying in the body process at lag j , and $\nu = p(z_{ij} = 2 | z_{ij-1} = 2)$. Their complements, $1 - \pi$ and $1 - \nu$, represent the probabilities of transitioning from body to tail, and tail to body, respectively.

173 The functions $\boldsymbol{\mu} = (\mu_1, \mu_2)$ provide the conditional expectations of the two processes
 174 assumed to describe the data. The models here use two canonical body and tail de-
 175 velopment models: the chain ladder model for body loss development factors ([Mack](#),
 176 [England and Verrall, 2002](#)), and an exponential decay curve following the gen-
 177 eralised Bondy model for the tail process ([CAS Tail Factor Working Party, 2013](#)).
 178 The full model specification is completed by choosing these functional forms, along
 179 with the functional form for the variance and the remaining prior distributions.

$$\begin{aligned}
 \mu_{1_{ij}} &= \log(\alpha_{j-1} y_{ij-1}) \quad \forall j > 1 \\
 \mu_{2_{ij}} &= \log(\omega^{\beta^j} y_{ij-1}) \quad \forall j > 1 \\
 \sigma_{ij}^2 &= \exp(\gamma_1 + \gamma_2 j + \ln(y_{ij-1})) \\
 \log \boldsymbol{\alpha}_{1:M-1} &\sim \text{Normal}(0, 1/1 : M - 1) \\
 \log \omega &\sim \text{Normal}(0, 1) \\
 \log \frac{\beta}{1 - \beta} &\sim \text{Normal}(0, 1) \\
 \boldsymbol{\gamma}_{1:2} &\sim \text{Normal}(0, 1)
 \end{aligned} \tag{3}$$

180 Due to the multiplicative autoregressive nature of typical loss development models,
 181 the first data point is not modelled, and the hidden Markov process is assumed to
 182 start in the body state. The $M - 1$ body link ratios are given by the vector $\boldsymbol{\alpha}$, and
 183 the tail link ratios are given by ω^{β^j} , for any j , allowing extrapolation out to arbitrary
 184 development periods. The parameter ω is constrained to be strictly greater than
 185 1.0, such that growth is monotonically increasing, and β is constrained to lie in the
 186 interval $(0, 1)$, to avoid tail factors growing without bound to ∞ . The expression for
 187 the variance encodes the assumption of less volatility at higher development periods,
 188 and is proportional to the losses at the previous time point.

189 The prior distribution on $\boldsymbol{\alpha}$ is regularised towards a link ratio of 1 in inverse proportion

to the development period. This assumption is imposed to ensure link ratios, when $z_{ij} = 2$ (i.e. the latent state is in the tail process), do not become unrealistically large due to the resulting non-identifiability. Even when the tail process is most likely, some samples could be generated from the body process because z_{ij} is a random variable, and large values of α_{ij-1} might have unwanted influence on the predictions. This is akin to the use of pseudo-priors in Bayesian mixture models (Carlin and Chib, 1995), which ensure parameters in the mixture not being sampled do not become implausible.

2.1.1 Model variants

Three variations of the hidden Markov model are used (Table 1). The base model (referred to as HMM) has homogeneous transition matrix Θ , and sets ν to zero. This implies that once the tail process is active, the model cannot switch back to the body process. This is the primary assumption underlying tail modelling generally: at some development point, the losses smoothly develop to ultimate. Secondly, the HMM- ν model estimates ν , allowing for tail processes to switch back to body processes. Some triangles may illustrate unexpected late-development volatility, at which point the more flexible body process is a better explanation of the data. Finally, the HMM-lag variant allows π to vary across development periods, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{j-1}, \dots, \pi_{M-1})$, on which an ordered assumption is imposed such that the probability of transitioning to the tail increases with development period, meaning $\pi_{j-1} < \pi_{j-2}$. Many extensions of these variants are possible, including the addition of covariates on the estimation of Θ or other parametric or non-parametric forms. This paper restricts focus on these three variants as they present the simplest use cases to compare to the two-step process. Although the HMM-lag and HMM- ν variant could be combined, early tests of the hidden Markov development model by the author indicated that this model risked unidentifiability without further data or covariates.

Name	Θ_{ij}
HMM	$\begin{pmatrix} \pi & 1 - \pi \\ 0 & 1 \end{pmatrix}$
HMM- ν	$\begin{pmatrix} \pi & 1 - \pi \\ \nu & 1 - \nu \end{pmatrix}$
HMM-lag	$\begin{pmatrix} \pi_{ij-1} & 1 - \pi_{ij-1} \\ 0 & 1 \end{pmatrix}$

Table 1: The three hidden Markov model transition matrix variants used in the examples.

216 **2.1.2 Estimation**

217 We fit the models in Stan ([Carpenter et al., 2017](#)) using Bayesian inference via Hamil-
 218 tonion Monte Carlo, via the `cmdstanpy` ([Stan Development Team, 2024b](#)) Python
 219 package and `cmdstan` ([Stan Development Team, 2024a](#)). Stan requires specifying a
 220 statement proportional to the joint log density of the data and parameters. For the
 221 above model, the log likelihood for a single data point, grouping all transition matrix
 222 parameters as $\psi = \{\pi, \nu\}$ and emission distribution parameters as $\phi = \{\alpha, \omega, \beta, \gamma\}$,
 223 can be factored as:

$$p(y_{ij}, z_{ij}, z_{ij-1}, \phi, \psi) = p(y_{ij} | z_{ij}, \phi)p(z_{ij} | z_{ij-1}, \psi)p(z_{ij-1} | \psi)p(\psi)p(\phi) \quad (4)$$

224 However, it is often more computationally efficient to marginalise the latent discrete
 225 parameters z_{ij} and z_{ij-1} out of the density, leading to a double summation over
 226 possible states for z_{ij} and z_{ij-1} :

$$\begin{aligned}
p(y_{ij}, \boldsymbol{\phi}, \boldsymbol{\psi}) &= \sum_{k=1}^K p(y_{ij} | z_{ij} = k, \boldsymbol{\phi}) p(\boldsymbol{\phi}) \\
&\quad \sum_{h=1}^K p(z_{ij} = k | z_{ij-1} = h, \boldsymbol{\psi}) p(z_{ij-1} = h | \boldsymbol{\psi}) p(\boldsymbol{\psi})
\end{aligned} \tag{5}$$

227 This recursive estimation over possible states is the forward algorithm (Rabiner,
 228 1989). After model fitting, the hidden states on the training data can be recov-
 229 ered using the Viterbi algorithm (Rabiner, 1989), which provides the most likely joint
 230 sequence of latent states that generated the data. For future data, new samples of
 231 z_{ij} can be taken from a categorical distribution with estimated parameters of the
 232 transition matrix Θ_{ij} for that data point.

233 2.2 Two-step approach

234 The hidden Markov development model is compared to a more traditional two-step
 235 modelling approach. Denote $\tau \in \{2, \dots, M\}$ the final body training data point, and
 236 $\boldsymbol{\rho} = (\rho_1, \rho_2) \in \{2, \dots, M\}$, where $\rho_1 < \rho_2$, a vector of tail start and end training
 237 window development points, respectively. While these constants could in theory vary
 238 over experience periods, there is typically insufficient data to do so. In the traditional
 239 approach, the chain ladder method is first fit to training data up to development
 240 period τ inclusive and predictions of the lower diagonal loss triangle are made up
 241 to and including τ , only. Secondly, the tail model is fit to data lying within the
 242 development period interval $[\rho_1, \rho_2]$, and predictions from τ made to some arbitrary
 243 development lag, j^* . The challenge and art of this two-step process is in finding a
 244 value for $\tau + 1$ that identifies the development lag where losses are plateauing in
 245 the tail, and finding values for $\rho_{1,2}$ that identify a suitable decaying curve of link
 246 ratios. While $\tau = \rho_1$ in some cases, more generally $\rho_1 \leq \tau$. Note, this presents a
 247 difference between the two-step and hidden Markov model approaches: the hidden

248 Markov model identifies clear body-to-tail switch-over points, whereas the the interval
249 of data delineated by $\rho_{1,2}$ in the two-step process might overlap the final body training
250 point, τ .

251 To maintain comparability to the hidden Markov model above, the two-step approach
252 is also implemented in Stan and both models estimated with a shared variance. In
253 real applications, uncertainty from body to tail models is typically ignored, but this
254 may unfairly penalise the two-step process compared to the hidden Markov models.
255 The two-step approach differs from equation 2 in only two ways:

$$y_{ij} \sim \begin{cases} \text{Lognormal}(\mu_1, \sigma_{ij}) & j \leq \tau \\ \text{Lognormal}(\mu_2, \sigma_{ij}) & \rho_1 \leq j \leq \rho_2 \end{cases} \quad (6)$$

$$\log \boldsymbol{\alpha}_{1:\tau-1} \sim \text{Normal}(0, 1)$$

256 where now the decision between the two models is decided by τ and $\rho_{1,2}$. The further
257 exception in the two-step approach is the use of a standard normal prior on the log-
258 scale link ratios, rather than the constrained prior used in the hidden Markov models.
259 The two-step model's $\tau - 1$ link ratios are all estimated directly and do not suffer
260 from non-identifiability.

261 2.3 Simulation-based calibration

262 To validate the hidden Markov model, we use both simulated examples (shown in the
263 results) to build intuition, alongside simulation-based calibration (Talts et al., 2018;
264 Modrák et al., 2023). Simulation-based calibration leverages the self-consistency of
265 the Bayesian joint distribution of parameters and data: fitting a Bayesian model to
266 datasets generated from its prior predictive distribution and averaging over datasets
267 should return the prior distribution.

268 Focusing on the base HMM model variant, 1000 full-triangles with $N = M = 10$ were
269 generated from the prior predictive distribution, and the HMM model fit to each
270 upper diagonal \mathcal{Y} . The prior distributions were the same as in equation 2, except
271 for the priors on (γ_1, γ_2) , which were given more informative normally-distributed
272 priors with locations and scales of $(-3, -0.25)$ and $(-1, 0.1)$, respectively. Due to
273 the multiplicative autoregressive forms in the location and scales of the likelihood in
274 equation 2, particularly large values for σ can cause overflow in the sampled data.

275 Each model was summarised by calculating the rank statistics of quantities of in-
276 terest. The rank statistic is the number of times a simulated value is greater than
277 the posterior values, and should be approximately uniformly distributed if the model
278 has been implemented correctly and is unbiased (Talts et al., 2018). To reduce the
279 autocorrelation in the posterior distributions, the posteriors were thinned to every
280 10th posterior draw. Rank statistics were calculated for each parameter in the HMM
281 model, as well as the joint log likelihood and an the ultimate loss prediction at data
282 point ($i = 1, j = 10$), since Modrák et al. (2023) recommend using test quantities
283 that average over the entire data space in evaluating SBC.

284 2.4 Datasets and model performance

285 We compared the hidden Markov development model to the two-step approach in two
286 different ways. Firstly, we estimated both models on the 200 industry paid loss trian-
287 gles from Meyers (2015), 50 triangles for the four lines of business: private passenger
288 auto (PP), workers compensation (WC), commerical auto (CA), and other occur-
289 rence (i.e. general) liability (OO). We removed triangles with zero loss values first,
290 leaving 49 CA triangles, 48 OO triangles, and 50 PP and WC triangles. Each trian-
291 gle covers 10 years of historical accident periods and development, allowing splitting
292 triangles into an upper diagonal of training data, \mathcal{Y} , and lower diagonal of test data,

²⁹³ $\tilde{\mathcal{Y}}$. As these are yearly triangles, and have been chosen for previous model validation
²⁹⁴ exercises ([Meyers, 2015](#)), it is reasonable to assume that the losses at development
²⁹⁵ period 10 are close to ultimate. For the two-step approach, we inspected the mean
²⁹⁶ and standard deviations of the empirical link ratios across triangles (shown in Figure
²⁹⁷ [2](#)), and selected $\tau = 6$ and $\rho = (4, 10)$. These were chosen given that the link ratios
²⁹⁸ showed smooth patterns of decay from approximately development period 4 onwards,
²⁹⁹ and the triangles were close, on average, to their values at period 10 by development
³⁰⁰ period $\tau + 1 = 7$.

³⁰¹ When fitting the models to industry triangles, a small number of hidden Markov and
³⁰² two-step models failed due to producing very large posterior predictions on out-of-
³⁰³ sample data, which numerically overflowed. The multiplicative autoregressive nature
³⁰⁴ of both the hidden Markov and two-step models mean that large predictions at one
³⁰⁵ time point can quickly compound to unrealistic and computationally-unstable values.
³⁰⁶ For this reason, we capped the predictions at 100 times the maximum value across
³⁰⁷ the training and test data for each triangle.

³⁰⁸ Secondly, we used five triangles presented in the relatively recent literature on papers
³⁰⁹ on loss development modelling, including tail development, that provide more histori-
³¹⁰ cal data than the industry triangles: the long-tailed liability and short-tailed property
³¹¹ quarterly triangles from medium-size insurers in [Balona and Richman \(2022\)](#), the an-
³¹² nual liability triangle in [Merz and Wüthrich \(2015\)](#), the Swiss annual liability triangle
³¹³ from [Gisler \(2009\)](#), and the annual liability triangle in [Verrall and Wüthrich \(2012\)](#).
³¹⁴ These triangles had between 17 and 22 periods of data. For each triangle, we used
³¹⁵ the latest diagonal as the test dataset to evaluate predictive performance. For the
³¹⁶ two-step modelling approach, we chose the τ and ρ constants by visual inspection of
³¹⁷ the empirical link ratios, selecting $(12, 5, 5, 5, 12)$ for τ values for each triangle, and
³¹⁸ $[(4, 16), (3, 20), (3, 21), (3, 16), (10, 21)]$ for ρ values, respectively. The link ratios are
³¹⁹ displayed in Figure [8](#).

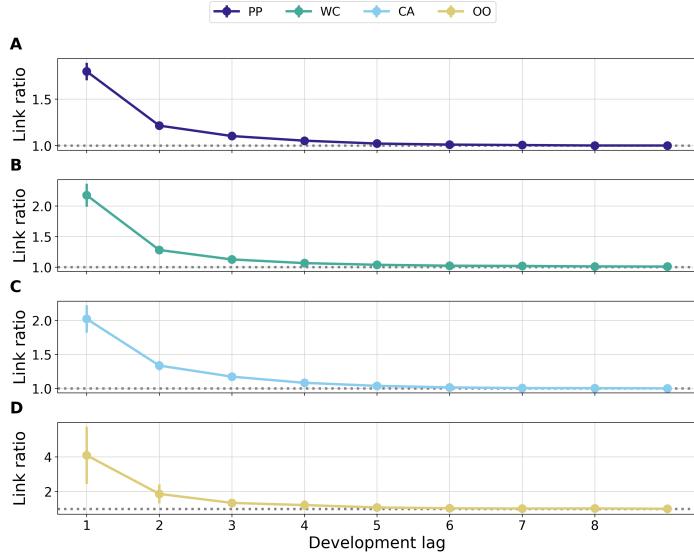


Figure 2: The empirical link ratios by line of business (panels A-D) in the Meyers (2015) dataset. Points indicate the mean across triangles, and vertical line segments show 2 standard deviations.

320 **2.4.1 Model performance**

321 Model performance can be split into two facets: accuracy and calibration. We sum-
 322 marised model accuracy using two metrics applied purely to the out-of-sample data:
 323 the expected log predictive density (ELPD) and the root mean square error (RMSE).
 324 The ELPD (Vehtari et al., 2017) is based on the logarithmic scoring rule, and is de-
 325 fined for a single triangle as the joint log likelihood of the out-of-sample data. In this
 326 way, it more heavily penalises models further from the true data generating process.
 327 Across each accident period i and development period j in $\tilde{\mathcal{Y}}$, we take the sum of log
 328 likelihood values marginalized across the posterior samples:

$$\begin{aligned}
\text{ELPD} &= \sum_{i=1}^N \sum_{j=N-i+1}^M \log p(\tilde{y}_{ij} | \mathcal{Y}) \\
&= \sum_{i=1}^N \sum_{j=N-i+1}^M \log \int p(\tilde{y}_{ij} | \theta) p(\theta | \mathcal{Y}) d\theta \\
&\approx \sum_{i=1}^N \sum_{j=N-i+1}^M \frac{1}{S} \sum_{s=1}^S \log p(\hat{y}_{ij}^{(s)} | \mathcal{Y})
\end{aligned} \tag{7}$$

where $p(\tilde{y}_{ij} | \mathcal{Y})$ is the posterior predictive distribution for the i th accident period at lag j , θ is used to generically refer to all model parameters, i.e. $\theta = \{\phi, \psi, z\}$, and the super-script in $\tilde{y}_{ij}^{(s)}$ denotes the s th sample from the posterior distribution with S total samples. The second sum over $N - i + 1$ development periods in the i th accident period assumes a typical loss triangle with a full lower diagonal of test data. To compare models, we calculated the difference in ELPD for each triangle, t , and its standard error, where the standard error of the difference is the square root of the product of *i*) the sample variance of log predictive density differences between models, and *ii*) the number of data points (Vehtari et al., 2017; Sivula et al., 2020). For the industry data, we combined ELPD values for each triangle by taking the mean of the ELPD differences and mean of the standard errors (Sivula et al., 2020). Approximate 95% confidence intervals were then derived by using a range of 2 standard errors around the estimate. Although the test data is known with certainty, it is still just a portion of accident periods and development lags for each triangle, and therefore uncertainty in ELPD was still calculated.

The RMSE was defined per out-of-sample data point in $\tilde{\mathcal{Y}}$

$$\text{RMSE}_{ij} = \sqrt{\frac{1}{S} \sum_{s=1}^S (\hat{y}_{ij}^{(s)} - \tilde{y}_{ij})^2} \tag{8}$$

where $\hat{y}_{ij}^{(s)}$ is the s th sample from the posterior predictive distribution. In contrast

346 to ELPD, RMSE penalises models that produce predictions further from the test
347 data points using a quadratic scoring rule, and may demonstrate different results
348 depending on the context. As with ELPD, the average differences in RMSE between
349 models per triangle were used to compare models, and their standard errors were
350 derived as the square root of the product of *i*) the sample variance of the differences
351 in RMSE and *ii*) the number of data points.

352 Model calibration was inspected using histograms of the percentiles of the true data
353 on the posterior predictive distributions, i.e. the empirical cumulative distribution
354 functions. Well-calibrated models' percentiles should be uniformly distributed, as in
355 the case of simulation-based calibration above.

356 3 Results

357 To build intuition for the two models, Figure 3 shows a numerical simulation of data
358 from a HMM model, and the results for the HMM model, in panel A, and the two-step
359 model, in panel B. Whereas the HMM model's posterior predictions switch from body
360 to tail depending on the particular latent state path, the two-step models have fixed
361 $\tau = 6$ and $\rho_{1,2} = (6, 10)$. As shown by the ELPD values above each plot, the HMM
362 variant outperforms the two-step approach for all experience periods except the first,
363 where our chosen value τ matches the HMM data-generating process exactly. In other
364 experience periods, the two-step approach generalises poorly. This illustrates that,
365 if τ is chosen correctly, and the tail model is trained on suitable development lags,
366 the two-step approach can provide more exact predictions, because uncertainty in the
367 latent state from the HMM model hurts predictive accuracy. However, if experience
368 periods differ in their body-to-tail switch-over dynamics, which is expected, then
369 overall performance suffers due to growing generalisation error.

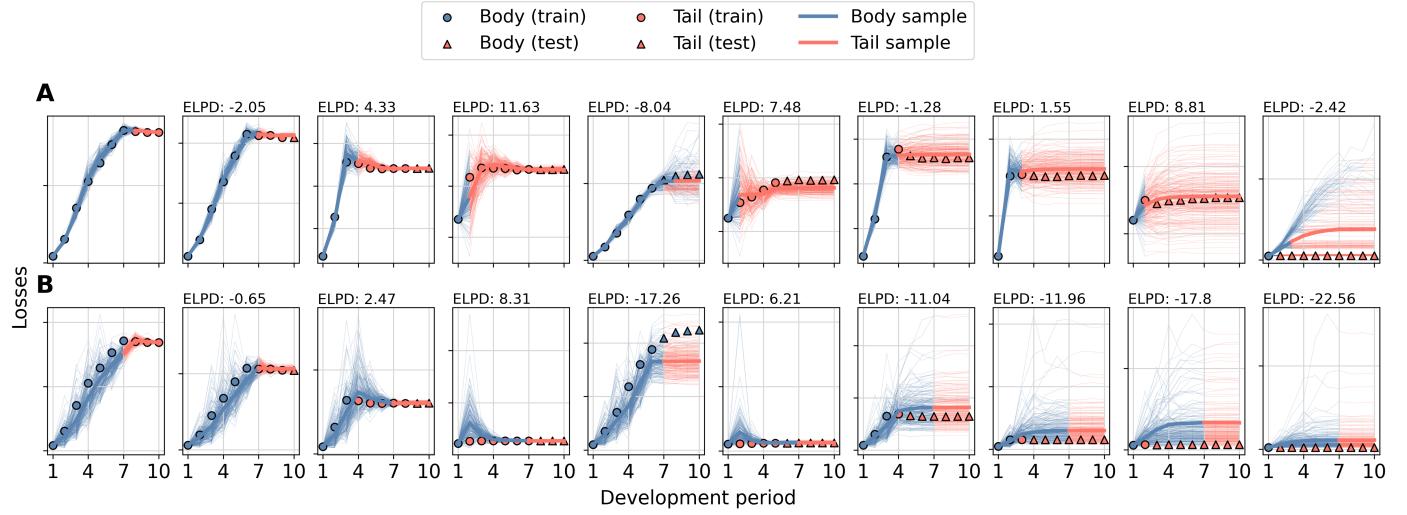


Figure 3: A simulated data example comparing the HMM to two-step process to build intuition. Panels A (HMM) and B (two-step) show 10 experience periods from a single loss triangle simulated from the HMM model. The true losses are shown as points, with colours identifying body (blue) or tail (orange) points. Circles denote training data and triangles test data. Thin lines are samples from the posterior predictive distributions, coloured by latent state, and the thicker line shows the mean paths. The ELPDs on the test data points in each experience period are shown above each plot. The two-step model was fit assuming $\tau = 6$ and $\rho_{1,2} = (6, 10)$.

370 Simulation-based calibration of the HMM model indicated no problems with model
371 calibration (Figure 4), with all histograms matching the assumptions of uniformity.
372 However, 10 models were removed for poor convergence, which typically occurred
373 when the simulated link ratios from the body process were higher than values expected
374 from real loss triangles. Given this occurred rarely, the priors were left unchanged,
375 although suitable prior distributions for Bayesian chain ladder models is an area with
376 a dearth of literature. For the 990 models, the average classification accuracy of the
377 recovered latent state values z , across both training and test data, was 97% with a
378 95% highest density interval (i.e. the 95% most likely values) of [91, 100].

379 The ELPD and RMSE differences between models (compared to the HMM base
380 model) are shown in Figure 5. When calculating ELPD, a small number of pointwise
381 log densities showed very small, negative values, indicating poor predictions on the
382 out-of-sample data leading to numerical instability. We decided to remove any log
383 predictive densities for all models that had values < -100 , which for a single out-
384 of-sample data point was particularly low, given that most ELPD values for a single
385 data point lie within [-5, 5]. This retained 99.91% of values for the PP results, 99.42%
386 of values for WC, 99.64% for CA, and 98.38% for OO. The full log densities are given
387 in the supplementary materials, as well as results for different levels of filtering.

388 For ELPD and RMSE, at least one of the hidden Markov model variants out-performed
389 the two-step approach, apart from ELPD for PP and WC lines of business, where a
390 small proportion of the two standard errors included zero (Figure 8). Overall, the
391 HMM- ν model attained 75% of the best average ELPD scores, and 50% of the best
392 average RMSE scores, meaning that allowing for tail processes to revert to the body
393 is important to making future predictions. Evaluating the predictions at the lag-10
394 out-of-sample values only, to mirror the estimation of ultimate loss, mirrored the re-
395 sults in Figure 5 apart from some minor changes in ranks of the HMM models (full
396 results supplied in the supplementary materials).

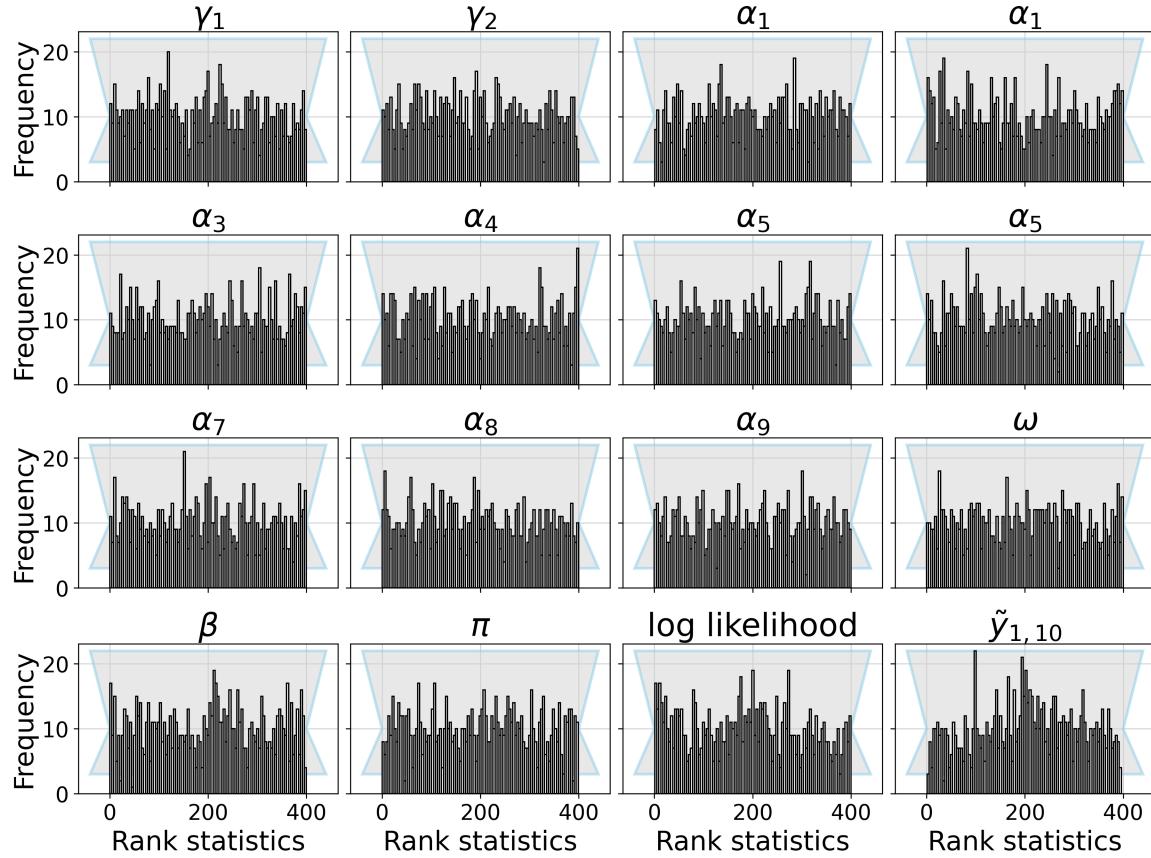


Figure 4: Histograms of simulation-based calibration rank statistics for the HMM model, with the 99% percentile interval from a discrete uniform distribution shown in the grey shaded band. Each histogram shows a key model parameter, and the final two panels show the ranks for the joint log likelihood and the first ultimate loss distribution. For each model, we sampled 4000 draws from the posterior distribution, and thinned the samples by 10 to remove any autocorrelation, meaning a maximum rank statistic of 400. Of the 1000 models, 10 were removed due to poor convergence.

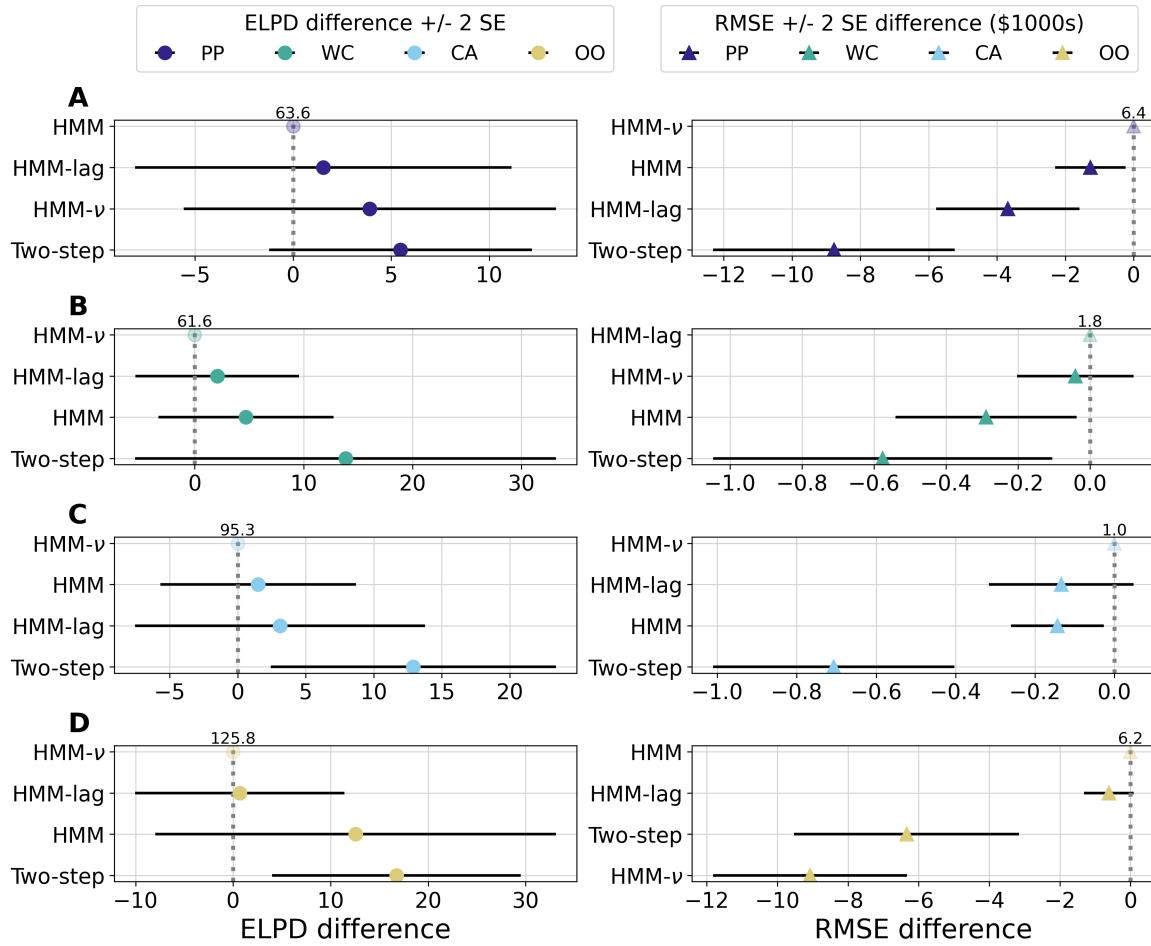


Figure 5: The ELPD (left column) and RMSE (in thousands of dollars; right column) differences (± 2 standard errors; SE) in order of performance for each model and line of business for the industry triangles. Rows A through D enumerate results for line of business separately. The best-performing model is shown at the top of each panel, with the absolute ELPD or RMSE value displayed above. Positive ELPD differences with an uncertainty interval that does not cross zero indicates a credible difference at the 95% level in favour of the top model. Negative RMSE differences with an uncertainty interval that does not cross zero indicates a credible difference in favour of the top model.

397 Model calibration histograms (Figure 6) indicated that both the hidden Markov mod-
398 els and two-step approaches typically have predictions that are too uncertain, indi-
399cated by the inverted-U shaped histograms. In WC, the hidden Markov models had
400 the most uniform percentiles, whereas the two-step approach showed signs of both
401 under- and over-estimation.

402 The hidden Markov model variants had different implications for body-to-tail switch-
403 over points depending on the particular line of business (Figure 7). The PP and CA
404 lines showed, in general, the quickest development to the tail state, at development
405 lag 2 for the HMM and HMM- ν models, and by development lag 3 for most accident
406 periods for the HMM-lag model. In contrast, WC stayed longest in the body state,
407 followed by OO, and both WC and OO lines demonstrated relatively equitable proba-
408bilities of being in the body and tail at later development periods. This is particularly
409 noticeable for the HMM- ν model, where the chance of returning to the body from
410 tail process was allowed.

411 For the five literature triangles, the lack of hold-out data meant that the uncertainties
412 around the ELPD and RMSE differences indicated more equal model performance
413 (Figure 8). The average ELPD and RMSE differences indicated that one of the hidden
414 Markov models often performed better than the two-step approaches. The two-step
415 approach out-performed the hidden Markov models for the [Merz and Wüthrich \(2015\)](#)
416 triangle, and ranked second for RMSE for three of the five triangles. The manual
417 selection of τ for the two-step process often closely aligned with the development lag
418 from the best-performing HMM where the probability of being in the tail was > 0.5 .
419 The one exception was the long-tailed liability triangle of [Balona and Richman \(2022\)](#),
420 where the chosen τ of 12 was much larger than the most likely switch-over lag from
421 the HMM variant of $j = 3$. However, $j = 3$ did closely match the choice of $\rho_1 = 4$.

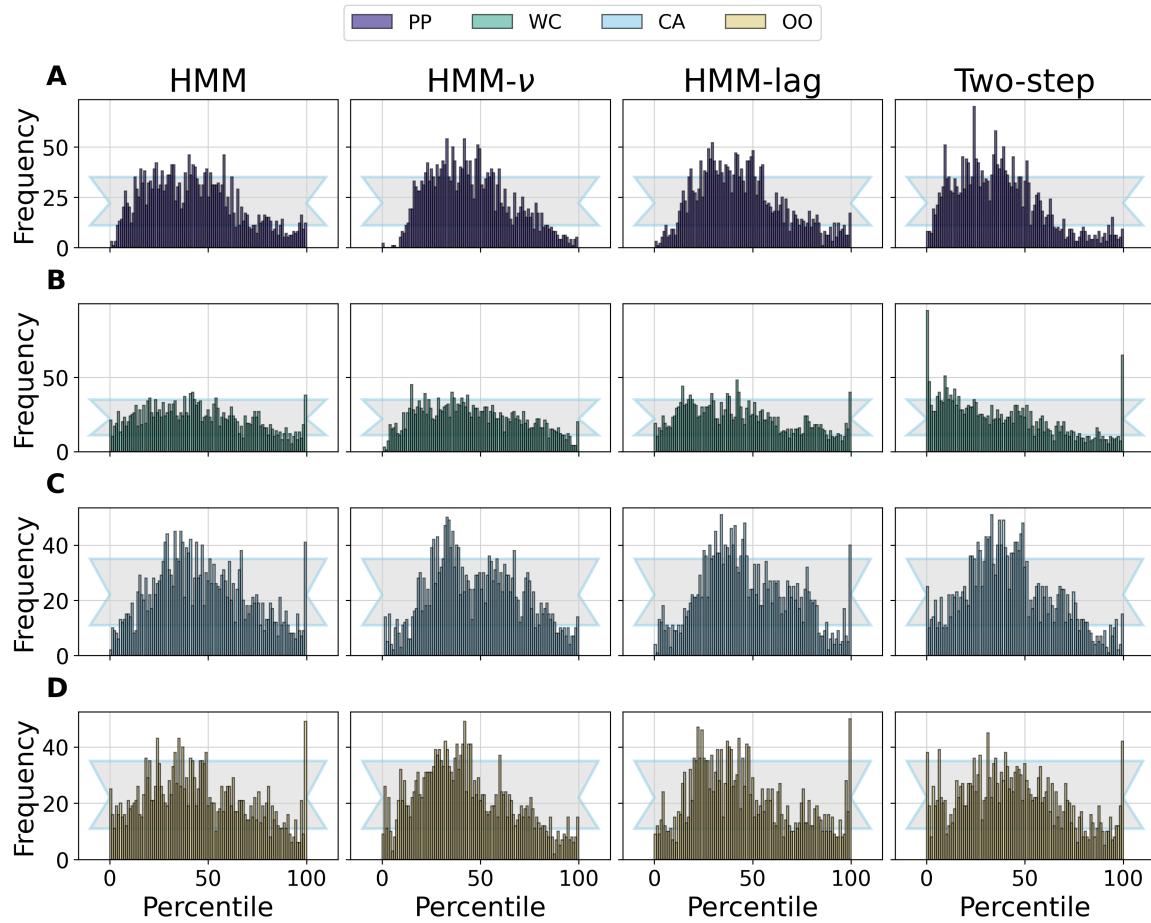


Figure 6: Percentiles of the true left-out values on the posterior distributions for each model and line of business (panels A through D) in the industry triangles. Grey shaded regions provide the 99% intervals of a discrete uniform distribution, for reference. Right-skewed histograms indicate under-estimation, left-skewed histograms indicate over-estimation, and inverted-U histograms indicate predictions that are uncertain.

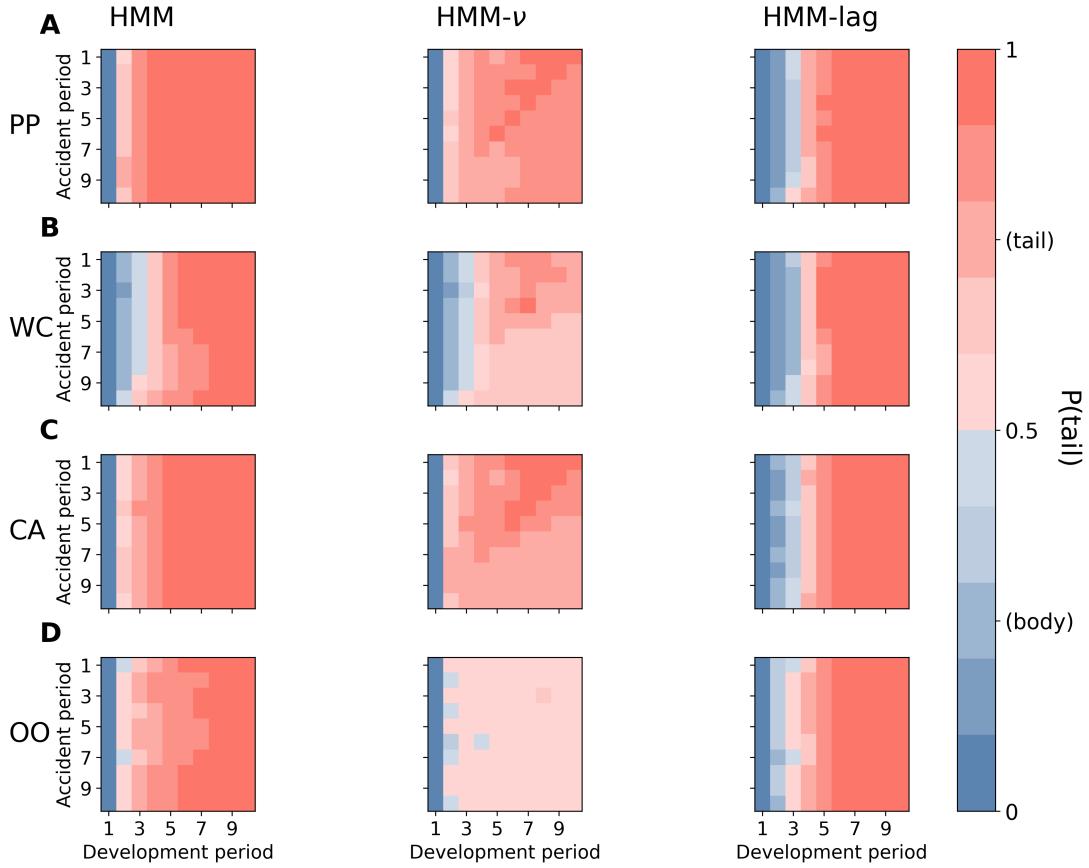


Figure 7: The average probability of being in the tail process for each hidden Markov model parameterization (columns) and line of business (rows A-D) across triangles in the industry data. Probabilities ≤ 0.5 are coloured in blue whereas probabilities > 0.5 are coloured in orange. More faded squares indicating smaller probabilities of being in body and tail processes, respectively.

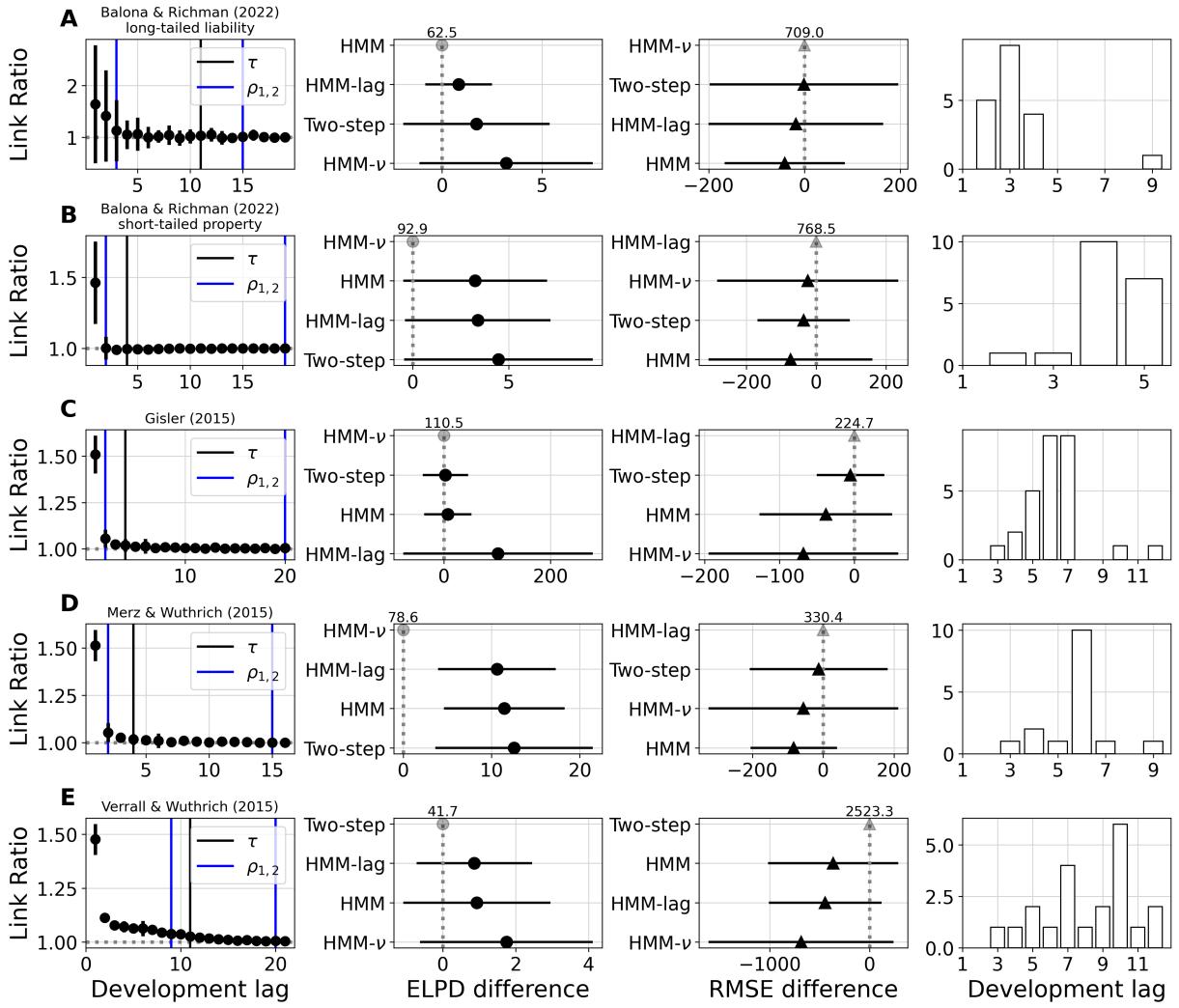


Figure 8: Results for the hidden Markov model and two-step approach on the five literature triangles (rows A - D). The first column shows the mean (± 2 standard deviations) of the empirical link ratios in the triangles. The black and blue vertical lines indicate τ and $\rho = \rho_{1:2}$, the tail start and generalised Bondy model training windows, respectively (see model definitions, above). The second and third columns show the ELPD and RMSE differences (± 2 SE) from the best performing model (top model in each panel), calculated from predictions on the latest diagonal of data (out-of-sample) in the loss triangles. The absolute ELPD and RMSE of the best-performing models are shown above the top model, for reference. The final column shows the development lags where the probability of being in the tail was > 0.5 for the highest ELPD hidden Markov model variant.

422 **4 Discussion**

423 This paper has proposed a hidden Markov loss development model for insurance
424 loss triangles that combines body and tail development models, and automates the
425 selection of body-to-tail switch-over points. Simulation-based calibration validated
426 the hidden Markov model implementation as being unbiased, and across a range
427 of different datasets, the hidden Markov model variants provided similar results to,
428 and often out-performed, the traditional two-step approach. The hidden Markov
429 models' automated detection of body and tail processes more gracefully captures
430 loss development dynamics that may vary over triangle experience periods, as well
431 as reduce analysts' degrees of freedom that make the traditional two-step approach
432 reliant on difficult-to-reproduce and variable subjective decisions.

433 The hidden Markov development model posits a clear data-generating process for
434 loss development dynamics. Although referred to here as ‘body’ and ‘tail’, these
435 two latent states might equivalently be thought of as flexible and smooth periods of
436 loss development, and can interchange depending on the context, as in the HMM- ν
437 model variant here. In this way, the hidden Markov model is not a strictly analogous
438 implementation of the two-step approach, as the two-step approach allows analysts
439 to choose tail model training data windows that overlap the body-to-tail cutoff point.
440 Thus, the same data points may be used in estimation of body and tail processes,
441 rather than a discrete cutoff point between the two. Despite this flexibility, the
442 two-step approach is not a single generative model, and should an analyst choose
443 values for τ and $\rho_{1,2}$ that are not representative of a particular experience period, the
444 predictions from such a model could be extremely biased, as shown in the example of
445 Figure 3. While the hidden Markov model may still make relatively poor predictions
446 for those experience periods with little data (e.g. see the last column of panel A in
447 Figure 3), uncertainty in the true latent state, \mathbf{z} , is more-accurately accounted for.

448 Although most of the hidden Markov model variants performed consistently better
449 on average than the two-step approach on the curated industry datasets of Meyers
450 (2015), the approaches were more similar on the five literature triangles. These two
451 sets of data present different case studies. The industry triangles have been selected
452 to encompass relatively large insurers with mostly stable loss dynamics (see Meyers,
453 2015, appendix A) over a period of 10 years. Due to the number of triangles, the
454 two-step approach's cutoff points, τ and $\rho_{1,2}$, were chosen based on average empirical
455 link ratios, which might not have been the best selection points for some triangles.
456 By contrast, the literature triangles encompass more accident periods per triangle
457 but also smaller books of business (e.g. the medium-sized triangles from Balona and
458 Richman, 2022), and more variability in the tail than present in the industry triangles.
459 Previous papers on loss development models combining body and tail dynamics have
460 not considered the breadth of triangles and lines of business used here. For instance,
461 England and Verrall (2001), Verrall and Wüthrich (2012), and Verrall and Wüthrich
462 (2015) all used a single triangle to illustrate their approaches, and Zhang et al. (2012)
463 used a dataset of 10 workers' compensation triangles and did not consider other lines
464 of business or more volatile triangles. Moreover, the previous papers did not compare
465 their approaches to the more common two-step approach applied in actuarial practice.
466 The datasets used here are provided alongside this article in the repository for ease
467 of access and comparison of other loss development modelling approaches.

468 Both the hidden Markov and two-step approaches demonstrated relatively poor cali-
469 bration on the out-of-sample data from the Meyers (2015) dataset (Figure 6). Primar-
470 ily, the out-of-sample predictions were often too uncertain, producing a predominance
471 of percentiles falling within the central range of possible percentiles. Few articles have
472 shown calibration plots from fully-Bayesian posterior distributions on out-of-sample
473 loss development data, so this pattern of calibration requires further inspection in the
474 literature. Meyers (2015) reports on calibration using the same data set, showing rel-
475 atively well-calibrated predictions. However, we note that Meyers (2015) calculated

476 the percentile of the total ultimate losses in each triangle on a lognormal distribu-
477 tion with mean and variance informed by the total ultimate losses from their models.
478 Thus, these are not directly comparable because the results above average percentiles
479 over each posterior predictive distribution in the test data, not a distribution informed
480 by just the average of posterior predictive distributions.

481 The hidden Markov models presented here could be extended in a number of use-
482 ful ways. Notably, the transition matrix probabilities might be parametric or non-
483 parametric functions of covariates, such as premium volume in each experience period
484 or inflation levels in each calendar period, or include hierarchical effects for experience
485 and development periods. The Bayesian framework, alongside the hidden Markov
486 models implemented here and available in the supplementary material, make these
487 extensions highly accessible. Additionally, the hidden Markov model framework is
488 general enough to include any body or tail model, not just the chain ladder and gen-
489 eralised Bondy forms. For instance, there are a number of inverse power curves to use
490 for tail modelling ([CAS Tail Factor Working Party, 2013; Evans, 2015; Clark, 2017](#)),
491 and extensions and variations on the chain ladder model have been commonplace
492 ([England and Verrall, 2002](#)). Although the analyses in this paper focused on paid
493 losses, the same models could be applied to estimates of reported losses (i.e. paid loss
494 plus estimates of reserve), or joint modelling of both paid and reported losses (e.g.
495 see [Zhang, 2010](#), for one approach).

496 5 Competing interests

497 The author declares no competing interests.

⁴⁹⁸ **References**

- ⁴⁹⁹ Al-Mudafer, M. T., Avanzi, B., Taylor, G., and Wong, B. (2022). Stochastic loss
500 reserving with mixture density neural networks. *Insurance: Mathematics and Eco-*
501 *nomics*, 105:144–174.
- ⁵⁰² Balona, C. and Richman, R. (2022). The actuary and IBNR techniques: a machine
503 learning approach. *Variance*.
- ⁵⁰⁴ Barnett, G. and Zehnwirth, B. (2000). Best estimates for reserves. In *Proceedings of*
505 *the Casualty Actuarial Society*, volume 87, pages 245–321.
- ⁵⁰⁶ Beard, R. (1960). Three R's of insurance: risk, retention and reinsurance. *Journal of*
507 *the Staple Inn Actuarial Society*, 15(6):399–421.
- ⁵⁰⁸ Bornhuetter, R. L. and Ferguson, R. E. (1972). The actuary and IBNR. In *Proceedings*
509 *of the casualty actuarial society*, volume 59, pages 181–195.
- ⁵¹⁰ Carlin, B. P. and Chib, S. (1995). Bayesian model choice via Markov chain Monte
511 Carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Method-*
512 *ology*, 57(3):473–484.
- ⁵¹³ Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M.,
514 Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2017). Stan: a probabilistic
515 programming language. *Journal of statistical software*, 76.
- ⁵¹⁶ CAS Tail Factor Working Party (2013). The estimation of loss development tail
517 factors: a summary report.
- ⁵¹⁸ Clark, D. (2017). Estimation of inverse power parameters via GLM. *Actuarial Review*,
519 *May-June 2017*.
- ⁵²⁰ Clarke, T. and Harland, N. (1974). A practical statistical method of estimating claims
521 liability and claims cash flow. *ASTIN Bulletin: The Journal of the IAA*, 8(1):26–37.

- 522 De Alba, E. (2002). Bayesian estimation of outstanding claim reserves. *North Amer-*
523 *ican Actuarial Journal*, 6(4):1–20.
- 524 England, P. D. and Verrall, R. J. (2001). A flexible framework for stochastic claims
525 reserving. In *Proceedings of the Casualty Actuarial Society*, volume 88, pages 1–38.
- 526 England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general
527 insurance. *British Actuarial Journal*, 8(3):443–518.
- 528 England, P. D., Verrall, R. J., and Wüthrich, M. V. (2019). On the lifetime and
529 one-year views of reserve risk, with application to IFRS 17 and Solvency II risk
530 margins. *Insurance: Mathematics and Economics*, 85:74–88.
- 531 Evans, J. (2015). A continuous version of Sherman’s inverse power curve model with
532 simple cumulative development factor formulas. In *Casualty Actuarial Society E-*
533 *Forum, Fall 2014-Volume*.
- 534 Friedland, J. (2010). Estimating unpaid claims using basic techniques. In *Casualty*
535 *actuarial society*, volume 201.
- 536 Fröhlich, A. and Weng, A. (2018). Parameter uncertainty and reserve risk under
537 Solvency II. *Insurance: Mathematics and Economics*, 81:130–141.
- 538 Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y.,
539 Kennedy, L., Gabry, J., Bürkner, P.-C., and Modrák, M. (2020). Bayesian workflow.
540 *arXiv preprint arXiv:2011.01808*.
- 541 Gesmann, M. and Morris, J. (2020). Hierarchical compartmental reserving models.
542 In *Casualty Actuarial Society*, page 4.
- 543 Gisler, A. (2009). The insurance risk in the SST and in Solvency II: modelling and
544 parameter estimation. *Available at SSRN 2704364*.

- 545 Hesselager, O. (1994). A Markov model for loss reserving. *ASTIN Bulletin: The*
546 *Journal of the IAA*, 24(2):183–193.
- 547 Kunce, J. and Chatterjee, S. (2017). A machine-learning approach to parameter
548 estimation. *Virginia: CAS*.
- 549 Kuo, K. (2019). Deeptriangle: a deep learning approach to loss reserving. *Risks*,
550 7(3):97.
- 551 Lally, N. and Hartman, B. (2018). Estimating loss reserves using hierarchical Bayesian
552 Gaussian process regression with input warping. *Insurance: Mathematics and Eco-*
553 *nomics*, 82:124–140.
- 554 Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T. A., Watanabe, Y. Y.,
555 Murgatroyd, M., and Papastamatiou, Y. P. (2017). Analysis of animal accelerome-
556 ter data using hidden Markov models. *Methods in Ecology and Evolution*, 8(2):161–
557 173.
- 558 Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder
559 reserve estimates. *ASTIN Bulletin: The Journal of the IAA*, 23(2):213–225.
- 560 Mack, T. (1994). Which stochastic model is underlying the chain ladder method?
561 *Insurance: mathematics and economics*, 15(2-3):133–138.
- 562 Merz, M. and Wüthrich, M. V. (2015). Claims run-off uncertainty: the full picture.
563 *Swiss Finance Institute Research Paper*, (14-69).
- 564 Meyers, G. (2015). Stochastic loss reserving using Bayesian MCMC models. Casualty
565 Actuarial Society Arlington, VA.
- 566 Modrák, M., Moon, A. H., Kim, S., Bürkner, P., Huurre, N., Faltejsková, K., Gelman,
567 A., and Vehtari, A. (2023). Simulation-based calibration checking for Bayesian
568 computation: the choice of test quantities shapes sensitivity. *Bayesian Analysis*,
569 1(1):1–28.

- 570 Munroe, D., Zehnwirth, B., and Goldenberg, I. (2018). Solvency capital requirement
571 and the claims development result. *British Actuarial Journal*, 23:e15.
- 572 Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications
573 in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.
- 574 Scurfield, H. (1968). Motor insurance statistics. *Journal of the Staple Inn Actuarial
575 Society*, 18(3):207–236.
- 576 Sherman, R. E. (1984). Extrapolating, smoothing and interpolating development
577 factors. In *Proceedings of the Casualty Actuarial Society*, volume 71, pages 122–
578 155.
- 579 Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology:
580 undisclosed flexibility in data collection and analysis allows presenting anything as
581 significant. *Psychological science*, 22(11):1359–1366.
- 582 Sivula, T., Magnusson, M., Matamoros, A. A., and Vehtari, A. (2020). Uncertainty
583 in Bayesian leave-one-out cross-validation based model comparison. *arXiv preprint
584 arXiv:2008.10296*.
- 585 Stan Development Team (2024a). **CmdStan**: the command line interface to stan.
- 586 Stan Development Team (2024b). **CmdStanPy**: the python interface to cmdstan.
- 587 Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2018). Validat-
588 ing Bayesian inference algorithms with simulation-based calibration. *arXiv preprint
589 arXiv:1804.06788*.
- 590 Taylor, G., McGuire, G., and Greenfield, A. (2003). Loss reserving: past, present and
591 future. *University of Melbourne Centre of Actuarial Studies Research Paper*, (109).
- 592 Taylor, G. C. (1977). Separation of inflation and other effects from the distribution
593 of non-life insurance claim delays. *ASTIN Bulletin: The Journal of the IAA*, 9(1–
594 2):219–230.

- 595 Taylor, G. C. and Ashe, F. R. (1983). Second moments of estimates of outstanding
596 claims. *Journal of Econometrics*, 23(1):37–61.
- 597 Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation
598 using leave-one-out cross-validation and waic. *Statistics and computing*, 27:1413–
599 1432.
- 600 Verrall, R. J. and Wüthrich, M. V. (2012). Reversible jump Markov chain monte carlo
601 method for parameter reduction in claims reserving. *North American Actuarial
602 Journal*, 16(2):240–259.
- 603 Verrall, R. J. and Wüthrich, M. V. (2015). Parameter reduction in log-normal chain-
604 ladder models. *European Actuarial Journal*, 5:355–380.
- 605 Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insur-
606 ance*. John Wiley & Sons.
- 607 Zhang, Y. (2010). A general multivariate chain ladder model. *Insurance: Mathematics
608 and Economics*, 46(3):588–599.
- 609 Zhang, Y., Dukic, V., and Guszcza, J. (2012). A Bayesian non-linear model for
610 forecasting insurance loss payments. *Journal of the Royal Statistical Society Series
611 A: Statistics in Society*, 175(2):637–656.