Hierarchical Compartmental Models for Loss Reserving

Jake Morris, FIA

Abstract

Motivation. This paper proposes a triangle-based stochastic reserving framework for parsimoniously describing insurance claims generation, reporting and settlement processes with intuitive parameters.

Method. Deterministic compartmental models are explored as extensible tools to describe and project the insurance claims process using a small number of parameters, including a measure of case reserve robustness. A Schedule-P reserving case study illustrates the application of a nonlinear hierarchical ("mixed-effects") framework to fit compartmental models to outstanding and cumulative paid claims development triangles, simultaneously. This allows one or more of the claims process parameters to vary by claims cohort in accordance with a statistical distribution. An optional Bayesian implementation facilitates the robust incorporation of external information and judgment into the projection of reserve uncertainty.

Results. A flexible stochastic reserving framework is established, with benefits including the ability to explicitly account for reporting and/or settlement rate changes, make inferences about components of the claims process and scenario test future process changes using information gathered across the business.

Conclusions. Hierarchical compartmental models can describe and project the insurance claims process in an optional level of detail for the purpose of setting reserves.

Availability. Frequentist model R code is contained in Appendix E, Bayesian model OpenBUGS code is contained in Appendix F and an illustration spreadsheet is available at: http://www.casact.org/pubs/forum/16sforum/.

Keywords. Stochastic loss reserving, compartmental reserving models, claims process modeling, hierarchical models, nonlinear mixed-effects, Bayesian modeling, MCMC.

1. INTRODUCTION

A variety of triangle-based stochastic reserving techniques have been proposed for estimating future general insurance claims payments, ranging from generalized linear models (England and Verrall, 2002) to nonlinear hierarchical models (Guszcza, 2008). Methods incorporating both paid and incurred information have been explored (Martínez-Miranda, Nielsen and Verrall, 2012; Quarg and Mack, 2004), which provide richer inference and improved interpretability. Furthermore, Bayesian methods (Zhang, Dukic and Guszcza, 2012; Meyers, 2007; England and Verrall, 2005; Verrall, 2004) have become increasingly ubiquitous; providing flexibility and the ability to robustly incorporate judgment into uncertainty projections.

This paper explores a new triangle-based (and optionally-Bayesian) stochastic reserving *framework* which considers the relationship between exposure, case reserves and paid claims. By doing so, it enables practitioners to build communicable models that are consistent with their understanding of the insurance claims process. Furthermore, it supports the identification and quantification of claims process characteristics to provide tangible business insights.

1.1 Research Context

Compartment(al) models (Sheppard, 1948) are extensible tools for describing the transfer of material between components of a system over time. For a sufficient volume of claims, the insurance claims process can be represented by a small number of compartments and intuitive parameters. The parameters describe aggregate claims movements between compartments and the ultimate loss ratio (ULR), decomposed into a reported loss ratio and a measure of case reserve robustness.

Motivated by Guszcza (2008), a nonlinear hierarchical modeling framework is proposed for fitting compartmental loss reserving models to claims triangles, allowing one or more of the model parameters (and hence development patterns) to vary by claims cohort in accordance with a statistical distribution. This enables flexible and parsimonious compartmental models to be fitted to reported outstanding claims and cumulative paid claims development triangles, simultaneously.

An optional Bayesian implementation (akin to Zhang, Dukic and Guszcza, 2012) allows external information and judgment to be incorporated into reserve uncertainty projections. Additionally, Markov chain Monte Carlo (MCMC) techniques facilitate model flexibility, and consequently, specific features such as the correlation between successive observations and calendar shocks can be accounted for.

1.2 Objective

Hierarchical compartmental reserving models have parallels with the hierarchical growth curves put forward by Guszcza (2008). In contrast to monotonic growth curves however, compartmental models can be fitted to cumulative paid claims *and* outstanding claims reserves, simultaneously. Since outstanding claims typically rise and fall over time, negative incurred claims development is supported. Furthermore, explicit modeling of outstanding claims may reduce the subjectivity inherent in the selection of a growth curve for tail extrapolation. Finally, relating compartmental model parameters back to the claims process provides intuitive control over the level of model complexity.

In contrast to Zhang, Dukic and Guszcza (2012), the corresponding Bayesian implementation enables prior beliefs to be more readily incorporated into process-based model parameters. This allows drivers of uncertainty to be isolated. Additionally, Bayesian hierarchical compartmental models have the flexibility to handle negative development for reserve uncertainty projections contrary to many existing GLM-type methods (England and Verrall, 2002).

Furthermore, compared to existing methods that utilize both paid and incurred data (e.g. Martínez-Miranda, Nielsen and Verrall, 2012; Quarg and Mack, 2004), a compartmental approach ensures consistency between estimated paid and incurred claims.

Although this paper proposes a triangle-based approach, methods incorporating individual claims data (e.g. Antonio and Plat, 2014; Parodi, 2013) exhibit a number of desirable properties, including the ability to reflect underlying claims processes. Such methods typically require a combination of models to be parameterized however, whereas a compartmental framework allows claims process characteristics to be quantified using a single structural model. Additionally, hierarchical model diagnostic tests can help to mitigate the risk of overfitting the data and reducing extrapolation validity.

1.3 Outline

The remainder of the paper proceeds as follows:

- Section 2 will introduce compartmental modeling theory, hierarchical compartmental models and Bayesian hierarchical compartmental models.
- Section 3 will define a compartmental model for the claims process. Parameter interpretations will be discussed and a number of practical extensions will be explored.
- Section 4 will contain a triangle reserving case study detailing the application of frequentist and Bayesian hierarchical compartmental models to a Schedule-P dataset.
- Section 5 will present a brief overview of future development areas.
- Section 6 will summarize the paper's findings.

Appendices will contain various supplementary materials including the case study data, frequentist modeling R code, and Bayesian modeling OpenBUGS code.

2. COMPARTMENTAL MODELS

A system is said to be a compartment(al) system when its entities can be grouped into a finite number of connected homogeneous components, known as compartments (Sheppard, 1948). They are often used to describe how entities/materials change location or state over time. The set of all possible compartments in a system is called the *state-space*, and the phenomena under study in each compartment are described by *state-variables* (Blomhøj, Kjeldsen and Ottesen, 2014).

Compartmental models can be deterministic or stochastic, containing discrete or continuous state-variables in discrete or continuous time. In deterministic models, the behavior of the quantities within the system is dictated solely by their past behavior and the rules that govern the model. In contrast, stochastic models imply a distribution of possible behaviors (Brauer, 2008). A useful feature of compartmental models is that complexity can be controlled by adjusting the number of compartments and/or their corresponding inflows and outflows.

The focus of this paper will be a practical claims reserving application of **deterministic, continuous state-variable and continuous-time compartmental models**. The rationale is as follows:

- Compartmental models describing exposure, reported outstanding claims and cumulative paid claims (where the latter two are simultaneously fitted) have not yet been introduced into the loss reserving literature.
- Deterministic models are practical to implement, and their simplicity results in clear and communicable claims process parameters.
- The hierarchical framework proposed in Section 4 increases mathematical complexity to the extent that at present, appropriate hierarchical stochastic compartmental reserving models are not easily implementable in conventional software.

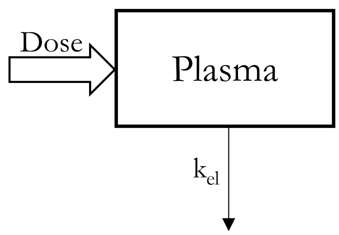
Sections 2.1 and 2.2 will contain overviews of deterministic and stochastic compartmental models, and Section 2.3 will introduce hierarchical compartmental models.

2.1 Deterministic compartmental models

Deterministic compartmental models have many possible applications. One of which is to describe the transport of material through biological systems, where compartments have physiological interpretations. For example, "compartmental pharmacokinetic models" are commonly used to describe the continuous transfer of an administered drug into, within and out of a patient. State-spaces typically comprise blood plasma and body tissues/organs, with state-variables denoting their drug concentration-time (or amount-time) profiles.

Deterministic, continuous-time models can be expressed as linear systems of ordinary differential equations (ODEs), with state-variables expressed as differentials of time. Analytical state-variable solutions are linear combinations of exponential terms describing the estimated amounts of material in each compartment at each time.

A one compartment pharmacokinetic model with state-space $\{Plasma(t)\}$ for a direct intravenous drug dose can be written schematically as follows:



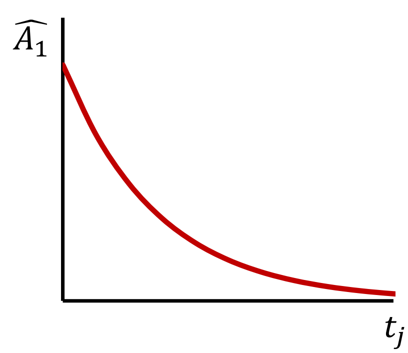
Alternatively, the model can be written as a single ODE, where the state-variable $\{A_1(t)\}$ denotes the amount of drug in the blood plasma at time t, and the positive "rate elimination constant" $\{k_{el}\}$ describes how quickly the drug is eliminated from the body. It is assumed that elimination of the drug is constant and directly proportional to its amount (first-order kinetics):

$$\begin{gathered} dA_1/dt = -k_{el}A_1 \\ A_1(0) = Dose \end{gathered} \tag{2.1}$$

A patient's blood plasma amount-time profile $A_1(t)$ can be measured by repeatedly sampling their blood over the time following a drug dose. The rate parameter k_{el} can then be estimated by solving the ODE and fitting the model to the patient's amount-time observations. Denoting y_j as the jth drug amount measurement for a patient, we can specify a nonlinear regression (Seber and Wild, 1989) as

$$\begin{gathered} A_1(t_j) = y_j = Dose \cdot e^{-k_{el}t_j} + \varepsilon_j \\ \varepsilon_j \sim N(0, \sigma^2) \end{gathered} \tag{2.2}$$

where σ^2 is the variance of the discrepancy between the model fit and the drug amount measurements. For illustrative purposes, an estimated blood plasma amount-time profile $\widehat{A_1}(t_j)$ for a given dose and rate of elimination is as follows:



2.2 Stochastic compartmental models

In contrast to deterministic compartmental models, stochastic compartmental models introduce uncertainty external to the history of the modeled process by assuming that one or more of the states are random variables. This may be achieved, for example, by adding probabilistic state transfer mechanisms to an existing deterministic structure (Rescigno and Segre, 1966).

Three example forms of stochastic compartmental models and their corresponding properties are:

1) *Discrete-time Markov chain models:* discrete state-variables, discrete time steps
2) *Continuous-time Markov chain models:* discrete state-variables, continuous time scale
3) *Stochastic differential equation (SDE) models:* continuous state-variables and time scale

Hachemeister (1980) provides a loss reserving application of discrete-time Markov chain models. Analogously, Orr (2007), Hasselager (1994) and Norberg (1993) provide loss reserving applications of continuous-time Markov chain models.

2.3 Hierarchical compartmental models

Section 2.1 describes how a deterministic compartmental model can be used to estimate a drug amount-time profile for a single patient. However, in practice drug developers wish to make inferences about a population of individuals that may eventually take a particular drug. Assuming a drug has been administered to a group of individuals and expressing y_{ij} as the jth drug amount measurement ($j = 1 \; to \; n_i$) for the ith individual ($i = 1 \; to \; M$), we could use nonlinear regression to **fit a separate compartmental model to each individual:**

$$y_{ij} = Dose_i \cdot e^{-k_{el_i} t} + \varepsilon_{ij} \qquad \textbf{(2.3)}$$

However, this modeling approach may result in many parameters relative to the number of data points available for modeling, reducing the credibility of each estimated parameter.

An alternative approach is to pool all individuals' concentration measurements and **fit one compartmental model with a single parameter to all individuals combined:**

$$y_{ij} = Dose_i \cdot e^{-k_{el} t} + \varepsilon_{ij} \qquad \textbf{(2.4)}$$

Although k_{el} is likely to be estimated with greater precision than each k_{el_i} in Eq. (2.3), it is unlikely to result in an accurate fit to each individual due to between-patient variability e.g. differing metabolisms.

The approach commonly used in pharmacokinetic modeling in addition to other life and social sciences is **nonlinear hierarchical modeling**, which has previously been advocated for loss reserving by Guszcza (2008). **Hierarchical** (or *mixed-effects*) **models allow some model parameters to be fixed across individuals and some to vary by individual**. More generally, they allow parameters to vary by any natural data grouping. For example, for estimating insurance claims reserves Guszcza (2008) proposes claims cohorts (individual accident years) as a grouping for cumulative paid claims.

A hierarchical framework allows model parameters to vary by the assumed data grouping in accordance with a statistical distribution defined by a mean and variance only. This reduces the number of estimable parameters compared to the first modeling approach outlined above. Conversely, because the modeler can select which parameters vary by individual, each individual can be described in greater detail compared to the second modeling approach outlined above.

Hierarchical/mixed-effects models are said to allow data-sparse individuals to "borrow strength" from data-rich individuals. For parameters that vary by individual, an individual's parameter estimate is a weighted average of:

1) The estimated average parameter value across all individuals; and
2) The estimated individual parameter value for the individual.

The weight given to the individual parameter value is proportional to the individual's data volume. To illustrate how nonlinear hierarchical models are structured, Eq. (2.3) can be rewritten as

$$y_{ij} = Dose_i \cdot e^{-(\overline{k_{el}}+(k_{el_i}-\overline{k_{el}}))\cdot t} + \varepsilon_{ij} \tag{2.5}$$

where $\overline{k_{el}}$ represents the average rate of elimination across all individuals. Denoting $\overline{k_{el}}$ as β, and $k_{el_i} - \overline{k_{el}}$ as b_i (Pinheiro and Bates, 2000), this becomes

$$\begin{gathered} y_{ij} = Dose_i \cdot e^{-(\beta+b_i)\cdot t} + \varepsilon_{ij} \\ \varepsilon_{ij} \sim N(0, \sigma^2), \qquad b_i \sim N(0, \psi^2) \end{gathered} \tag{2.6}$$

where β is referred to as a *fixed-effect* and b_i as a *random-effect*, which has its own probability sub-model. A shared distribution for the random-effects induces a correlation between individuals, which may be an appropriate assumption if they are assumed to come from a wider population. σ^2 represents the within-subject variability, whereas ψ^2 represents the between-subject variability. For any number of individuals being modeled, only three parameters (β, σ and ψ) need to be estimated.

Two key reasons for using a hierarchical framework are **parsimony** and **flexibility**. These features may be useful for loss reserving where data are incomplete and sometimes limited for modeling purposes, requiring descriptive models that do not overfit.

Antonio and Zhang (2014) provide a detailed exploration of nonlinear hierarchical models for insurance data.

2.3.1 Bayesian hierarchical compartmental models

A modeler may want to incorporate external information and/or judgment into a compartmental model to account for information not contained within the modeled dataset. For example, in drug development there may be other data-rich drug administration studies from which to base parameter prior distributions. For the hierarchical model outlined in Eq. (2.6) it could be assumed that

$$\log(\beta) \sim N(\overline{\beta}, \gamma^2) \tag{2.7}$$

where β denotes the fixed-effect for the rate of drug elimination, and $\overline{\beta}, \gamma^2$ denote the prior mean and variance of $\log(\beta)$ respectively, which are specified by the modeler rather than estimated. Bayes' rule can then be used to estimate the posterior distribution of the fixed-effect as

$$\begin{gathered} p(\beta | y_{ij}) \propto p(\beta)\, p(y_{ij} | \beta) \\ \equiv p(\beta | y_{ij}) \propto p(\beta)\, \mathcal{L}(\beta; y_{ij}) \\ posterior \ \propto prior \times likelihood \end{gathered} \tag{2.8}$$

where $p(\cdot)$ is a probability density function, β is the "random" parameter for which we wish to make inferences, y_{ij} is the "fixed" jth observation for individual i and $\mathcal{L}(\cdot)$ is the likelihood function. The posterior distribution is a *credibility weighting* of the prior distribution and likelihood function, where the weight placed on prior beliefs is inversely proportional to the volume of modeled data.

As highlighted by Zhang, Dukic and Guszcza (2012), this approach can be useful for loss reserving where it is often essential for a practitioner to **incorporate judgment** into reserve projections to allow for information not contained within the modeled data. Additionally, Bayesian methods allow us to **quantify reserve uncertainty** consistently with the definition stated by the 2005 Casualty Actuarial Society Working Party on Quantifying Variability in Reserve Estimates:

'Given . . . our current state of knowledge, what is the probability that [the entity's] final payments will be no larger than the given value'.

This can be framed mathematically using Bayesian statistics. Denoting ULR_i as the ultimate loss ratio (and parameter of interest) for the ith claims cohort and $Incurred_{ij}$ as the jth cumulative incurred claims observation for the ith claims cohort, the posterior density of ULR_i *given* $Incurred_{ij}$ is

$$p(ULR_i | Incurred_{ij}) \propto p(ULR_i)\, \mathcal{L}(ULR_i; Incurred_{ij}) \tag{2.9}$$

which provides an estimate of ULR parameter uncertainty. It is straightforward to incorporate process uncertainty into this posterior, from which a distribution for final payments can be derived consistently with the above definition. Finally, Bayesian models **increase flexibility** because they require only that model parameters and the relationships between them are specified.

A detailed exposition of Bayesian methods and their applications is given by Gelman *et al.* (2013).

3. COMPARTMENTAL MODELS FOR LOSS RESERVING

To specify a deterministic, continuous state-variable and continuous-time compartmental model for the insurance claims process, a state-space must be defined. The selection of a possible state-space is illustrated by considering the insurance claims process over development time for a *cohort* of claims e.g. an accident year:

1) Once a group of insurance policies have been written and incepted, they are exposed to the risk of making claims. Therefore an initial **"Exposed to Risk"** state is defined.
2) For some proportion of exposed policies, claim events will occur and be reported to the insurer. Claims are typically case reserved and classed as being outstanding until settled, defining a second state: **"Claims Outstanding"**.
3) A proportion of all reported outstanding claims will be settled with a payment amount, which defines a **"Claims Paid"** state.

The state-space is therefore $\{$Exposed to Risk(t), Claims Outstanding(t), Claims Paid$(t)\}$. The states-variables in turn denote the amount of remaining exposure, the monetary amount of claims outstanding, and the cumulative monetary amount of claims paid at development time t.

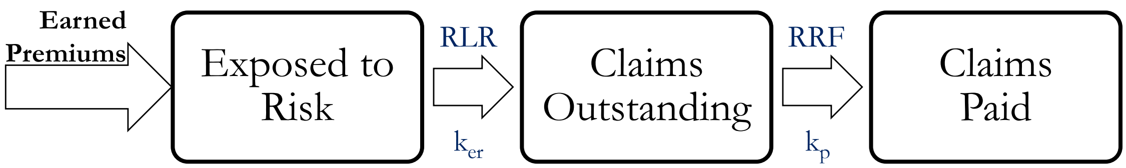
Assuming that the above process is a forward process only, i.e. no material re-opening of paid claims, a model schematic can be written as follows:

Exposure reduces over time as groups of claims are reported and become paid at some proportion of their outstanding amounts. This reduces the claim amounts outstanding (eventually to 0 as t becomes large) and ensures consistency between paid and incurred claims estimates. Although this model is for claim amounts, an adapted version could be defined for claim numbers (not shown).

To initialize the process, a suitable measure of exposure must be chosen as an input variable. For an accident year/quarter cohort of claims, earned premiums could be used (Guszcza, 2008; Clark, 2003). Alternatively, a pure exposure measure could be chosen in line with the original pricing basis (see Section 3.2).

Independently of the chosen exposure measure, the timing of policies coming on-risk during the claims cohort should be considered. If policies coming on-risk during an accident year/quarter are largely replaced by similar policies coming off-risk, i.e. steady-state conditions, a practitioner could input all exposures into to the system at development time 0. This is the approach taken for the case study presented in Section 4. Similarly, it would be acceptable to input all exposures at time 0 if all (or a large proportion of) policies that could give rise to a claim in the cohort are on-risk at the start of the cohort (e.g. accident quarter). However, if exposure materially fluctuates during the lifetime of the cohort, a more sophisticated approach is required to match the input exposures with the cohort's development times at which the policies come on-risk (see Section 3.2).

For an accident year/quarter cohort of claims, the use of (ultimate) earned premiums as an exposure measure provides an appealing parameter set. A schematic for what will be termed the "**baseline model**" is shown below, followed by its corresponding parameter definitions:



- **Reported Loss Ratio** ("RLR"): the *proportion* of premiums that become reported claims.
- **Rate of earning and reporting** ("k_{er}"): the *rate* at which claim events occur and are subsequently reported to the insurer.
- **Reserve Robustness Factor** ("RRF"): the *proportion* of outstanding claims that eventually become paid by the insurer.
- **Rate of payment** ("k_p"): the *rate* at which outstanding claims are paid by the insurer.

Therefore this model is defined in terms of proportions $\{RLR, RRF\}$ and rates $\{k_{er}, k_p\}$. For the continuous-time assumption to hold, a sufficient number of policies must be written to give rise to a steady "flow" of reported and paid claims over development time.

Denoting the state-space {Exposed to Risk(t), Claims Outstanding(t), Claims Paid(t)} as $\{EX(t), OS(t), PD(t)\}$, the above model can be written as follows:

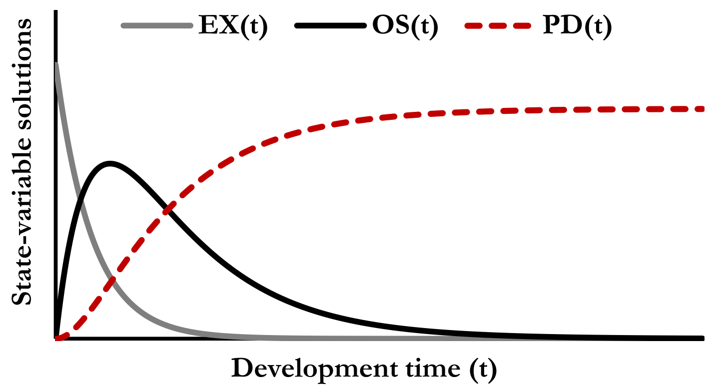
$$
\begin{aligned}
dEX/dt &= -k_{er} \cdot EX \\
dOS/dt &= k_{er} \cdot RLR \cdot EX - k_p \cdot OS \\
dPD/dt &= k_p \cdot RRF \cdot OS
\end{aligned}
\qquad \textbf{(3.1)}
$$

Compartment initial conditions are $EX(0) = earned\ premiums = P$, $OS(0) = 0$ and $PD(0) = 0$, assuming steady-state exposure. Each parameter is assumed to be constant over development time t; however, this assumption is relaxed in Section 3.2.

Analytical state-variable solutions can be obtained using Laplace transforms (Gustav, 1974):

$$EX(t) = Pe^{-k_{er}t}$$
$$OS(t) = \frac{P \cdot RLR \cdot k_{er}}{k_{er} - k_p} \cdot \left(e^{-k_p t} - e^{-k_{er}t}\right) \qquad \textbf{(3.2)}$$
$$PD(t) = \frac{P \cdot RLR \cdot RRF}{k_{er} - k_p} \cdot \left(k_{er} \cdot \left(1 - e^{-k_p t}\right) - k_p \cdot \left(1 - e^{-k_{er}t}\right)\right)$$

The paid claims solution is analogous to a "growth curve", as put forward for loss reserving by Clark (2003). For a given set of parameters, the state-variables in the above system can be visualized over development time t as follows:



Although exposure may be impractical to track over time, outstanding and cumulative paid claims are typically observable at specific development time points, albeit incomplete for reserving purposes. Nonlinear regression techniques can be used to fit Eq. (3.2) to outstanding and cumulative paid claims simultaneously to derive parameter estimates and project the claims process to ultimate.

3.1 Parameter interpretations

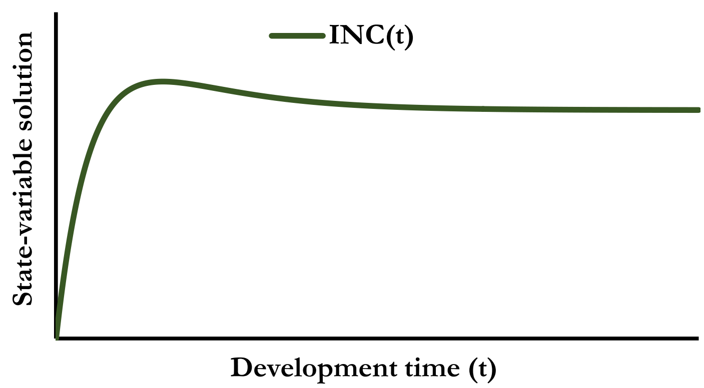
The two rate parameters k_{er} ($k_{er} > 0$) and k_p ($k_p > 0$) determine the monetary value of remaining exposures reported and outstanding claims paid respectively, per infinitesimal unit of time (with units t^{-1}). The term "k_{er}" is used to reflect that a policy exposed to risk must have a claim **e**vent occur before it is **r**eported, and may also be termed a rate of reporting (from exposure). It follows that higher magnitude rate parameters imply faster claims reporting/payment. However, if the model were to contain these rate parameters alone then all exposure would eventually convert to paid claims, resulting in a ULR equal to 100% (for a premium-based exposure measure).

To allow a range of possible ultimate loss ratios it is necessary to specify at least one proportion parameter, similar to Clark (2003). The two proportion parameters RLR ($RLR > 0$) and RRF ($RRF > 0$) determine the percentage of exposures that become reported claims and the percentage of outstanding claims that become paid claims respectively. The RRF parameter therefore indicates the average level of case over- or under-reserving for a cohort of claims. If case handlers persistently under-reserve, this would imply an RRF of over 100% and vice versa. An RRF of 100% indicates perfect case reserving on average amongst a cohort of claims. However, this may be the result of some claims being heavily over-reserved and some claims being heavily under-reserved, cancelling each other out in aggregate.

Persistent over-reserving is often associated with an incurred development pattern that rises and falls. Claims incurred at development time t $\{INC(t)\}$ can be derived under the model as

$$INC(t) = OS(t) + PD(t) \tag{3.3}$$

and visualized (for an RRF less than 100%):



Thus the model is able to capture negative incurred increments. Under the model, estimated cumulative incurred and paid claims tail development is defined by the extrapolation of estimated outstanding claims to zero (driven by k_p), and the estimated RRF. A convenient result is that the estimated ultimate loss ratio (ULR) can be directly obtained as

$$ULR = RLR \cdot RRF \tag{3.4}$$

The reason for this can be observed by equating the paid loss ratio (PLR) at development time t to the product of the RLR and RRF parameter definitions for a homogeneous group of claims, which are reported and subsequently paid together, i.e.

$$\frac{PD(t)}{P} = \frac{OS(t-v)}{P} \cdot \frac{PD(t)}{OS(t-v)} \tag{3.5}$$

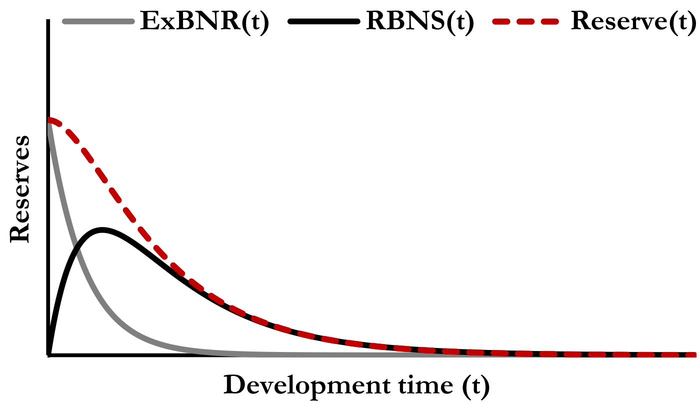
where t denotes development time within the claims cohort and v represents the elapsed time between the homogeneous group of claims being reported outstanding and paid. It is assumed that the premiums (P) for their underlying policies are written before time $t-v$. It follows that the RLR numerator and RRF denominator of the right hand side cancel out, and the PLR converges to the ULR for sufficiently large t.

3.1.1 ExBNR vs. RBNS

Using the compartmental model above it is possible to derive an exposed but not reported ("ExBNR") reserve and reported but not settled ("RBNS") reserve at development time t. The term "ExBNR" reflects the loss of claim event timing information once claims are grouped into outstanding and paid claims cohorts, and contains incurred but not reported ("IBNR") *plus* unearned claims:

$$\begin{aligned} ExBNR(t) &= EX(t) \cdot RLR \cdot RRF \\ RBNS(t) &= OS(t) \cdot RRF \\ Reserve(t) &= ExBNR(t) + RBNS(t) \end{aligned} \tag{3.6}$$

The reserves contain "IBNER" (incurred but not enough reported), indicated by the appearance of the reserve robustness factor (RRF). They can be visualized over development time for a given set of parameters as follows:



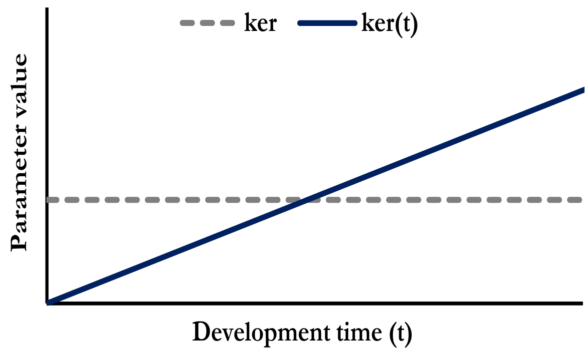
Taking the definition of IBNR to be ultimate losses less incurred losses to date, Eq. (3.4) can be used to define ultimate losses as $P \cdot RLR \cdot RRF$, and hence $IBNR(t) = P \cdot RLR \cdot RRF - INC(t)$. When $EX(0) = P$, Eq. (3.6) provides an alternative derivation: $IBNR(t) = Reserve(t) - OS(t)$.

3.2 Model extensions

Compartmental models are extensible, allowing practitioners to adjust them in line with their understanding of the claims process for the class of business being modeled. Matching model extensions to underlying processes may also enable models to be more easily communicated to stakeholders.

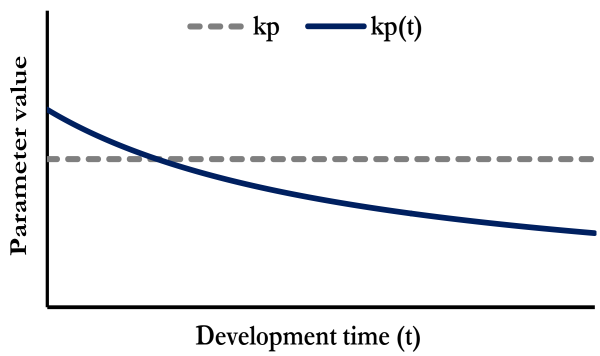
Parameters within the model have thus far been assumed to be constant and independent of development time. However, it may be desirable for one or more of the model parameters to depend on development time. For example, allowing the rate of reporting k_{er} to increase with development time t could capture delays between claim events and reports:

$$k_{er}(t) = \beta_{er} \cdot t \tag{3.7}$$



Alternatively, liability claims outstanding in later development periods may be those in/awaiting litigation. To reflect a potentially slower rate of settlement and subsequent payment for such claims, a nonlinear rate of payment $k_p(t)$ could be specified as follows:

$$k_p(t) = \frac{\beta_{p,1}}{\beta_{p,2} + t} \tag{3.8}$$



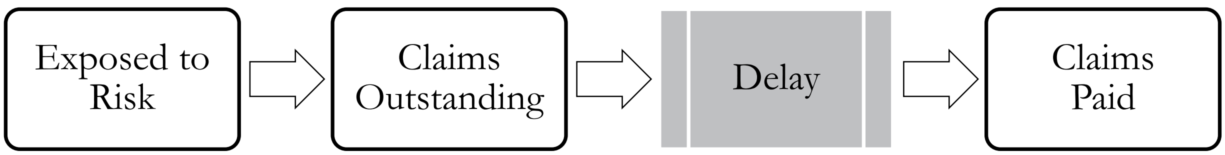
This function assumes that rate of payment reductions *decrease* over development time. Substituting the above rate functions into Eq. (3.1) gives

$$\begin{aligned} dEX/dt &= -\beta_{er} \cdot t \cdot EX \\ dOS/dt &= \beta_{er} \cdot t \cdot RLR \cdot EX - \{\beta_{p,1}/(\beta_{p,2} + t)\} \cdot OS \\ dPD/dt &= \{\beta_{p,1}/(\beta_{p,2} + t)\} \cdot RRF \cdot OS \end{aligned} \tag{3.9}$$

Implied development patterns for the baseline model and extended model incorporating the above functions are outlined in Appendix A.

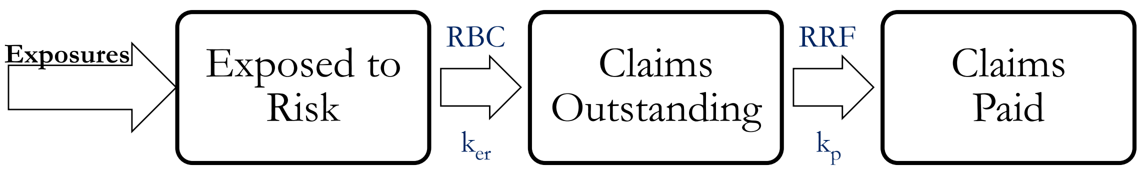
The proportion parameters could also be expressed as functions. For example, case reserves may be less robust at later development times for claims facing uncertain litigation. In practice, there are numerous plausible functions for describing how the claims process parameters are observed or perceived to behave; however, these will not be explored further in this paper.

It is also possible to increase the number of compartments to reflect claims sub-processes. For example, a bodily injury claims sub-class may exhibit a marked delay between claims being reported and subsequently being settled while damages are being quantified. This could be modeled using a "delay" state as follows:



Other potentially modelable sub-processes include calendar shocks (Section 4.2.2), reopened claims, third party claims payment recoveries, reinsurance recoveries, latent claims etc. However, available data may limit the degree to which complexity can be increased.

As noted in Section 3, it may be appropriate to initialize a compartmental reserving model with a non-premium measure of exposure. In which case the baseline schematic can be rewritten as follows:



The parameter interpretations for this model are largely unchanged; however, the reported loss ratio is replaced by a reported burning cost:

- **Reported Burning Cost** ("RBC"): the *proportion* of exposures that become reported claims.

The ultimate burning cost (UBC) can be obtained from the RBC and RRF parameters (analogously to the ULR in Eq. (3.4)) as

$$UBC = RBC \cdot RRF \tag{3.10}$$

This parameterization could be useful for pricing. The anticipated exposure for a cohort of new business could be multiplied by a selected UBC (allowing for changes in underwriting, claims environment, reserve robustness etc.) to derive an estimated loss cost. This may form the risk premium or be a precursor to a full frequency-severity analysis, for example.

Finally, compartmental reserving models can be generalized to describe exposure accumulation for cases where steady-state conditions do not hold (see Section 3). This can be achieved by continuously inputting portions of premium/exposure into the system over a period of time.

For example, if claims are grouped into an underwriting cohort then ultimate premiums can be projected to derive a discrete incremental writing pattern. For a cohort's premium and claims data observed at *discrete* development times $r\Delta, r \in \{0,1,2, \dots\}$ after the commencement of the underwriting cohort, $PPN[r]$ can be defined as the proportion of ultimate premiums written uniformly over the period $r\Delta \rightarrow r\Delta + \Delta$. It follows that $\sum_0^\infty PPN[r] = 1$. The input to the exposure compartment (denoted by $\overrightarrow{EX}$) over each *continuous* time increment $t \rightarrow t + \delta t$ (where δ is infinitesimally small) can then be set to

$$\overrightarrow{EX}(t \rightarrow t + \delta t) = \ ultimate\ premiums \cdot \delta t \cdot PPN\left(\left\lfloor \frac{t}{\Delta} \right\rfloor\right) \qquad \textbf{(3.11)}$$

where $\lfloor \cdot \rfloor$ denotes the floor or "next smallest integer value". If there is substantial time between policies being written and subsequently incepting (i.e. bound but not incepted "BBNI" policies), then the aforementioned writing pattern could be replaced by an inception pattern.

3.3 Limitations

As discussed in Section 3, all exposures can be input to the compartmental reserving system at time 0 under steady-state conditions. However, if steady-state conditions do not hold and material exposure fluctuations are not taken into account (e.g. using the approach outlined above), these will be absorbed into the reporting rate parameter k_{er}. This could lead to misleading k_{er} comparisons if the model is fitted to multiple claims cohorts, and additionally, may result in poor model fits.

Equation (3.5) illustrates a key assumption of deterministic compartmental reserving models: at a given time, claims within each compartment are assumed to be well-mixed and homogeneous i.e. they are assumed to behave uniformly and in accordance with a single set of parameters. In reality, each *individual* claim is likely to have a distinct RLR, k_{er}, RRF and k_p from every other claim. However, for an aggregated cohort of claims values it is only necessary for the *average* behavior of the cohort to be in line with the model parameters at each time, which may be a reasonable assumption for a high volume of claims within a particular claim size range.

A limitation of this approach is that a cohort with many heterogeneous individual claims (e.g. low-frequency high-value claims) or erratic case reserve fluctuations may not be well reflected by a deterministic compartmental model. To model a heterogeneous cohort, one could cap claims values within the cohort at a specified threshold and apply a frequency-severity or alternative approach for losses above the threshold. Other data segmentation techniques may be appropriate or, alternatively, the differing behavior of individual claims may be more accurately reflected by a stochastic or semi-stochastic compartmental model, as outlined in Appendix B.

A practical limitation is that some claims cohorts will have limited development histories, preventing a credible deterministic compartmental reserving model from being fitted due to a high ratio of parameters relative to data points. This limitation is addressed in Section 4.

3.4 Illustration

A spreadsheet containing a parameter-adjustable discretized compartmental reserving model is available at: http://www.casact.org/pubs/forum/16sforum/. This illustrates the dynamics of how the amounts in each compartment are determined over time for both constant and non-constant rate parameters. Additionally, it allows both steady-state and accumulating exposure.

4. HIERARCHICAL COMPARTMENTAL RESERVING CASE STUDY

The preceding Sections explore a deterministic compartmental reserving model for a single cohort of claims (e.g. an accident year). However, reserves are typically set for several cohorts of claims, often grouped into triangles. Cohorts are likely to have some shared characteristics; for example, due to the same underwriters and claims handling philosophy. However, they are also likely to exhibit differences; for example, due to changes in underlying risk profiles and differing claims environments.

The nonlinear hierarchical approach outlined in Section 2.3 allows for individual claims cohort characteristics when the data are credible, while allowing less mature cohorts to borrow strength from more mature cohorts. This can help to achieve parsimony. Following Guszcza (2008), triangles are viewed as "longitudinal" datasets, where claims cohorts are individuals and cumulative losses at various development times are a series of observations for each individual.

Frequentist and Bayesian hierarchical compartmental models will be fitted to a sample loss reserving dataset obtained at: http://www.casact.org/research/index.cfm?fa=loss_reserves_data. The selected workers' compensation dataset comprises both outstanding and cumulative paid claims development data grouped by accident years 1988-1997 and development years 1-10, together with earned premiums by accident year. The dataset contains both upper triangles (calendar years 1988-1997) and lower triangles of data (calendar years 1998-2006). The upper triangles and earned premiums as at 12/31/1997 are as follows:

Outstanding Claims ($'000s)

AY	Prem	1	2	3	4	5	6	7	8	9	10
1988	104	53	41	32	25	17	13	10	7	2	1
1989	89	54	37	27	20	14	10	7	3	2	
1990	86	55	37	28	18	11	7	4	3		
1991	99	61	42	26	15	9	6	4			
1992	105	66	46	31	22	12	8				
1993	119	68	51	40	22	17					
1994	111	62	47	32	24						
1995	78	57	49	35							
1996	64	57	42								
1997	48	41									

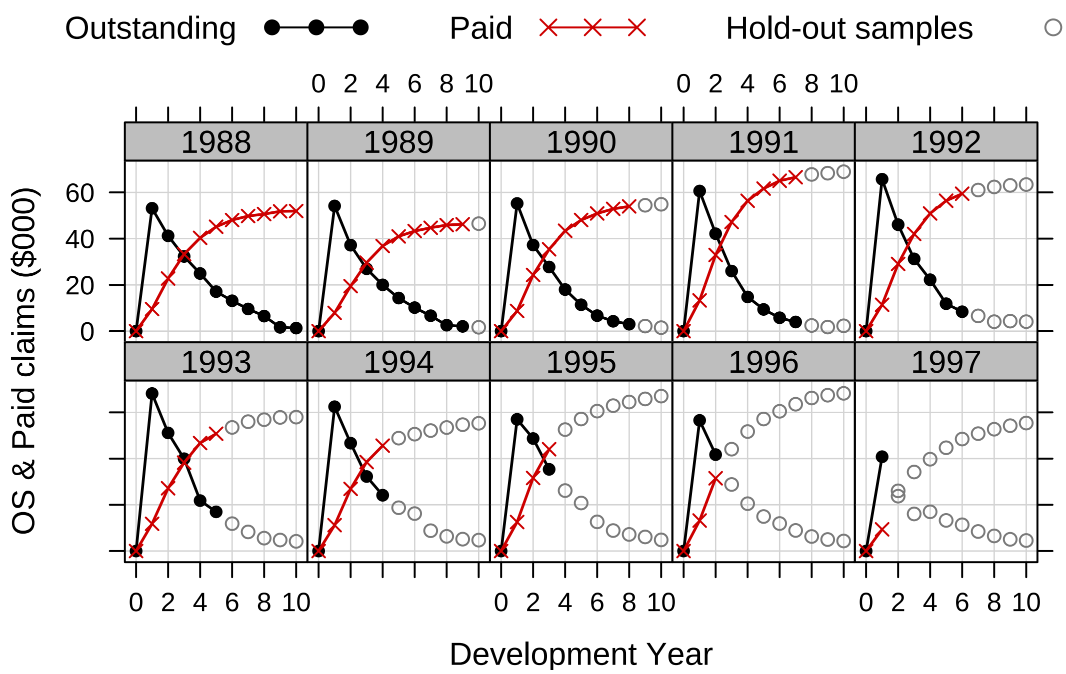
Cumulative Paid Claims ($'000s)

AY	Prem	1	2	3	4	5	6	7	8	9	10
1988	104	10	23	33	40	45	48	50	51	52	52
1989	89	8	19	30	37	41	43	45	46	46	
1990	86	9	24	35	43	48	51	53	54		
1991	99	13	33	47	56	62	65	67			
1992	105	11	29	42	51	56	59				
1993	119	12	27	38	47	51					
1994	111	11	27	38	46						
1995	78	13	32	44							
1996	64	13	31								
1997	48	9									

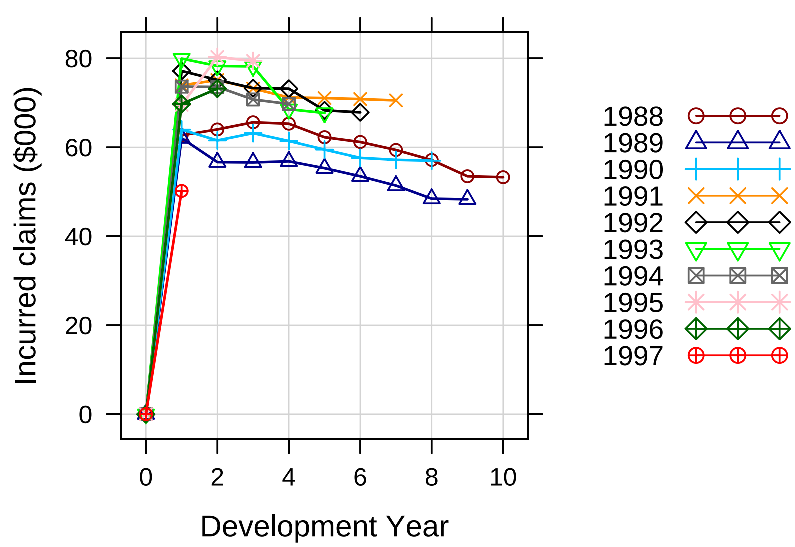
Claims development data should initially be visualized by accident year to establish whether:

1) A compartmental model is appropriate i.e. whether there is a detectable process; and
2) There are any claims process characteristics that can be identified from the outset.

The below plots suggest a clear process between claims being reported and subsequently paid, therefore a compartmental model may be appropriate. The data are also relatively stable, suggesting that the baseline compartmental model outlined in Section 3 is an appropriate starting point.



Incurred development has a clear downwards trend typical of over-stated case reserves at some point during the development history, i.e. $RRF < 1$:



In Section 4.1 a frequentist hierarchical compartmental model will be fitted, assessed for goodness of fit and improved as necessary. A Bayesian implementation will be explored in Section 4.2. For both exercises, model predictability will be tested against the lower triangle hold-out samples.

4.1 Frequentist modeling

The motivations for exploring frequentist hierarchical compartmental models (and point estimates) before their Bayesian counterparts are as follows:

- Best estimate reserves are of principle stakeholder interest, followed by reserve uncertainty;
- Fewer modeling assumptions are required and thus model building is less time consuming; and
- Model run times are relatively quick, allowing models to be tested, interpreted and improved upon relatively quickly.

The baseline compartmental model ODEs (Section 3) are

$$\begin{aligned} dEX/dt &= -k_{er} \cdot EX \\ dOS/dt &= k_{er} \cdot RLR \cdot EX - k_p \cdot OS \\ dPD/dt &= k_p \cdot RRF \cdot OS \end{aligned} \tag{4.1}$$

with initial conditions $EX(0) = earned\ premiums = P$, $OS(0) = 0$ and $PD(0) = 0$ (assuming steady-state exposure − see Section 3). To ensure that compartmental model parameter estimates are positive, we reparameterize using the logarithm of the parameters $\{lk_{er} = \log(k_{er}), lRLR = \log(RLR), lk_p = \log(k_p), lRRF = \log(RRF)\}$ to give an initial "structural" model:

$$\begin{aligned} dEX/dt &= -\exp(lk_{er}) \cdot EX \\ dOS/dt &= \exp(lk_{er}) \cdot \exp(lRLR) \cdot EX - \exp(lk_p) \cdot OS \\ dPD/dt &= \exp(lk_p) \cdot \exp(lRRF) \cdot OS \end{aligned} \tag{4.2}$$

This model can be specified in a format compatible with the R software (R Core Team, 2016) package "nlmeODE" (Tornøe *et al.*, 2004a) and combined with a grouped data object (see Appendices D and E). The data comprise upper triangles of outstanding and cumulative paid claims together with compartment initial conditions (earned premiums) by accident year, as outlined above.

To fit a hierarchical compartmental model based on the above ODEs, it must be decided which of the model parameters should have random-effects and therefore vary by accident year. For this case study, the components of the ultimate loss ratio (the reported loss ratio and reserve robustness factor) will be assumed to vary by accident year to define a baseline hierarchical model.

The Eq. (4.2) outstanding and cumulative paid claims state-variable solutions for accident year $i = 1 \text{ to } 10$ and development year $j = 1 \text{ to } 11 - i$ can be denoted $f_{OS}(P_i, \boldsymbol{\phi}_i, t_j)$ and $f_{PD}(P_i, \boldsymbol{\phi}_i, t_j)$ respectively, where P_i is the earned premium for accident year i. Stacking response variables for outstanding claims OS_{ij} and cumulative paid claims PD_{ij} into a single response variable $\boldsymbol{y}_{ij} = (OS_{ij}, PD_{ij})^T$ enables a nonlinear hierarchical "statistical" model to be specified (Model 1):

$$\boldsymbol{y}_{ij} = \boldsymbol{f}(P_i, \boldsymbol{\phi}_i, t_j) + \boldsymbol{\varepsilon}_{ij}$$

$$\boldsymbol{y}_{ij} = \begin{bmatrix} OS_{ij} \\ PD_{ij} \end{bmatrix}, \quad \boldsymbol{f}(P_i, \boldsymbol{\phi}_i, t_j) = \begin{bmatrix} f_{OS}(P_i, \boldsymbol{\phi}_i, t_j) \\ f_{PD}(P_i, \boldsymbol{\phi}_i, t_j) \end{bmatrix}, \quad \boldsymbol{\varepsilon}_{ij} = \begin{bmatrix} \varepsilon_{ij}{}^{OS} \\ \varepsilon_{ij}{}^{PD} \end{bmatrix}$$

$$\boldsymbol{\phi}_i = \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \\ \phi_{4i} \end{bmatrix} = \begin{bmatrix} lk_{er} \\ lRLR_i \\ lk_p \\ lRRF_i \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} 0 \\ b_{1i} \\ 0 \\ b_{2i} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b}_i \tag{4.3}$$

$$\boldsymbol{b}_i \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_1^2 & 0 \\ 0 & \psi_2^2 \end{bmatrix}\right), \qquad \boldsymbol{\varepsilon}_{ij} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & \lambda^2 \end{bmatrix}\right)$$

The fixed-effects $\boldsymbol{\beta}$ represent the mean values of the *logarithm* of the claims process parameters across a theoretical "population" of accident years, and the random-effects $\boldsymbol{b}_i$ represent the deviations of the individual accident year parameters $\boldsymbol{\phi}_i$ from their mean values. The random-effects are assumed to be independent for different accident years and the within-accident-year errors $\boldsymbol{\varepsilon}_{ij}$ are assumed to be independent for different (i, j), and independent of the random-effects (Pinheiro and Bates, 2000). The variance of random-effect $b_{qi} \in \boldsymbol{b}_i$ is denoted $\psi_q{}^2$. The within-accident-year variances are denoted σ^2 and $\lambda^2\sigma^2$ for outstanding and cumulative paid claims respectively.

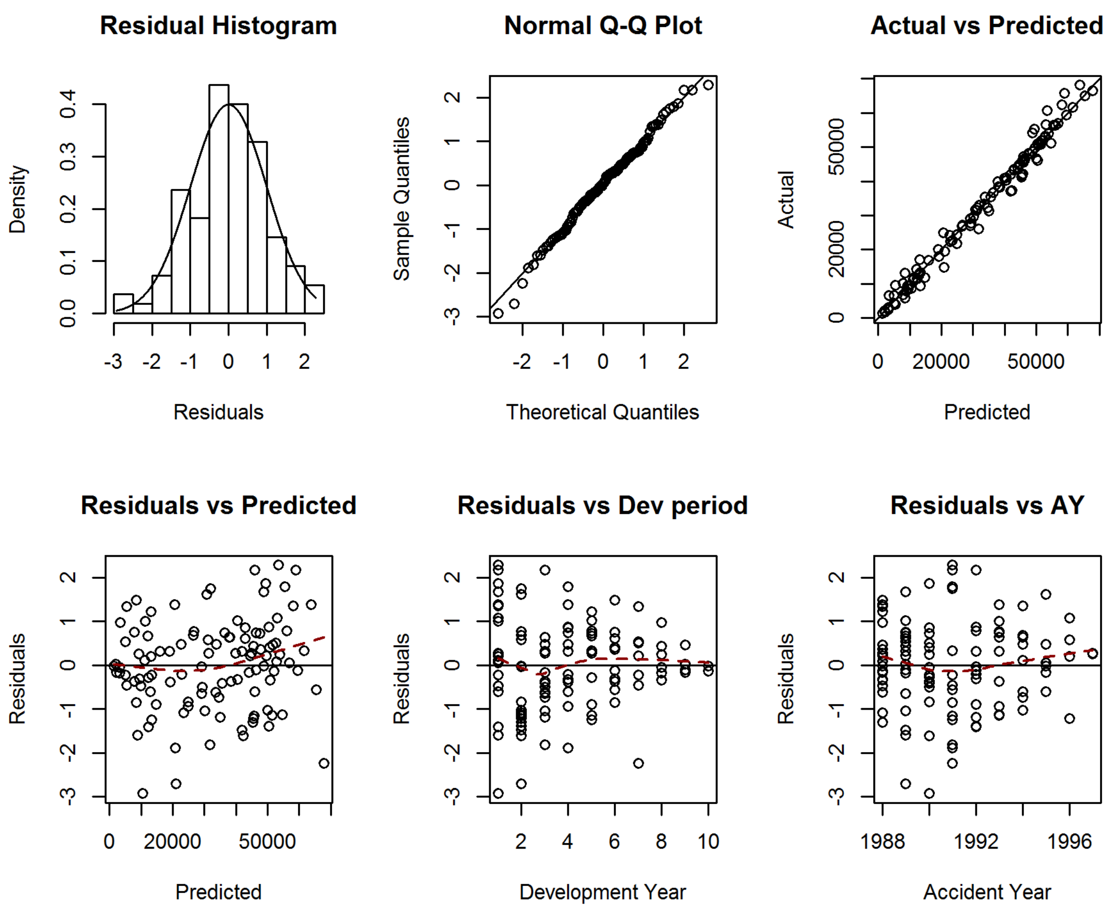
Initial fixed-effect parameter estimates are required to begin optimization, which could be obtained using a self-starting algorithm (Appendix C) or selected judgmentally as follows:

- Development year 1 outstanding claims are observed to be a high proportion of earned premiums. Therefore the reported loss ratio initial value has been selected as 100%, i.e. all premiums are assumed to convert to reported claims.
- The early outstanding loss peaks indicate a fast rate of reporting, so an initial value of 1.5 has been selected. This results in a value of claims reported in the first development year equal to $(1 - e^{-1.5}) \cdot P \cdot RLR = 78\% \cdot P \cdot RLR$.
- The downwards incurred development trend indicates large case reserve redundancies ($RRF < 1$), therefore a value of 0.75 has been selected.
- The rate of payment is observed to be slower than the rate of reporting, justifying a selected initial value equal to half the rate of reporting (0.75).

The above model can be combined with the previously outlined ODE system and fitted to the outstanding and cumulative paid triangles concurrently using the R package "nlme" (Pinherio *et al.*, 2016). Convergence is achieved in seconds. Appendix E contains the R model code and output.

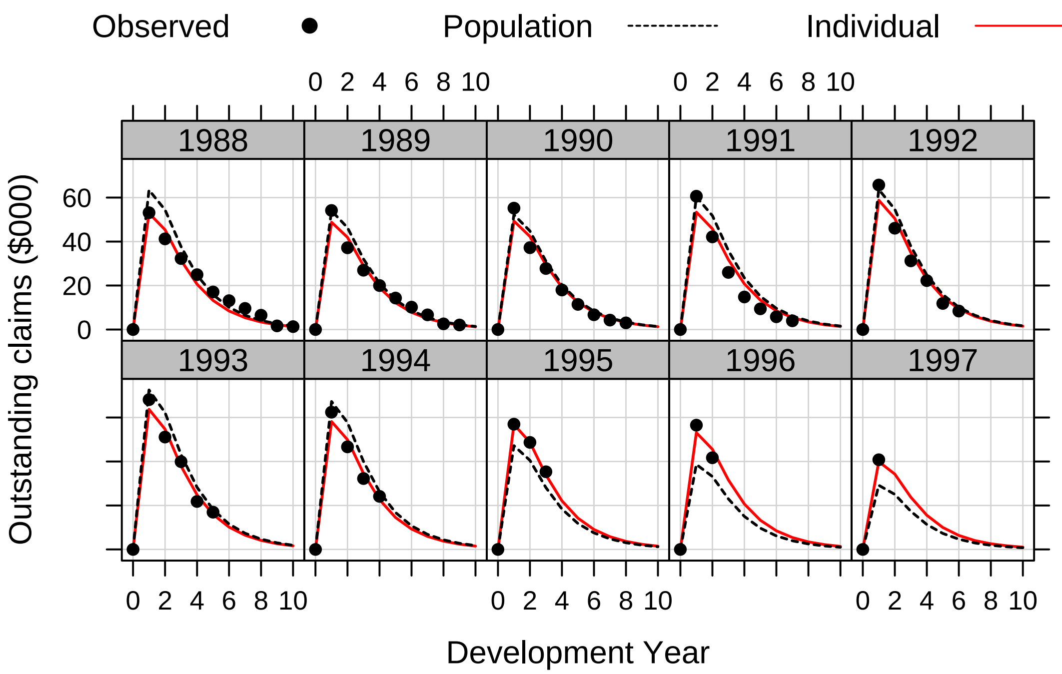
The estimated random-effect standard deviations ($\hat{\psi}_q$) relative to the fixed-effects ($\hat{\beta}_p \in \widehat{\boldsymbol{\beta}}$) for $lRLR_i$ ($\hat{\psi}_1 = 0.19$; $\hat{\beta}_2 = 0.03$) and $lRRF_i$ ($\hat{\psi}_2 = 0.13$; $\hat{\beta}_4 = -0.41$) indicate significant variation by accident year, justifying the inclusion of the random-effects. The within-accident-year error standard deviation for paid claims fits is estimated to be $\hat{\lambda} = 18\%$ of the within-accident-year error standard deviation for outstanding claims fits, which seems reasonable since paid claims development is comparatively stable.

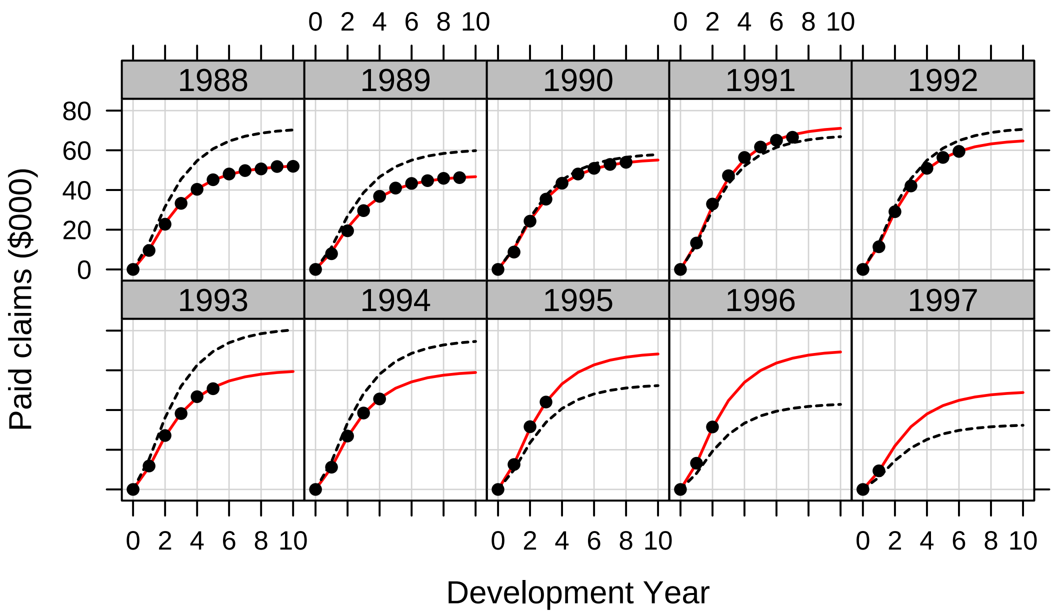
A set of diagnostic plots can be inspected to verify modeling assumptions and assess model fit:



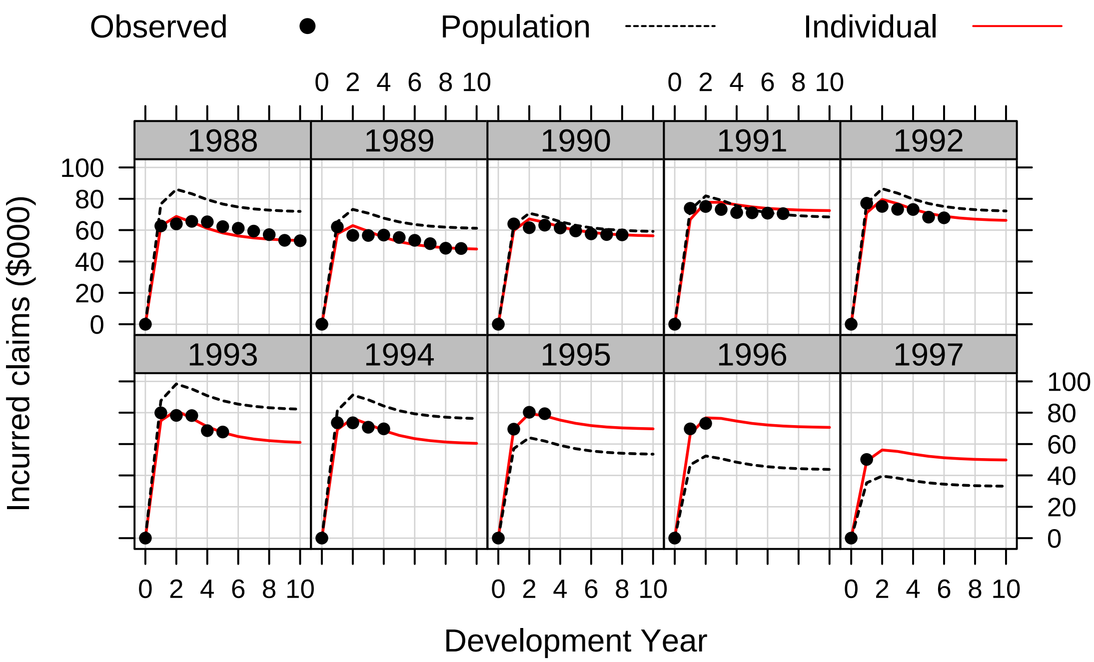
The upper left two plots indicate that the standardized residuals are approximately normal for this model, and the "Actual vs. Predicted" plot shows that the model fits the data reasonably well for most of the data range. However, some higher valued observations are under-predicted by the model, and the "Residuals vs. Predicted" plot highlights this. The remaining residual plots mostly lie between [-2, 2] and overlaid LOESS smoothers (Cleveland, 1979) suggest they are absent of trends.

To assess how well this model describes each accident year, we can plot the observed development data by accident year (circles) and superimpose the individual model fits (solid lines). To highlight the between-accident-year variability, the population-level fits (based on the fixed-effects and replicating a pooled model fit – see Section 2.3) are also shown (dashed lines):





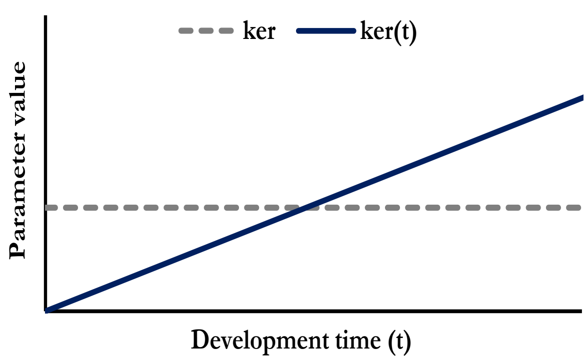
The population fits demonstrate that the model would not accurately describe claims development if parameters were fixed across accident years. The individual cumulative paid claims fits are reasonable, but the outstanding claims fits systematically under-predict the peak observations. It appears that claims are modeled to be reported over a longer time period than the data suggests.



4.1.1 Development time-dependent reporting rate

To attempt to improve the fits, we can adjust the structural model. Selecting a rate of reporting that speeds up over time may reduce the overall modeled reported time and reflect any reporting delays (see Section 3.2):

$$k_{er}(t) = \beta_{er} \cdot t$$



To incorporate this rate of reporting into the model, we can define $l\beta_{er} = \log(\beta_{er})$ and re-specify the compartmental model's ODE system as follows:

$$
\begin{aligned}
dEX/dt &= -\exp(l\beta_{er}) \cdot t \cdot EX \\
dOS/dt &= \exp(l\beta_{er}) \cdot t \cdot \exp(lRLR) \cdot EX - \exp(lk_p) \cdot OS \\
dPD/dt &= \exp(lk_p) \cdot \exp(lRRF) \cdot OS
\end{aligned} \tag{4.4}
$$

This structural model can be specified in R using the code in Appendix E.

Revising the definition of $\boldsymbol{f}(P_i, \boldsymbol{\phi}_i, t_j)$ to reflect the state-variable solutions of Eq. (4.4), we can write down a second hierarchical model (Model 2):

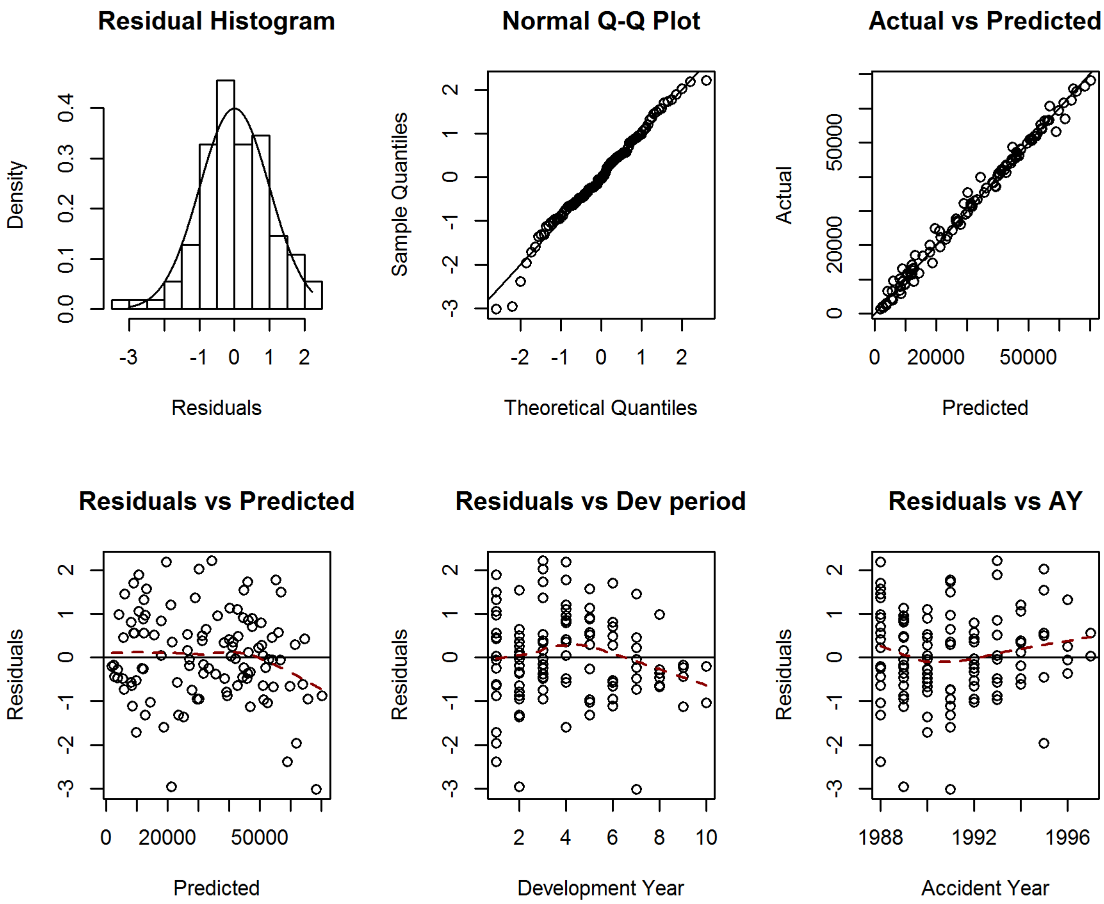
$$
\begin{gathered}
\boldsymbol{y}_{ij} = \boldsymbol{f}(P_i, \boldsymbol{\phi}_i, t_j) + \boldsymbol{\varepsilon}_{ij} \\
\boldsymbol{y}_{ij} = \begin{bmatrix} OS_{ij} \\ PD_{ij} \end{bmatrix}, \quad \boldsymbol{f}(P_i, \boldsymbol{\phi}_i, t_j) = \begin{bmatrix} f_{OS}(P_i, \boldsymbol{\phi}_i, t_j) \\ f_{PD}(P_i, \boldsymbol{\phi}_i, t_j) \end{bmatrix}, \quad \boldsymbol{\varepsilon}_{ij} = \begin{bmatrix} \varepsilon_{ij}{}^{OS} \\ \varepsilon_{ij}{}^{PD} \end{bmatrix} \\
\boldsymbol{\phi}_i = \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \\ \phi_{4i} \end{bmatrix} = \begin{bmatrix} l\beta_{er} \\ lRLR_i \\ lk_p \\ lRRF_i \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} 0 \\ b_{1i} \\ 0 \\ b_{2i} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b}_i \\
\boldsymbol{b}_i \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_1^2 & 0 \\ 0 & \psi_2^2 \end{bmatrix}\right), \qquad \boldsymbol{\varepsilon}_{ij} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & \lambda^2 \end{bmatrix}\right)
\end{gathered} \tag{4.5}
$$

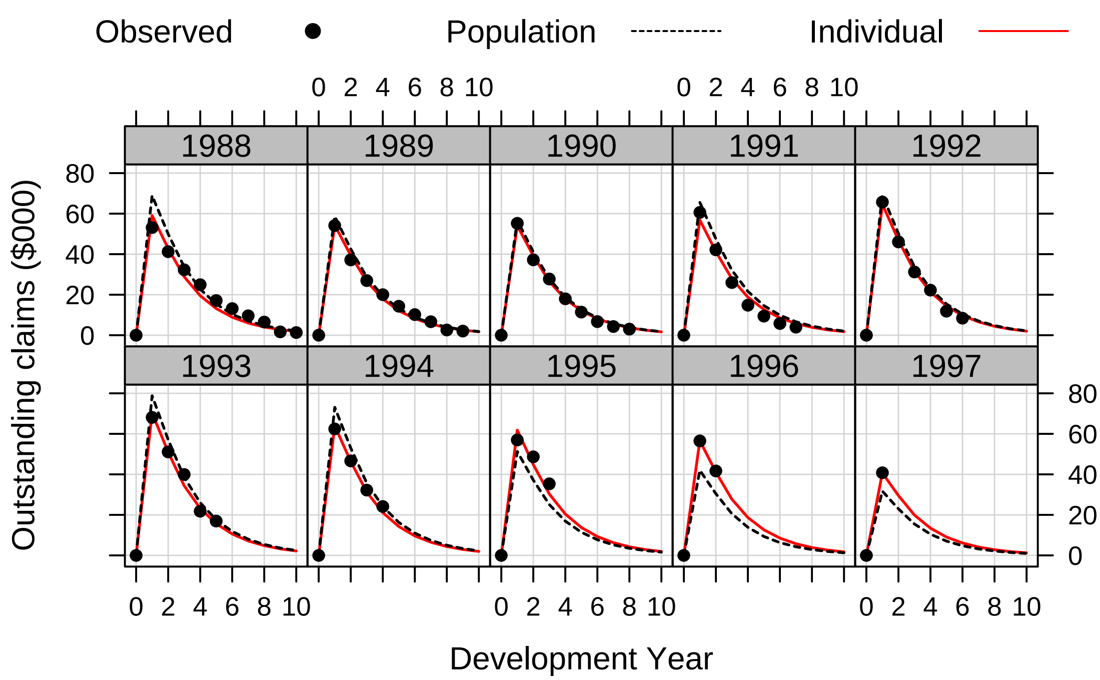
The model form, number of estimable parameters and statistical assumptions are unchanged from the previous model. However, rather than estimating the logarithm of the rate of reporting, we are estimating the logarithm of the linear coefficient for how the rate of reporting increases over development time, i.e. $l\beta_{er}$ replaces lk_{er}.

To ensure that outstanding claims are modeled to be reported over a shorter time-frame than previously, the starting value for $l\beta_{er}$ has been set to 5. This implies a reporting rate that is approximately 1.5 times faster than the Model 1 estimated rate at development year 0.5. The remaining initial parameter values have been set to the estimated fixed-effects in the previous model (to 2 decimal places). The model code and numerical output is contained in Appendix E.

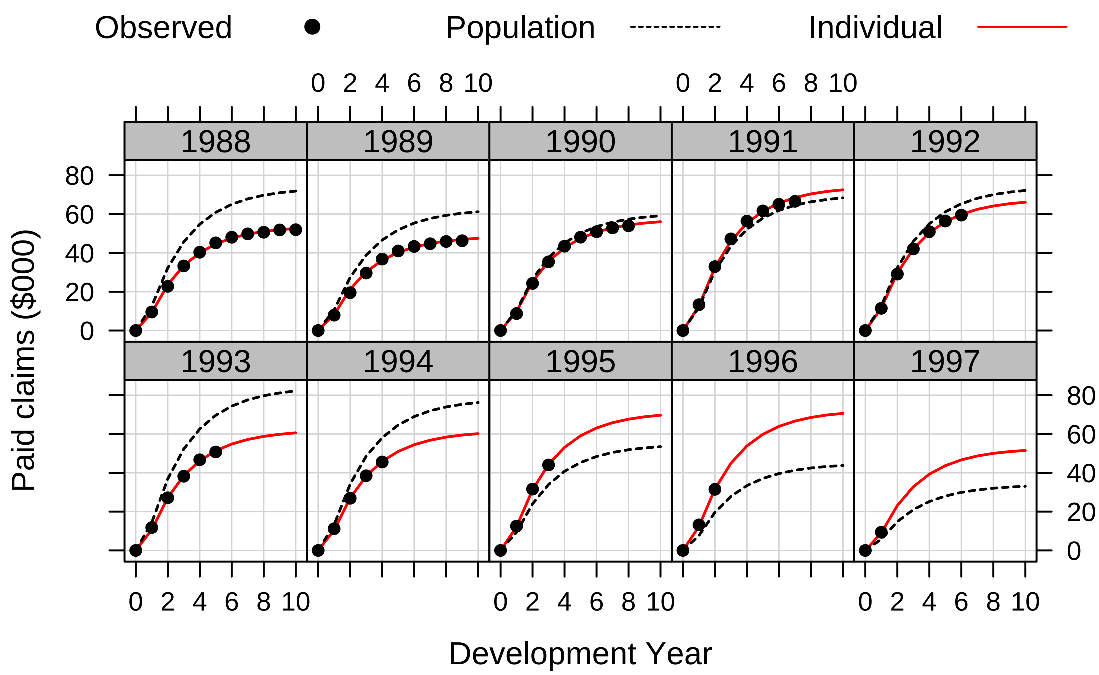
Under Model 2, the within-accident-year error standard deviation for cumulative paid claims fits is estimated to be $\hat{\lambda} = 25\%$ of the within-accident-year error standard deviation for the outstanding claims fits (up from $\hat{\lambda} = 18\%$), which may be due to an improvement in outstanding claims model fits.

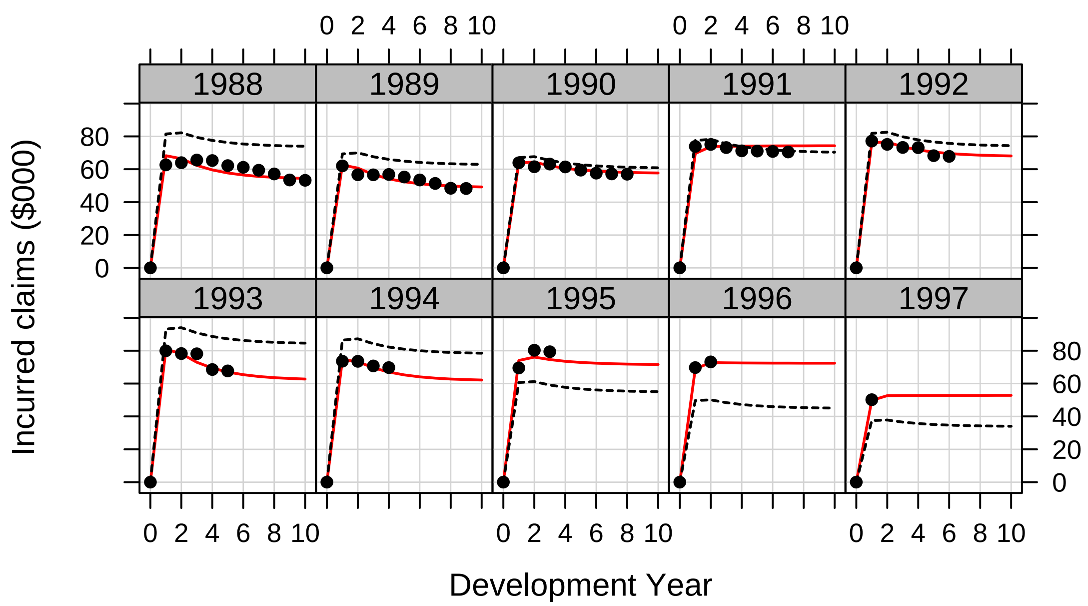
The "Actual vs. Predicted" plot below suggests that this model fits the data more closely than the last; however, the residuals exhibit a minor violation of normality. Furthermore, the "Residuals vs. Development Year" plot has a downwards trend across later development periods, indicating a small degree of over-prediction. Few data points drive this trend however, and therefore it may not be significant.


Residual Histogram
Density
Residuals
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
Actual vs Predicted
Actual
Predicted
Residuals vs Predicted
Residuals
Predicted
Residuals vs Dev period
Residuals
Development Year
Residuals vs AY
Residuals
Accident Year


Observed
Population
Individual
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
Outstanding claims ($000)
Development Year

The inclusion of a time-dependent rate of reporting has resulted in a more accurate description of the outstanding claims peaks. However, for the 1991 accident year there is evidence of continued over prediction, perhaps due to a differing rate of payment for this year.





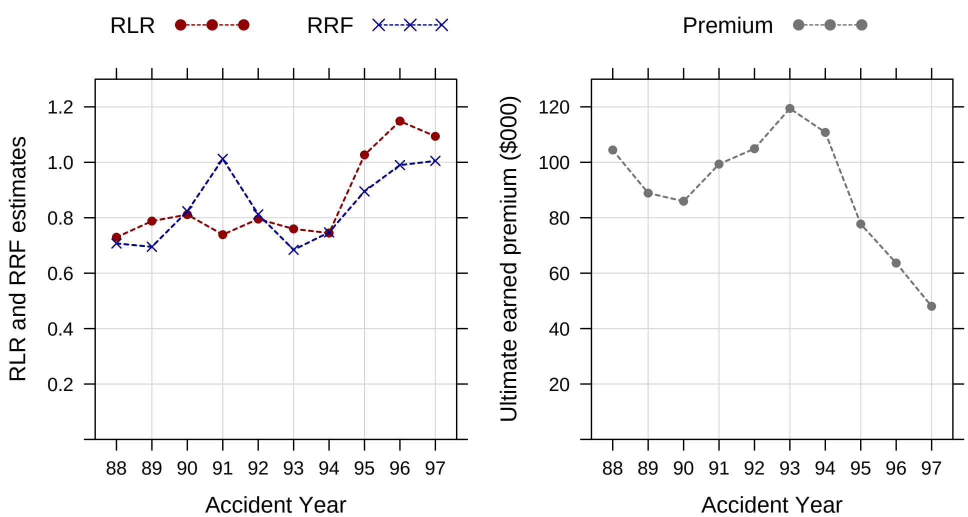
Paid claims are slightly over-predicted for later development periods, consistent with the residual plots. However, the incurred fits are improved due to the more accurate description of outstanding claims. A statistical comparison of this model against the last shows that the information criterion statistics (AIC and BIC) have both reduced. Therefore we may deduce that Model 2 is preferred to Model 1 and inspect it in greater detail (see Appendix E)

Approximate 95% confidence intervals show that the fixed-effects $\boldsymbol{\beta}$ (the mean-level logarithm of the compartmental reserving model parameters) are statistically significantly different from zero at the 1% level. Furthermore, the estimated fixed-effects correlation matrix contains a strong negative correlation between the rate of reporting and rate of payment parameters (-0.72). This seems intuitive: if few claims are reported over a given time period, a case handling team is likely to be better equipped to handle each payment more quickly than if many claims are reported over an equivalent time period.

At this stage the structural model could justifiably be selected as final. However, for other datasets further modifications may be required, such as those outlined in Section 3.2.

4.1.2 Random-effects correlation

In a hierarchical framework there are various possible statistical model modifications. For example, correlations between random-effects can be explored. The graphs below show the Model 2 estimated RLR_i and RRF_i parameters for each accident year alongside earned premiums for illustration:



The first plot suggests a positive correlation between the reported loss ratio and reserve robustness factor parameters by accident year, indicative of a case reserving cycle effect, i.e. more conservative case reserves (low RRF_i) in a hard market (low RLR_i) to create cushions for the future (Line *et al.*, 2003). The model estimates market softening between 1994 and 1997 (increasing RLR_i); a conclusion supported by reducing premium volumes across these years.

Additionally, case reserves are estimated to be increasingly robust between 1993 and 1997, which corroborates the reducing downward trend for incurred model fits across these years. There is a discrepancy for the 1991 accident year where the data appear to be exhibiting over-reserving, yet the model does not recognize this.

To estimate the correlation between the random-effects for $lRLR_i$ and $lRRF_i$, we can update the random-effects variance-covariance matrix to define a third model (Model 3 in Appendix E):

$$\mathbf{\Psi} = \begin{bmatrix} \psi_1^2 & \psi_{12} \\ \psi_{21} & \psi_2^2 \end{bmatrix}$$

The covariance between random-effect q and random-effect $r \neq q$ is denoted ψ_{qr}. The updated model estimates a strong and statistically significant positive correlation between the estimated reported loss ratio and reserve robustness factor random-effects (0.78).

To assess whether this model is significantly improved from the last, a likelihood ratio test can be carried out. The resultant p-value of 0.013 indicates that the hypothesis that the correlation between the random-effects is zero can be rejected at the 5% level (but not at the 1% level). We may therefore marginally prefer Model 3 to Model 2, particularly if we wish to make inferences about the correlation between $lRLR_i$ and $lRRF_i$ to assess case reserve cycle strength.

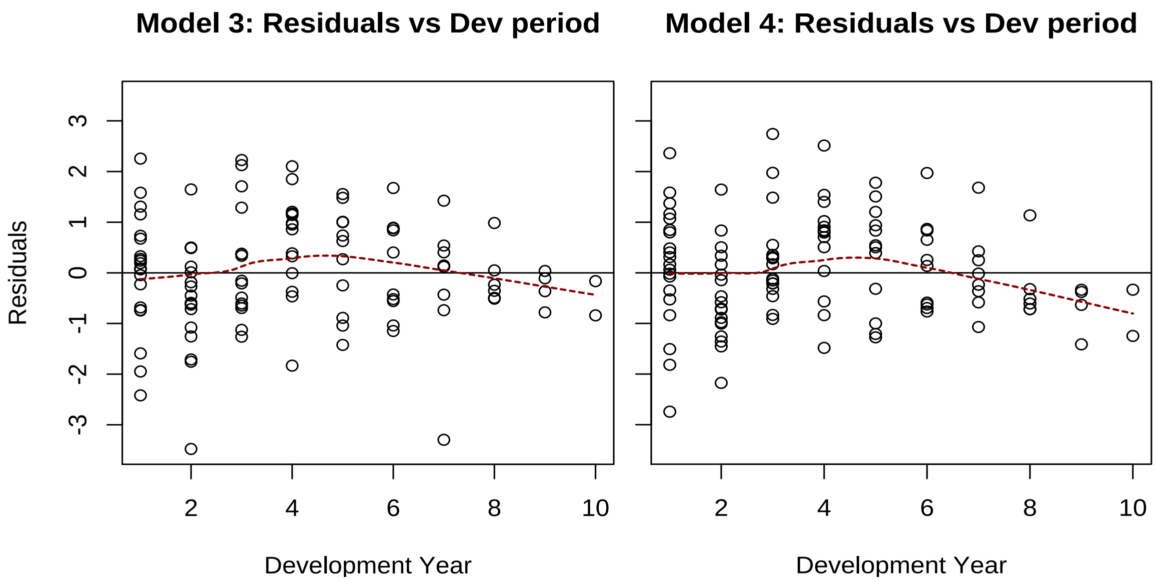
We could add random-effects for the remaining compartmental model parameters to define a fourth model. For example, a "block-diagonal" random-effects variance-covariance structure (Pinheiro and Bates, 2000) allows the rate of payment to vary by accident year independently of $lRLR_i$ and $lRRF_i$, resulting in differing payment patterns by accident year:

$$\mathbf{\Psi} = \begin{bmatrix} \psi_1^2 & \psi_{12} & 0 \\ \psi_{21} & \psi_2^2 & 0 \\ 0 & 0 & \psi_3^2 \end{bmatrix}$$

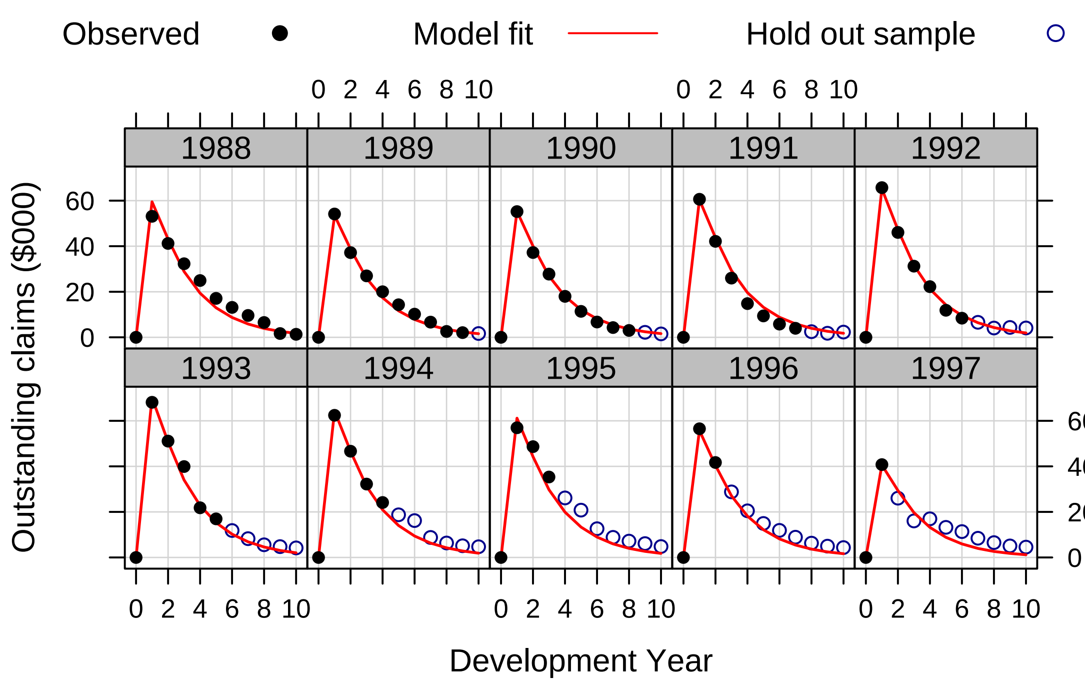
The statistical comparisons for Model 4 against the previous models (Appendix E) show a reduced BIC and significant likelihood ratio test for the new random-effect, suggesting that Model 4 should be preferred. However, the "Residuals vs. Development Year" diagnostic comparisons below tell a different story. Although Models 3 and 4 both produce a downwards residual trend which indicates a degree of over-prediction, Model 4's trend is stronger.

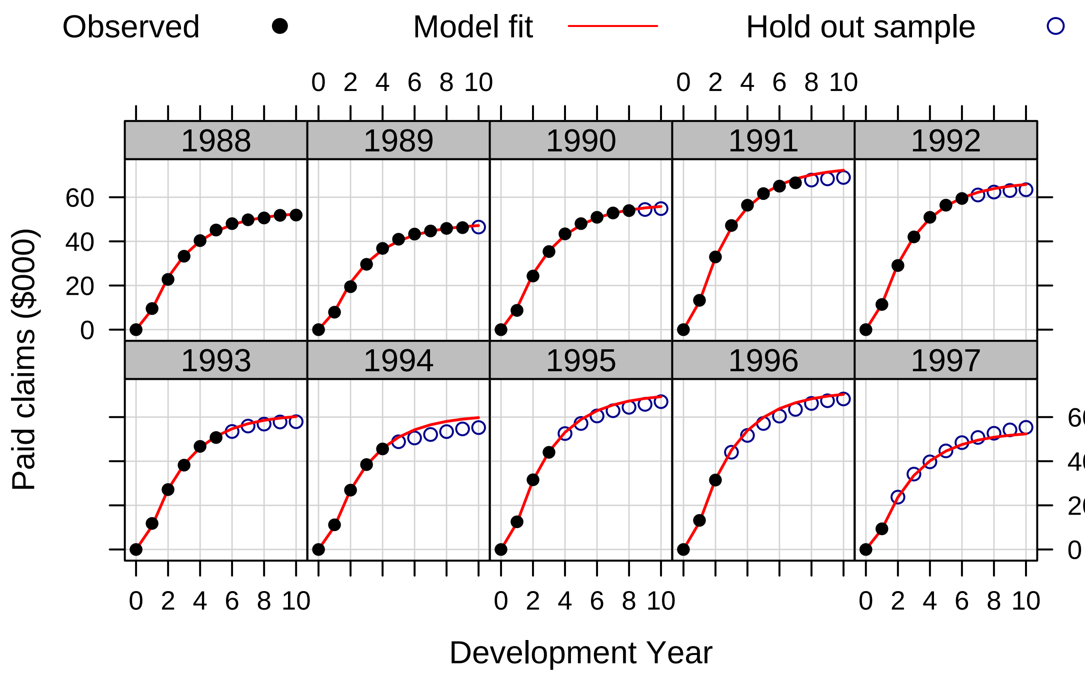
A double-log transformation $\log\{\boldsymbol{y}_{ij}\} = \log\{\boldsymbol{f}(P_i, \boldsymbol{\phi}_i, t_j)\} + \boldsymbol{\varepsilon}_{ij}$ reduces the downwards trend for both models, but residual normality is consequently violated (not shown).

Although the residual plot for Model 3 has two outliers, we may judge this model more suitable for best estimate reserving purposes if it is considered less likely to over-project ultimate losses. On this basis Model 4 will be rejected in favor of Model 3 (noting that either model could be justifiably selected).

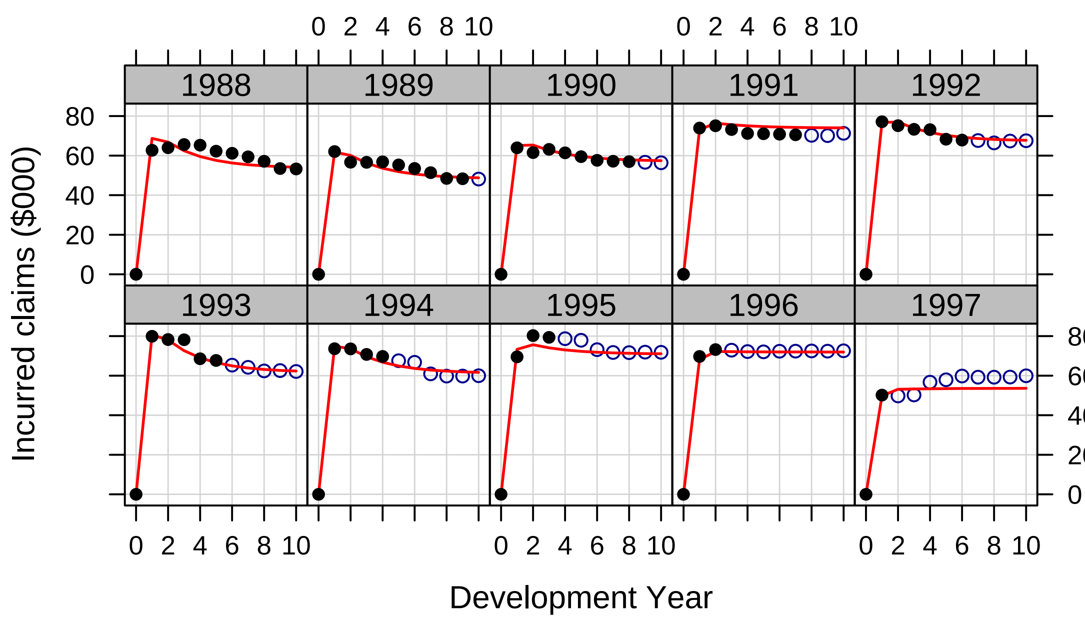


While it may be possible to further improve model fits by experimenting with alternative initial parameter values, Model 3 appears adequate based on the residual plot above and individual accident year fits (see below). We can therefore select Model 3 as final and compare its projections against the lower triangle hold-out samples (open circles) as follows:





Outstanding claims extrapolations have generally under-estimated actual outstanding development, while the cumulative paid claims extrapolations have generally over-estimated actual cumulative paid development.

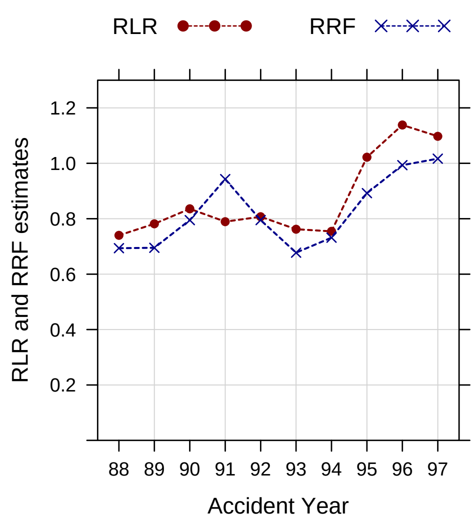


The under- and over-projections largely offset each other for the incurred extrapolations, although the aforementioned 1991 accident year development fit issue has propagated into the extrapolation.

In contrast to the historical upper triangle development, hold-out sample outstanding claims have taken longer to converge to zero and hold-out sample cumulative paid claims increases have tapered relative to their initial "growth". These characteristics suggest a **slow-down in the rate of payment in the hold-out sample**, perhaps consistent with the nonlinear rate of payment function defined in Section 3.2. It could be that the estimated softening market (see graph below) led to tighter cashflow and slower payments (Line *et al.*, 2003). There may also be reserve robustness improvements in later hold-out development years. Additionally, the residual plots at the fitting stage displayed some evidence of over-prediction, which could partially account for the paid claims over-projections.

Although the modeled dataset showed insufficient evidence of a payment rate reduction over time, had cashflow tightening been anticipated as a result of the estimated softening market, a practitioner could have scenario tested slowdowns in the rate of payment for the purpose of setting reserves.

In addition to payment rate reductions, case reserve robustness appears to have increased between the 1993 and 1996 accident years, shown by negative incurred development flattening across these years. Furthermore, the 1997 accident year appears to have exhibited under-reserving (or late reporting/claim re-openings) in contrast to the over-reserving trend seen in previous years. The compartmental model estimated increasing reserve robustness between 1993 and 1996, and a small amount of under-reserving for 1997. This is despite there being only two observations available for modeling 1997, resulting in a fit principally reliant on data-rich years which exhibited over-reserving:



The hierarchical compartmental reserving (CR) modeled development time 10 and ultimate incurred claims (time ∞, given by $P_i \times \widehat{RLR}_i \times \widehat{RRF}_i$) are shown below, alongside the Munich chain ladder (MCL; Quarg and Mack, 2004) and basic chain ladder (CL) incurred method results (without tail factors) by accident year.

To compare the predictability of each method, the percentage differences from the actual time 10 incurred claims are shown with the closest estimate(s) highlighted:

AY	Time 10 Incurred	CR Incurred t=10	CR Incurred t=∞	MCL Incurred	CL Incurred	var(CR)	var(MCL)	var(CL)
1988	53,261	54,149	*53,611*	53,261	53,261	2%	0%	0%
1989	48,162	48,769	*48,288*	47,640	48,109	1%	-1%	0%
1990	56,368	57,447	*57,112*	57,132	54,697	2%	1%	-3%
1991	71,274	74,028	*73,926*	72,016	65,550	4%	1%	-8%
1992	67,515	67,718	*67,323*	66,276	61,847	0%	-2%	-8%
1993	62,122	62,331	*61,664*	60,035	60,658	0%	-3%	-2%
1994	59,974	61,670	*61,160*	59,663	60,521	3%	-1%	1%
1995	71,829	71,073	*70,878*	69,426	66,815	-1%	-3%	-7%
1996	72,573	71,970	*71,959*	69,680	61,118	-1%	-4%	-16%
1997	59,939	53,597	*53,617*	49,977	42,242	-11%	-17%	-30%
Total	**623,017**	**622,751**	***619,537***	**605,106**	**574,819**	**0%**	**-3%**	**-8%**

The following conclusions can be drawn:

- The compartmental model produces the closest time 10 incurred loss estimates in total;
- The superior estimation accuracy of the compartmental approach for less mature accident years can be accredited to the model estimating increasingly robust case reserve setting (driven by a softening market – see above); and
- The Munich chain ladder method recognizes a shift in case reserve robustness by utilizing paid claims development. However, the basic chain ladder method does not, resulting in heavily under-estimated time 10 incurred losses.

In practice the Bornhuetter-Ferguson method (1972) may be used for the less mature years, possibly closing the estimation accuracy gap. Although not shown, the compartmental modeled ultimate paid and incurred estimates are equal whereas the Munich chain ladder estimates differ.

Thus far we have only considered point estimates. However, a compartmental framework enables scenario testing of one or more of the claims process parameters to generate a range of possible ultimate claims. For example, case reserving philosophy or settlement approaches could be discussed with the relevant case handlers/claims teams to establish a range of plausible RRF and/or k_p parameters.

Prediction errors could be assessed analytically or using bootstrapping techniques (England and Verrall, 1999). Additionally or alternatively, a hierarchical compartmental reserving model could be specified in a fully Bayesian framework, which will be explored in the following Sections.

4.2 Bayesian modeling

We may wish to implement the selected frequentist model within a Bayesian framework for the reasons outlined in Section 2.3.1. In particular:

- Judgment and information external to the claims triangle data can be robustly incorporated;
- Reserve uncertainty can be quantified as part of the fitting process; and
- Flexibility enables additional model features to be incorporated with relative ease.

We can specify a Bayesian hierarchical model by rewriting the paid and outstanding compartmental model differential equation solutions in Eq. (4.5) to explicitly state parameters with random-effects ($\boldsymbol{\phi}_i$) and without random-effects ($\boldsymbol{\eta}$). The Bayesian implementation will incorporate autoregressive sub-models for outstanding and cumulative paid claims residuals to reduce recurrent under/over prediction (Zhang, Dukic and Guszcza, 2012):

$$
\begin{gathered}
OS_{ij} = f_{OS}\left(P_i, \boldsymbol{\phi}_i, \boldsymbol{\eta}, t_j\right) + \varepsilon_{ij}^{OS} \\
PD_{ij} = f_{PD}\left(P_i, \boldsymbol{\phi}_i, \boldsymbol{\eta}, t_j\right) + \varepsilon_{ij}^{PD} \\
\boldsymbol{\phi}_i = \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \end{bmatrix} = \begin{bmatrix} lRLR_i \\ lRRF_i \end{bmatrix} \qquad\qquad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} lk_{er} \\ lk_p \end{bmatrix} \\
\begin{bmatrix} \phi_{1i} \\ \phi_{2i} \end{bmatrix} \sim N_2\left(\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \boldsymbol{\Psi} = \begin{bmatrix} \psi_1^2 & \psi_{12} \\ \psi_{21} & \psi_2^2 \end{bmatrix}\right) \\
\varepsilon_{ij}^{OS} = \rho_{OS}\varepsilon_{ij-1}^{OS} + \delta_{ij}^{OS} \qquad\qquad \varepsilon_{ij}^{PD} = \rho_{PD}\varepsilon_{ij-1}^{PD} + \delta_{ij}^{PD} \\
\delta_{ij}^{OS} \sim N\{0, \sigma_{OS}^2(1-\rho_{OS}^2)\} \qquad\qquad \delta_{ij}^{PD} \sim N\{0, \sigma_{PD}^2(1-\rho_{PD}^2)\} \\
\varepsilon_{i1}^{OS} \sim N(0, \sigma_{OS}^2) \qquad\qquad \varepsilon_{i1}^{PD} \sim N(0, \sigma_{PD}^2)
\end{gathered} \tag{4.6}
$$

The statistical assumptions are analogous to the selected frequentist model and similarly, $lRLR_i$ and $lRRF_i$ are assumed to vary by accident year with co-dependency. Residual autocorrelation terms are denoted ρ_{OS} and ρ_{PD}, and model process error is captured by the residual error terms ε_{ij}^{OS} and ε_{ij}^{PD}.

Normal prior distributions have been assigned to the implied *fixed-effects*. Similarly to the frequentist model these are the means of $lRLR_i$ and $lRRF_i$ (denoted $\boldsymbol{\theta}$), together with $l\beta_{er}$ and lk_p (denoted $\boldsymbol{\eta}$):

$$
\begin{gathered}
\boldsymbol{\theta} \sim N_2(\overline{\boldsymbol{\theta}}, \boldsymbol{\Omega}) \\
\boldsymbol{\eta} \sim N_2(\overline{\boldsymbol{\eta}}, \boldsymbol{\Pi})
\end{gathered} \tag{4.7}
$$

In Eq. (4.7), $\overline{\boldsymbol{\theta}}$ and $\boldsymbol{\Omega}$ denote the prior mean and variance-covariance matrix of $\boldsymbol{\theta}$, whereas $\overline{\boldsymbol{\eta}}$ and $\boldsymbol{\Pi}$ denote the prior mean and variance-covariance matrix of $\boldsymbol{\eta}$.

The prior means for the fixed-effects have been set to the estimated fixed-effects in the selected frequentist model, and the prior variance-covariance matrices describing uncertainty in the fixed-effects have been set to replicate the frequentist estimated fixed-effects confidence intervals:

$$\begin{aligned} \bar{\boldsymbol{\theta}} &= \{-0.15, -0.21\}^T & \quad \boldsymbol{\Omega} &= \begin{bmatrix} 0.0513^2 & 0 \\ 0 & 0.0506^2 \end{bmatrix} \\ \bar{\boldsymbol{\eta}} &= \{1.7, -0.9\}^T & \quad \boldsymbol{\Pi} &= \begin{bmatrix} 0.0392^2 & 0 \\ 0 & 0.0124^2 \end{bmatrix} \end{aligned} \quad \textbf{(4.8)}$$

These priors imply fixed-effect independence; however, their posterior distributions can demonstrate dependence. A prior distribution has also been assigned to the variance-covariance matrix of $\boldsymbol{\phi}_i$ ($\boldsymbol{\Psi}$) i.e. the variance of the implied *random-effects*, describing the magnitude of variability for the accident year varying (log) proportion parameters $lRRF_i$ and $lRLR_i$:

$$\boldsymbol{\Psi} \sim W_2^{-1}(\boldsymbol{\Sigma}, \nu) \quad \textbf{(4.9)}$$

W_2^{-1} is an inverse-Wishart distribution (and conjugate prior) with 2×2 scale matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom (Gelman *et al.*, 2013). The frequentist analysis results have not been used to inform this prior. Instead, a vague prior has been set to allow the variance-covariance matrix to be principally estimated from the data. The prior *inverse* scale matrix and degrees of freedom have been set as

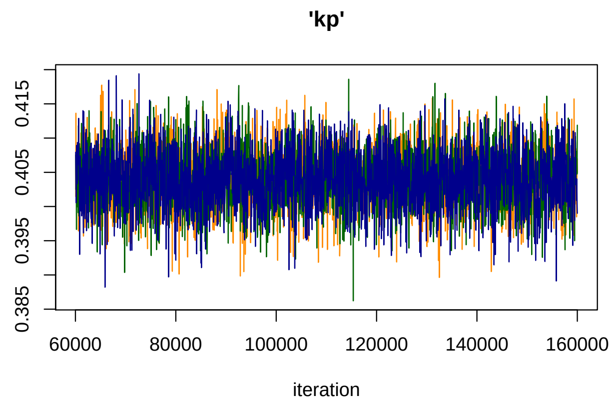
$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}, \nu = 2 \quad \textbf{(4.10)}$$

where the degrees of freedom are as low as possible while still maintaining a proper distribution (Johnson and Kotz, 1972). Although this prior is vague in its description of accident year variability magnitude, the off-diagonal elements have been set to give a 0.80 positive correlation between $lRLR_i$ and $lRRF_i$ (recall that the estimated correlation in the selected frequentist model was 0.78).

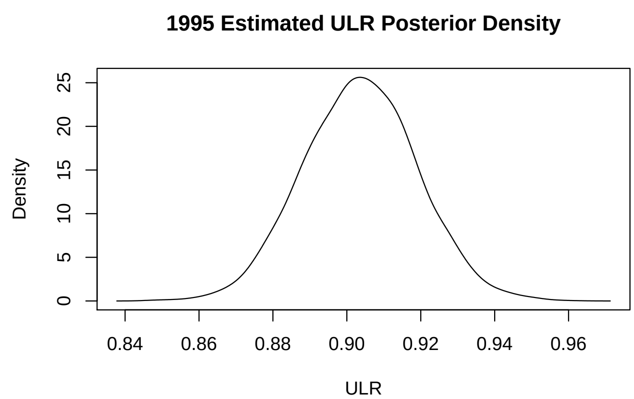
Vague priors have been assigned to the remaining model parameters. Priors for the standard deviations of the within-accident-year errors have been selected to comfortably cover the standard deviations estimated in the selected frequentist model. Finally, priors for the correlation terms of the autoregressive processes have been set to cover the minimum and maximum correlation values:

$$\begin{aligned} \sigma_{OS} &\sim U(0{,}10000) & \quad \rho_{OS} &\sim U(-1{,}1) \\ \sigma_{PD} &\sim U(0{,}5000) & \quad \rho_{PD} &\sim U(-1{,}1) \end{aligned} \quad \textbf{(4.11)}$$

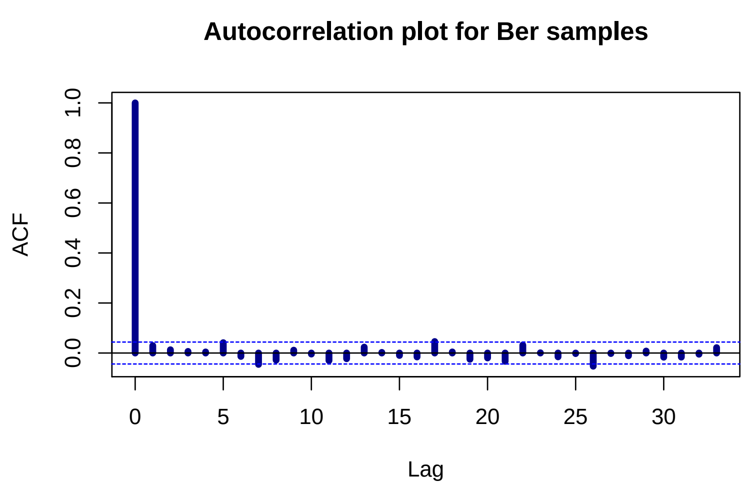
Using OpenBUGS (**B**ayesian inference **U**sing **G**ibbs **S**ampling; Lunn *et al.*, 2000), three parallel Markov chains were run with 60,000 burn-in iterations per chain, followed by 100,000 iterations per chain. To reduce sample autocorrelation, every 50th iteration of each chain was used, resulting in 2,000 simulated draws per chain and 6,000 samples in total. Various diagnostics checks were carried out to ensure that the simulation had converged to its approximate stationary distribution. Individual parameter estimation convergence was initially assessed and, as an example below, the values of k_p have been plotted over MCMC iterations by Markov chain.



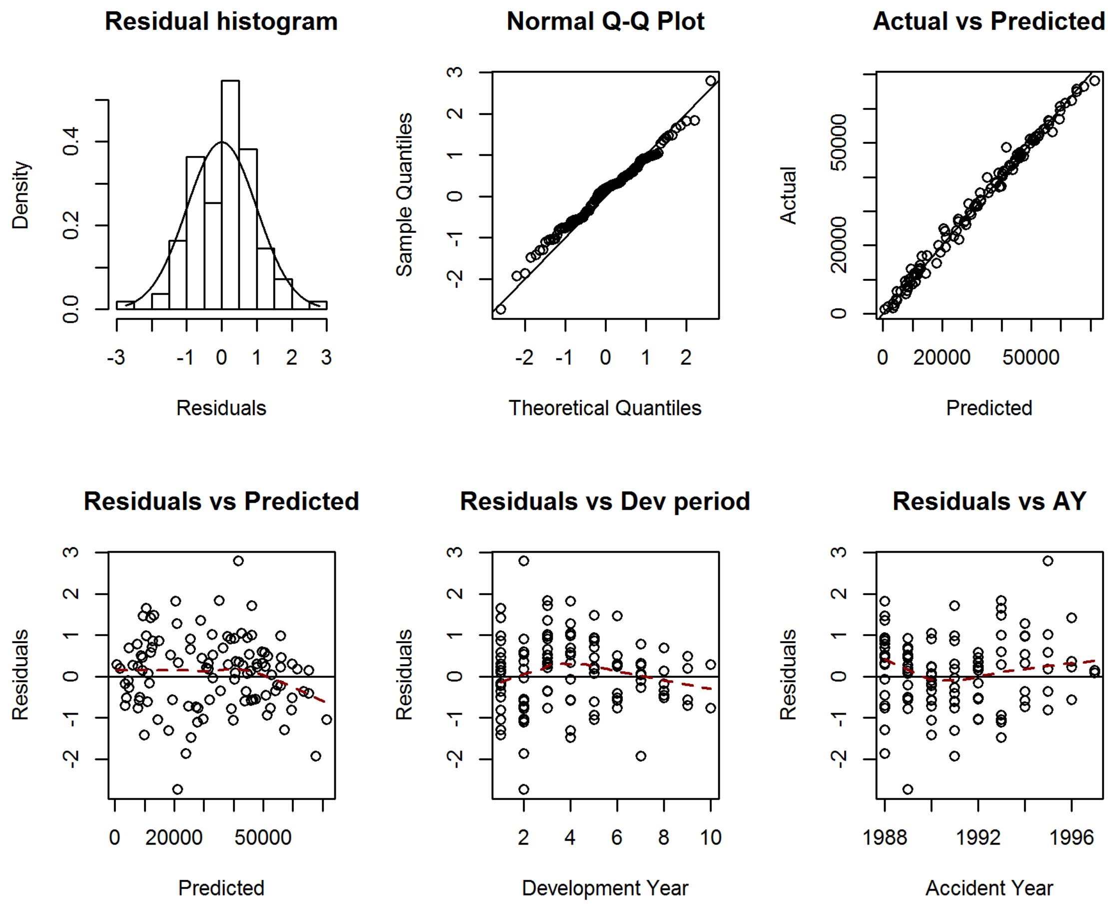
The similar and stable chains indicate that the posterior distribution of k_p has approximately converged to its stationary distribution. Densities for model parameters were also inspected. The estimated ultimate loss ratio posterior density for the 1995 accident year is as follows:



The density is smooth and bell-shaped, suggesting that convergence has been achieved. Finally, checks were carried out to assess sample autocorrelation. The plot below shows that the autocorrelation of the second chain β_{er} samples is not statistically different from zero:

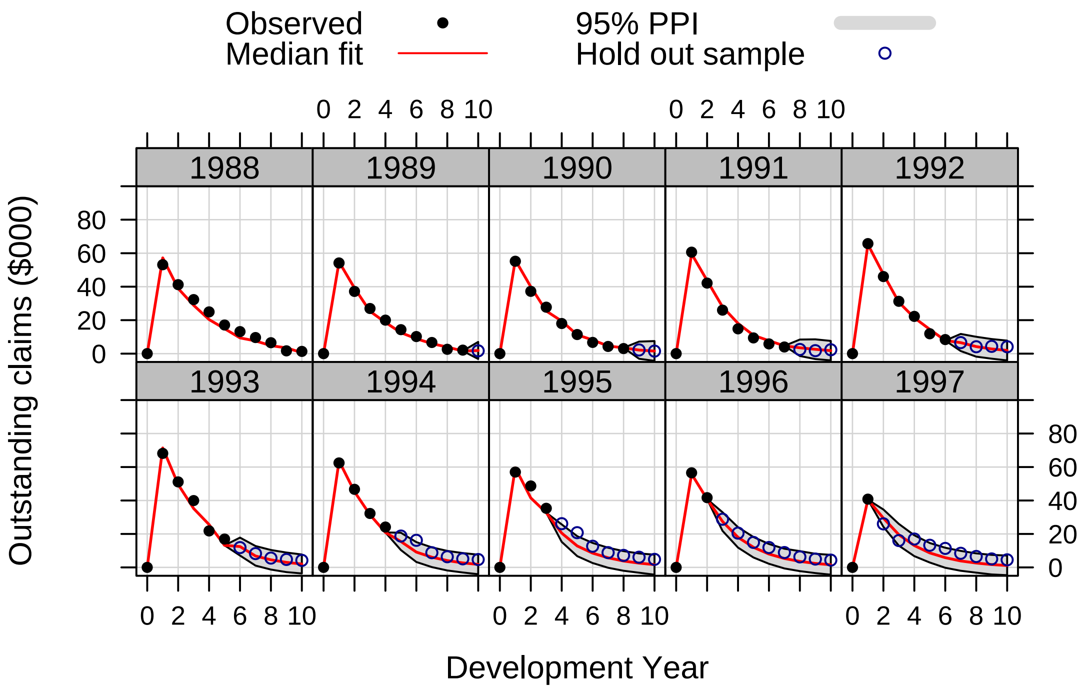


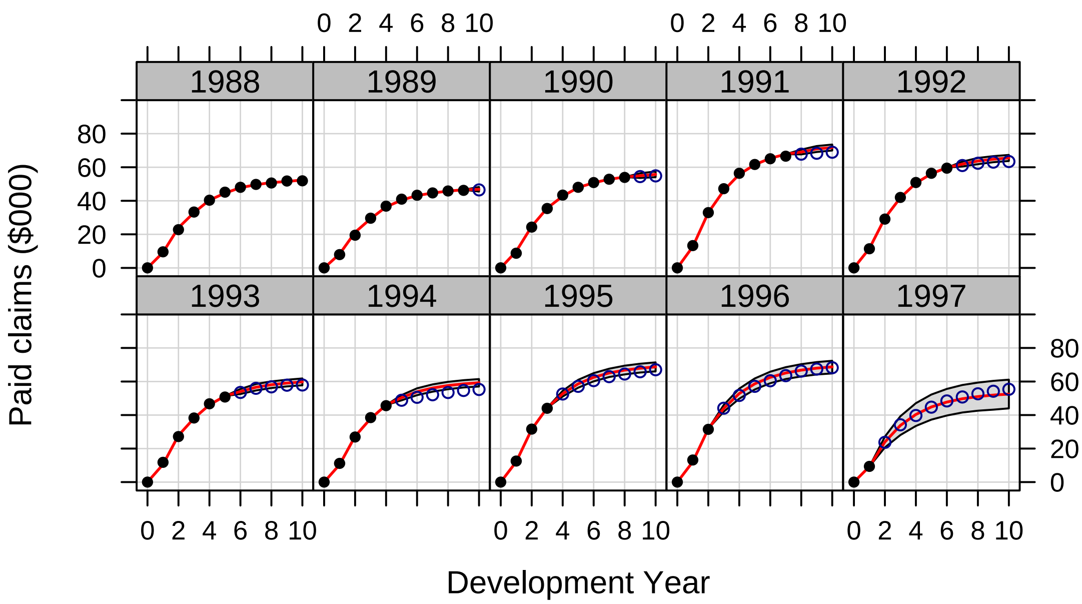
Given the diagnostics above (and various others not shown) the simulation appears to have reached approximate convergence and we can proceed to inspect the model diagnostic plots. Posterior densities are estimated for all parameters of interest, and therefore the diagnostic plots are based on estimated posterior density *medians*:



Residual normality approximately holds and the model fits are close to the observations. However, similarly to the frequentist model there is a downward trend in the "Residuals vs. Development Year" plot across later development years.

Similarly to Section 4.1, the individual accident year fits can be inspected. In the Bayesian setting however, for unobserved development years ($t_j \in i + j \geq 12$) 95% posterior predictive intervals ("PPIs") can be plotted (Gelman *et al.*, 2013). These show a range of prediction uncertainty due to both parameter and process uncertainty. Since this model is a Bayesian implementation of the selected frequentist model, we will compare the median fits, extrapolations and PPIs to the observed and hold-out sample development together:





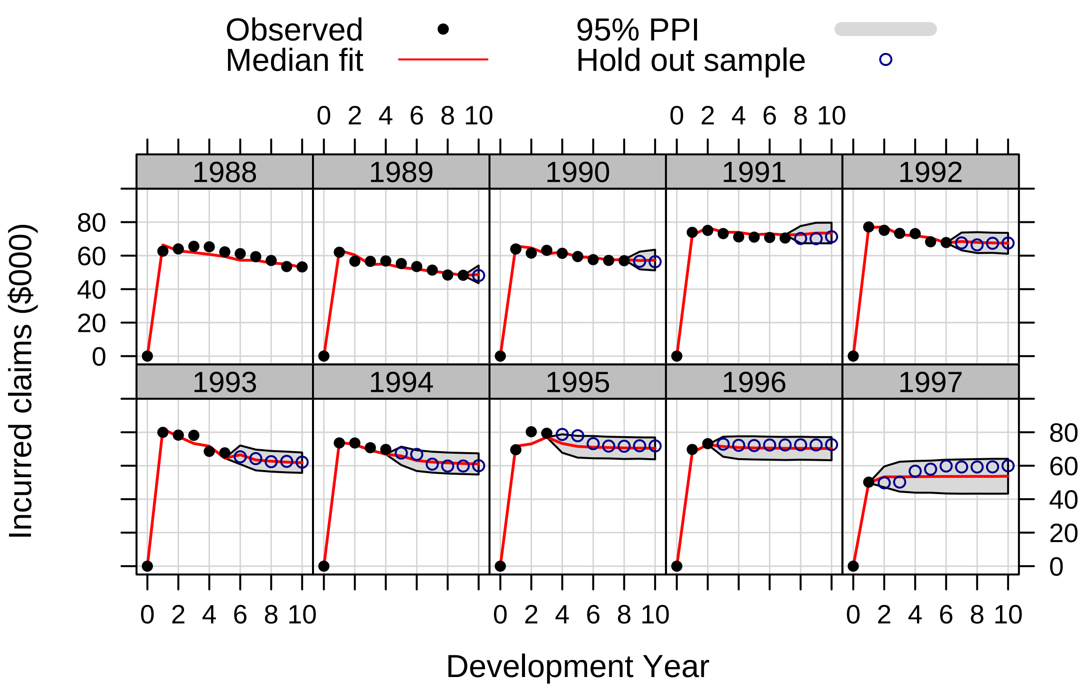
The median fits are similar to the selected frequentist model with some minor improvements. The PPIs are slightly wider for less mature accident years and contain the possibility of both under- and over- reserving. However, the outstanding claims PPIs do not converge to zero and even fall below zero in later development periods because of the residual normality assumption in Eq. (4.6).

To address this shortfall, a double-log transformed model form was tested:

$$\log(OS_{ij}) = \log\{f_{OS}(P_i, \boldsymbol{\phi}_i, \boldsymbol{\eta}, t_j)\} + \varepsilon_{ij}^{OS}$$
$$\log(PD_{ij}) = \log\{f_{PD}(P_i, \boldsymbol{\phi}_i, \boldsymbol{\eta}, t_j)\} + \varepsilon_{ij}^{PD} \quad \textbf{(4.12)}$$

Similarly to the equivalent frequentist model, this transformation resulted in a violation of residual normality. In particular, there were too many small residuals relative to larger residuals, which is characteristic of an overfitted model. Therefore the model was rejected.

The paid claims PPIs are generally narrower than the outstanding claims PPIs due to closer model fits: median $\hat{\sigma}_{PD} = 760$ and median $\hat{\sigma}_{OS} = 3151$. However, paid claims are over-projected similarly to the frequentist model, suggesting that a smaller residual error variance could instil false extrapolation confidence if possible future development period claims process shifts are not considered.



To assess posterior parameter uncertainty, we can review median parameter estimates and their 95% central posterior intervals {median [2.5%ile, 97.5%ile]} (Gelman *et al.*, 2013). For the 1997 accident year $\widehat{RLR}_{10} = 1.10\ [0.95, 1.25]$ and $\widehat{RRF}_{10} = 1.02\ [0.83, 1.23]$, suggesting that case reserve robustness is the main driver of ULR uncertainty ($\widehat{ULR}_{10} = 1.12\ [0.93, 1.30]$).

The estimated residual autocorrelations are $\hat{\rho}_{OS} = 0.58\ [0.30, 0.83]$ and $\hat{\rho}_{PD} = 0.55\ [0.27, 0.75]$, indicating moderate to strong serial correlation. The estimated accident year correlation between $lRLR_i$ and $lRRF_i$ is $\hat{\rho}_{lRLR_i lRRF_i} = \hat{\psi}_{12}/\{\hat{\psi}_1\hat{\psi}_2\} = 0.77\ [0.38, 0.93]$, indicating a strong case reserving cycle effect. However, the 95% posterior interval is quite wide and the extent of the estimated effect is significantly influenced by the variance-covariance matrix prior in Eq. (4.10).

Model predictive power can be evaluated by inspecting the 95% PPI hold-out sample coverages as follows:

95% PPI Coverage	**1-year ahead**	**10-years ahead**	**Total**
Outstanding	89%	100%	93%
Paid	100%	67%	82%
Incurred	100%	100%	98%

The outstanding and incurred claims PPI coverages are close to the nominal 95% rate across all time horizons. The poor coverage for the 10-year ahead and total paid claims hold-out samples can be attributed to over-projection, particularly for the 1994 accident year. Removing this year from the coverage calculations gives a 10-year coverage of 78% and total coverage of 93%.

The over-projections may be the result of hold-out sample rate of payment reductions (see Section 4.1.2) and/or differing rates of payment by accident year not reflected by the structural model and PPIs. Similarly to the frequentist setting, we could have scenario tested slow-downs in the rate of payment over development time. PPI coverage could possibly have been improved in practice by using informative priors for the random-effects or, alternatively, by increasing the number of random effects. The latter option will be explored in the following scenario.

4.2.1 Scenario 1: Fully random structure

We may be able to achieve a more accurate description of historical claims development by allowing *all* claims process parameters to vary by accident year:

$$\boldsymbol{\phi}_i = \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \\ \phi_{4i} \end{bmatrix} = \begin{bmatrix} lk_{er,i} \\ lRLR_i \\ lk_{p,i} \\ lRRF_i \end{bmatrix}$$

$$\begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \\ \phi_{4i} \end{bmatrix} \sim N_4 \left(\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}, \boldsymbol{\Psi} = \begin{bmatrix} \psi_1^2 & \psi_{12} & \psi_{13} & \psi_{14} \\ \psi_{21} & \psi_2^2 & \psi_{23} & \psi_{24} \\ \psi_{31} & \psi_{32} & \psi_3^2 & \psi_{34} \\ \psi_{41} & \psi_{42} & \psi_{43} & \psi_4^2 \end{bmatrix} \right) \quad \textbf{(4.13)}$$

$\boldsymbol{\Psi}$ contains ten estimable parameters, which may not be supported by this dataset. However, negligible posterior covariance terms can be enforced by setting a prior assumption that $lk_{er,i}$ and $lk_{p,i}$ vary independently of all other parameters (see below). The assigned prior distributions and assumed parameter values are unchanged from the previous model, yet fewer priors are required because we do not need to distinguish between those parameters that do and do not vary by accident year.

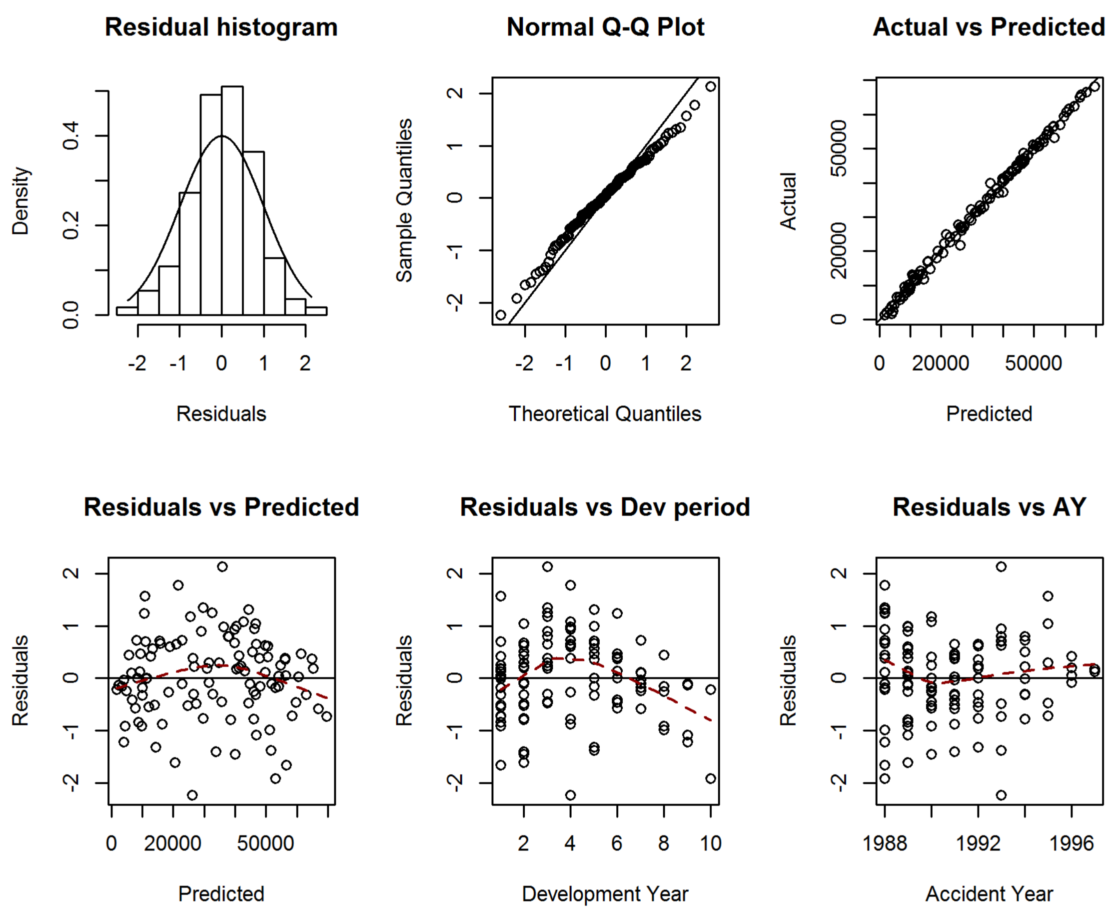
The fixed-effects prior assumptions are as follows:

$$\boldsymbol{\theta} \sim N_4(\overline{\boldsymbol{\theta}}, \boldsymbol{\Omega})$$
$$\overline{\boldsymbol{\theta}} = \{1.7, -0.15, -0.9, -0.21\}^T$$
$$\boldsymbol{\Omega} = \begin{bmatrix} 0.0392^2 & 0 & 0 & 0 \\ 0 & 0.0513^2 & 0 & 0 \\ 0 & 0 & 0.0124^2 & 0 \\ 0 & 0 & 0 & 0.0506^2 \end{bmatrix} \qquad \textbf{(4.14)}$$

The random-effects variance-covariance matrix prior assumptions are as follows:

$$\boldsymbol{\Psi} \sim W_4^{-1}(\boldsymbol{\Sigma}, \nu)$$
$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.8 \\ 0 & 0 & 1 & 0 \\ 0 & 0.8 & 0 & 1 \end{bmatrix}, \nu = 4 \qquad \textbf{(4.15)}$$

This assumes independence of the random-effects for $lk_{er,i}$ and $lk_{p,i}$. The remaining parameter priors, statistical assumptions and convergence arguments are unchanged from the previous model.

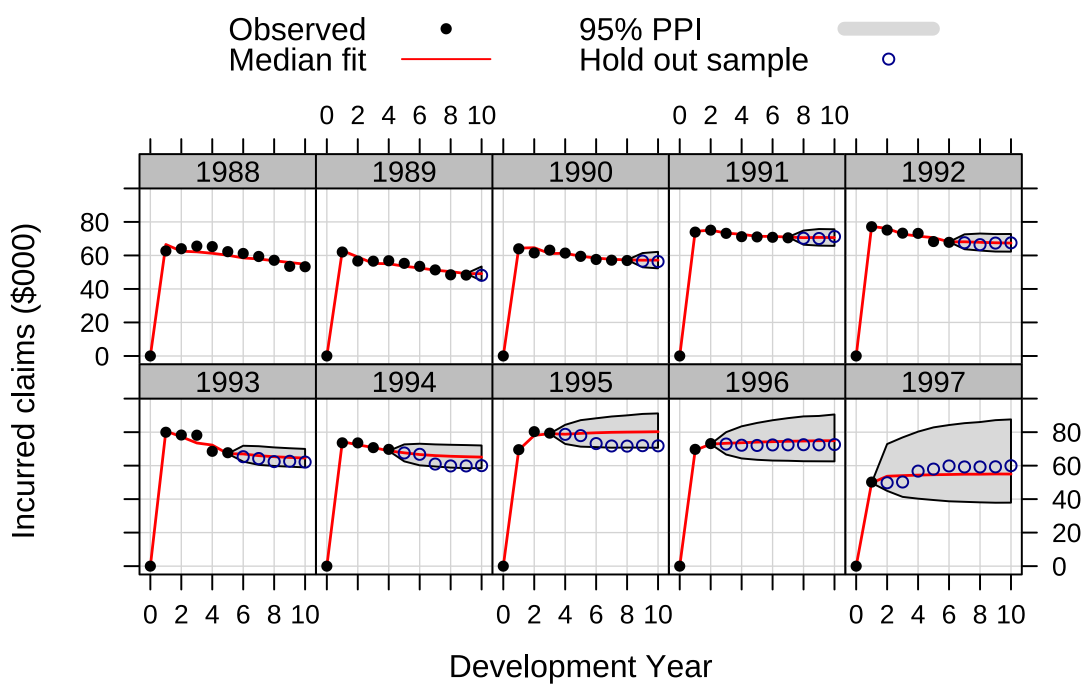


The diagnostics reveal that the fits are closer to the observations than the previous model. However, the residuals appear to be in violation of normality, indicating a degree of overfitting. Additionally, the downwards residual vs. development year trend is worsened (analogously to when complexity was increased in the frequentist modeling).

To assess this model against the last, we can compare each model's deviance information criterion (DIC) as follows:

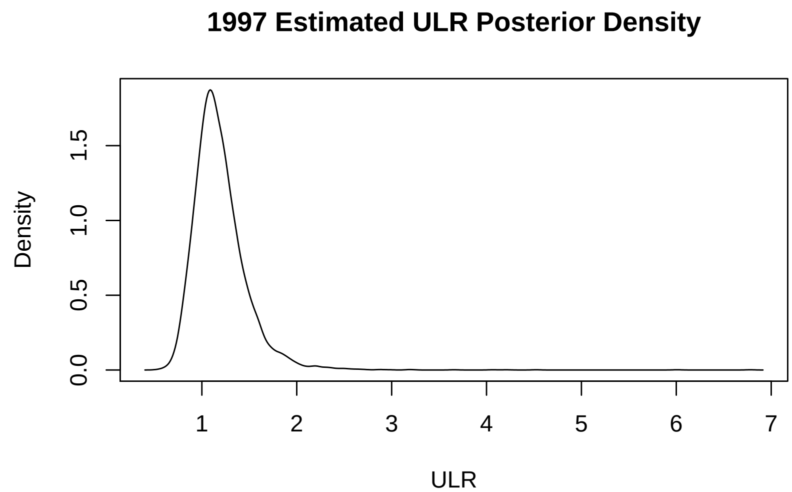
DIC	Outstanding	Paid
Bayesian Model 1	1031.0	879.6
Bayesian Model 2	1003.0	890.1

The DIC has decreased for the outstanding fits, indicating an improvement. However, it has increased for the paid fits which suggests that the model could be over-parameterized. There is an overall DIC reduction, and given the diagnostic plots a practitioner may select this model. In which case we will compare this model's incurred extrapolations against the hold-out samples as follows:



The fits more closely describe each individual year's incurred development relative to the previous model. Additionally, despite a number of years exhibiting over-reserving, the 1995 accident year fit assumes under-reserving on average (median $\widehat{RRF}_8 = 1.08$). This difference could be a feature of allowing all of the compartmental model parameters to vary by accident year according to a vague prior for $\boldsymbol{\Psi}$, enabling the model to place weight on the sharp incurred increase between development years 1 and 2. We could reduce the degree to which parameters vary across years (particularly less mature years where priors carry greater weight) by setting an informative variance-covariance prior.

The plots also show that the 1997 incurred density's mean is greater than its median. This is because the ULRs are assumed to be log-normally distributed (recall from Eq. (4.13) that $lRLR_i$ and $lRRF_i$ are assumed to be normally distributed).



The 95% PPI hold-out sample coverages (*with the previous model's stated in brackets*) are as follows:

95% PPI Coverage	1-year ahead	10-years ahead	Total
Outstanding	**100% (*89%*)**	100% (*100%*)	**100% (*93%*)**
Paid	**89% (*100%*)**	67% (*67%*)	**69% (*82%*)**
Incurred	100% (*100%*)	100% (*100%*)	**100% (*98%*)**

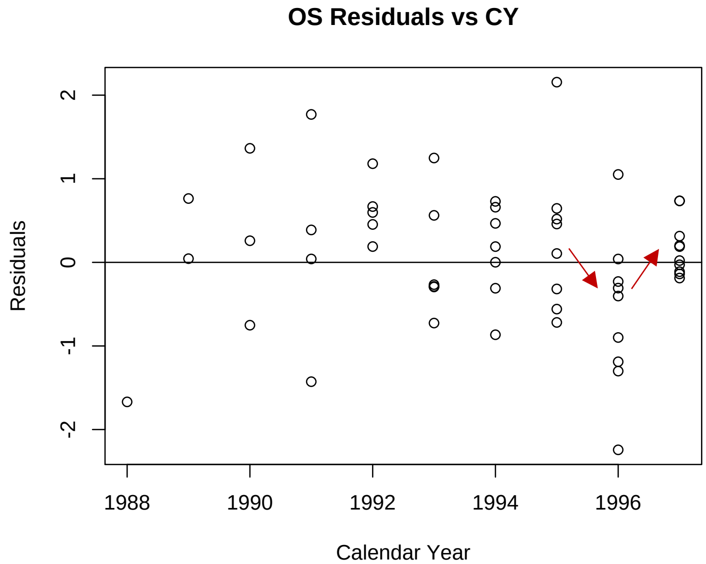
PPI 1-year ahead coverage has marginally improved for outstanding claims but worsened for paid claims. Outstanding and incurred claims coverages have improved to 100% across all time horizons. However, paid claims coverage has deteriorated owing to the model estimated (average) under-reserving for the 1995 accident year not materializing. As with the previous model's extrapolations, rate of payment reductions are not projected.

The OpenBUGS code for this model is contained in Appendix F.

The final scenario and area of model improvement that will be considered concerns the model fits by calendar year.

4.2.2 Scenario 2: Calendar shock sub-model

The outstanding claims residuals vs. calendar year plot for the previous model (shown below) displays a moderate downwards positional shift of residuals between the 1995 and 1996 calendar years, and upwards positional shift between 1996 and 1997.



This appears to be the result of outstanding claims exhibiting a step change at the same point in calendar time, perhaps due to a case reserve review during 1996. To capture the 1996 calendar shock within the model, we can define an indicator variable C_{ij} (for accident year $i = 1$ to 10 and development year $j = 1$ to $11 - i$) to mark the time before and after the calendar shock for each accident year:

$$C_{ij} = \begin{cases} 1, & i + j < 10 \\ 0, & i + j \geq 10 \end{cases} \tag{4.16}$$

To quantify the impact of the apparent case reserve review, we can then define an estimable proportional calendar shock impact variable, a_i, and restate OS_{ij} in Eq. (4.6):

$$OS_{ij} = f_{OS}(P_i, \boldsymbol{\phi}_i, t_j) \cdot (1 - C_{ij} \cdot a_i) + \varepsilon_{ij}^{OS} \tag{4.17}$$

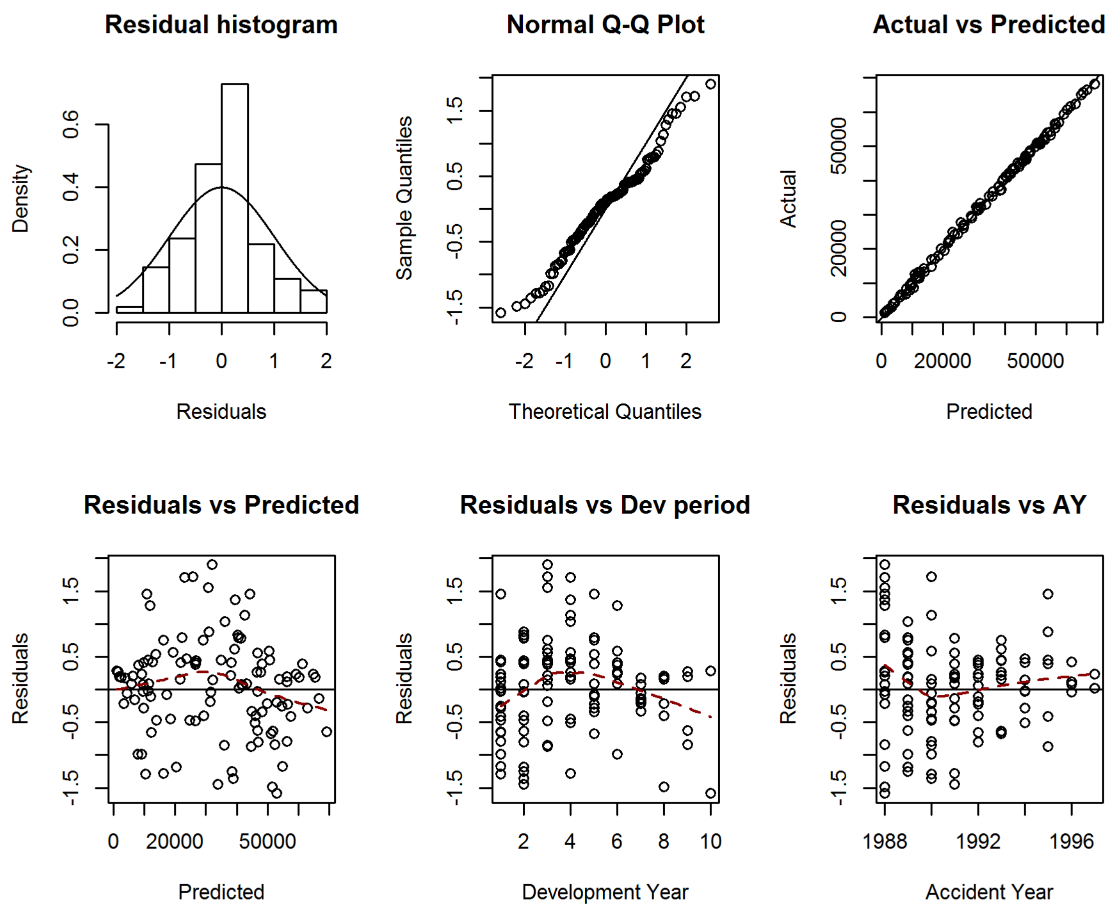
Therefore up until the end of the 1995 calendar year, the expected outstanding claims for accident year i and development year j are equal to $f_{OS}(P_i, \boldsymbol{\phi}_i, t_j)$ as before. However, once the review has taken place, outstanding claims are estimated to be $(1 - a_i)$% of their pre-shock values. Subsequent claims payments can be modeled to account for the shock as follows:

$$\begin{aligned} dEX/dt &= -\exp(l\beta_{er}) \cdot t \cdot EX \\ dOS/dt &= \exp(l\beta_{er}) \cdot t \cdot \exp(lRLR) \cdot EX - \exp(lk_p) \cdot OS \\ dPD/dt &= \exp(lk_p) \cdot \exp(lRRF) \cdot (1 - C_{ij} \cdot a_i) \cdot OS \end{aligned} \tag{4.18}$$

Upper and lower bounds for the estimated change in outstanding claims following the calendar shock have been set at 1% and 199% of the outstanding claims prior to the shock and assumed to be accident year independent. Candidate a_is are selected from a uniform distribution in the optimization:

$$a_i \sim U(-0.99, 0.99) \tag{4.19}$$

The remaining parameter priors, statistical assumptions and convergence arguments are unchanged from the previous model.



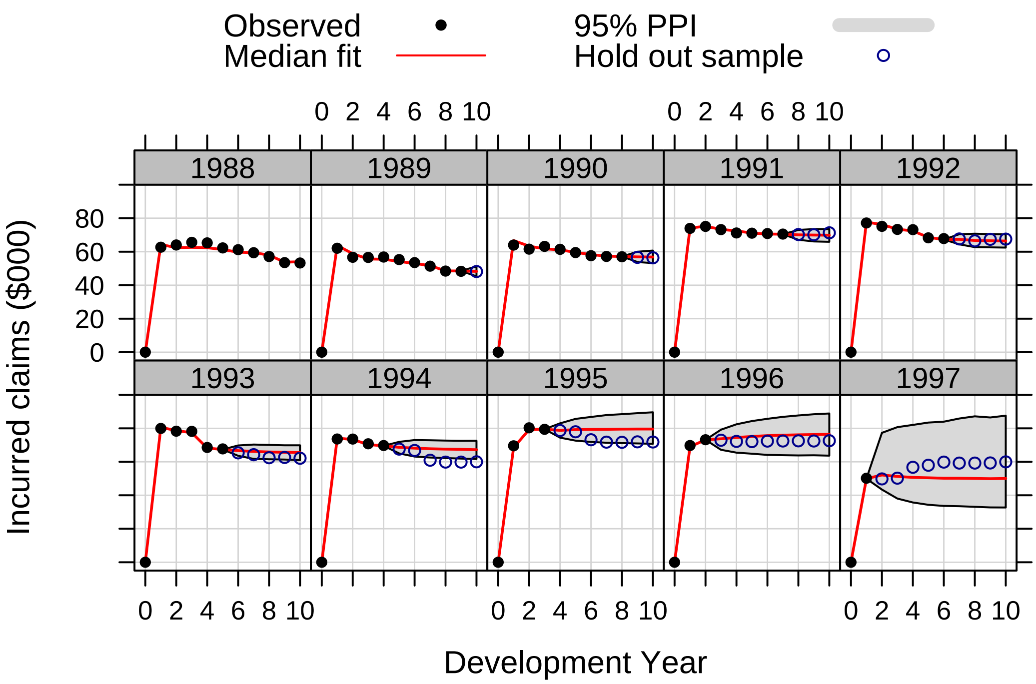
While the model fits are very close to the observations, there appears to be a serious violation of residual normality. The residual histogram shows that this model has far too many small magnitude residuals relative to mid-size residuals than expected under a standard normal distribution; an indication that claims are being *overfitted* by the model (similarly to the double-log model in Eq. (4.12)).

DIC	**Outstanding**	**Paid**
Bayesian Model 1	1031.0	879.6
Bayesian Model 2	1003.0	890.1
Bayesian Model 3	930.3	899.4

The DIC has substantially reduced for outstanding claims but has increased again for paid claims.

Combined with the above diagnostics, this suggests that the model is over-parameterized.

Although this model is not advisable for reserving purposes, its incurred extrapolations have been compared to the hold-out samples for illustrative purposes as follows:



The 95% PPI hold-out sample coverages (*with the previous model's stated in brackets*) are as follows:

95% PPI Coverage	**1-year ahead**	**10-years ahead**	**Total**
Outstanding	100% (*100%*)	100% (*100%*)	98% (*100%*)
Paid	89% (*89%*)	67% (*67%*)	**71% (*69%*)**
Incurred	100% (*100%*)	**89% (*100%*)**	**91% (*100%*)**

This model describes historical claims development more accurately than all previous models, yet incurred claims PPI coverage has reduced to its lowest level. It appears that explicitly modeling the outstanding claims calendar shock removes it from the modeled process error. Consequently, potential future calendar shocks are less likely to be adequately covered by the PPIs (such as the apparent 1999 shock affecting the 1994, 1995 and 1997 accident years).

The OpenBUGS code for this model is contained in Appendix F.

Although a more complex model could be built for future calendar shocks, this would not resolve the existing overfitting issue.

5. DISCUSSION

A hierarchical framework draws statistical strength across individuals, which can facilitate parsimony. However, as the case study demonstrates, this does not imply that the resultant model *will* be parsimonious. Diagnostic scrutiny is essential when selecting a hierarchical model for estimating reserves and their uncertainty.

Clark and Rangelova (2015) illustrate the importance of capturing accident year/development year interactions, and recommend that statistical methods allow intervention points for adjustment of intermediate results. In a hierarchical compartmental framework, an optional number of random-effects describe accident year development pattern differences based on intuitive parameters. The parameters themselves can be modeled to vary over development time. This flexibility allows the description of accident year/development year interactions such as changes in reporting/settlement rates and case reserve robustness, in addition to calendar shocks.

Although not demonstrated in the case study, continuous calendar trends such as inflation can be modeled within a compartmental framework. If a continuous "force of inflation" δ is assumed then expected claims payments $f_{PD}(P_i, \boldsymbol{\phi}_i, t_j)$ can be revised to include the inflation factor:

$$f'_{PD}(P_i, \boldsymbol{\phi}_i, t_j) = f_{PD}(P_i, \boldsymbol{\phi}_i, t_j) \cdot e^{(i+j-2)\delta} \tag{5.1}$$

As detailed by Zhang, Dukic and Guszcza (2012), the first calendar year $i + j = 2$ is treated as a "base" and subsequent expected calendar year payments are inflated by a factor $e^{(i+j-2)\delta}$, where δ is estimated or pre-specified. A similar approach could be taken to inflate outstanding claims or, alternatively, the differential equation system itself could be adjusted.

The deterministic compartmental model assumption of a smooth and detectable claims process relies upon claims cohort homogeneity for a volume of claims. This may not always be the case, and therefore further research is required to establish the validity and value of hierarchical semi-stochastic compartmental reserving models (Appendix B).

Other possible areas for future research include:

- The use of compartmental models to capture specific sub-processes such as legal shocks, catastrophes, latent claims, reopened claims, reinsurance recoveries and salvage/subrogation, to name but a few.
- Exploring the value of covariate models based on separate data sources. For example, if a claims handling team increased in size then one might expect the rate of payment to increase also.

- Establishment of a library of reporting/payment rate vs. development time functions along with their corresponding development profile properties.
- Simultaneous compartmental reserving for multiple insurance companies, e.g. by adding an extra level of hierarchy to describe company variation (Zhang, Dukic and Guszcza, 2012).

Many of the aforementioned extensions could be naturally incorporated within a Bayesian framework. Additionally, the Bayesian implementation itself could be further refined by considering alternative prior distributions. For example, prior dependence of random-effect variance and correlation terms could be controlled by using the separation strategy proposed by Barnard, McCulloch and Meng (2000).

Further work is required to evaluate the benefits of a compartmental approach compared to established methods, particularly for the estimation of reserve uncertainty.

6. CONCLUSIONS

This paper introduces a practical compartmental modeling framework for describing cumulative claims development. In particular, by considering the claims process over time as **Exposed to Risk → Claims Outstanding → Claims Paid**, an intuitive set of parameters have been defined which include a measure of case reserve robustness.

Cumulative paid claims model solutions are analogous to Clark's growth curve approach to loss reserving (2003). In contrast to growth curves which contain implicit tail factors, compartmental reserving model tail factors and hence ultimate projections are dictated by the extrapolation of outstanding losses to zero and estimated case reserve robustness. A number of possible model extensions have been explored to describe the nuances of the class of business being modeled, including changing reporting and/or settlement rates.

Following Guszcza (2008), a flexible nonlinear hierarchical framework is proposed to describe claims triangle data. Claims cohorts are viewed as individuals and cumulative losses are viewed as a series of observations for each individual. In contrast to Guszcza, cumulative paid triangles *and* outstanding claims triangles are fitted to, which enhances inference and interpretability. A probability sub-model allows a selection of the compartmental model parameters to vary by cohort and describe claims cohort pattern heterogeneity. Claims process trends can be identified and scenario tested, and parameter interpretability facilitates model discussion across the wider business.

A Bayesian implementation (similar to Zhang, Dukic and Guszcza, 2012) enables the robust incorporation of judgment and/or external information into claims projections. In addition to quantifying reserve uncertainty consistently with its definition, it offers additional model flexibility so that features such as residual autocorrelation and calendar effects can be explicitly accounted for.

Acknowledgments

I would like to thank Matt Locke for enduring numerous extremely useful conversations regarding the method and paper, Robert Ruiz for his advocacy, musings and mathematical rigor and Markus Gesmann for his valuable feedback and continual enthusiasm. I also thank Karl Goring and Matthew Killough for their insightful reviews. Additionally, I would like to thank Rob Murray for his encouragement and Lane Clark & Peacock for sponsoring the intermediate development of this research. Finally, I thank Jeff Courchene for approaching me to write this paper, and Nicki Power for graciously supporting several months' preoccupation. Any remaining errors are entirely my own.

Supplementary Material

A single-cohort compartmental reserving model illustration spreadsheet is available at http://www.casact.org/pubs/forum/16sforum/. The frequentist models in this paper were fitted using statistical software 'R', freely available at http://www.r-project.org. The R packages "nlmeODE" and "nlme" can be installed from the base R program. The Bayesian models in this paper were fitted using Bayesian Gibbs Sampling software 'OpenBUGS', freely available at http://www.openbugs.net. The case study dataset is freely available at http://www.casact.org/research/reserve_data/wkcomp_pos.csv (NAIC company code 337).

Appendix A: Implied development patterns

Implied continuous-time patterns of development are obtainable from the baseline compartmental reserving model solutions. Recall Eq. (3.2), which describes the claims process assuming that all exposure is input at time 0 and all model parameters are constant over development time t:

$$OS(t) = \frac{P \cdot RLR \cdot k_{er}}{k_{er} - k_p} \cdot \left(e^{-k_p t} - e^{-k_{er} t}\right)$$

$$PD(t) = \frac{P \cdot RLR \cdot RRF}{k_{er} - k_p} \cdot \left(k_{er} \cdot \left(1 - e^{-k_p t}\right) - k_p \cdot \left(1 - e^{-k_{er} t}\right)\right)$$

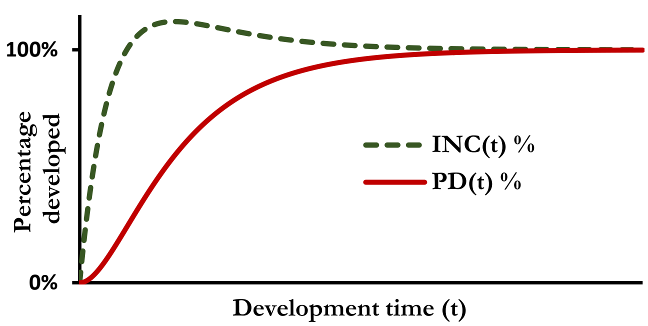
Given that $ULR = RLR \cdot RRF$, it follows that $P \cdot RLR \cdot RRF$ from the third ODE equates to the estimated ultimate losses. To derive the implied pattern of paid development at time t, we can therefore divide $PD(t)$ by $P \cdot RLR \cdot RRF$ to give

$$PD(t)\ \% = \frac{1}{k_{er} - k_p} \cdot \left(k_{er} \cdot \left(1 - e^{-k_p t}\right) - k_p \cdot \left(1 - e^{-k_{er} t}\right)\right) \quad \textbf{(A.1)}$$

Similarly, by summing $OS(t)$ and $PD(t)$, dividing by $P \cdot RLR \cdot RRF$ and simplifying, the implied incurred pattern of development can be derived as

$$INC(t)\ \% = \frac{k_{er} \cdot \left(e^{-k_p t} - e^{-k_{er} t}\right) + RRF \cdot \left(k_{er} \cdot \left(1 - e^{-k_p t}\right) - k_p \cdot \left(1 - e^{-k_{er} t}\right)\right)}{RRF \cdot \left(k_{er} - k_p\right)} \quad \textbf{(A.2)}$$

For a given set of parameters (with an $RRF < 1$), Eq. (A.1) and (A.2) can be visualized over development time as follows:



For perfect case reserving on average across a cohort of claims i.e. $RRF = 1$ (resulting in all claim amounts outstanding becoming paid claims), the incurred pattern in Eq. (A.2) simplifies and can be interpreted as an Exposed to Risk ("EtR") to reporting pattern:

$$INC(t)\ \% = EtR\ to\ Report(t)\ \% = 1 - e^{-k_{er} t} \quad \textbf{(A.3)}$$

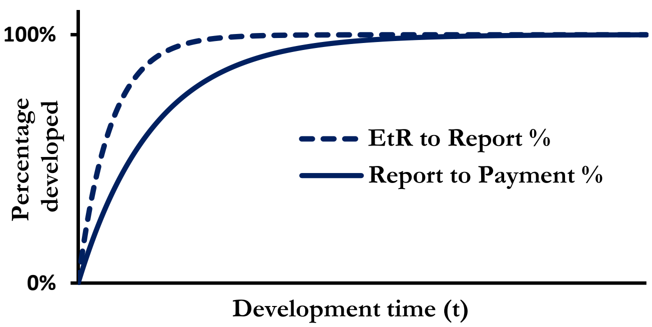
This result can also be obtained by letting $k_p \to 0$ in Eq. (3.2), and dividing $OS(t)$ by $P \cdot RLR$. To derive a report to payment pattern, it could be assumed that all exposures are initialized into the outstanding compartment at time 0. This results in a model that is defined in terms of two parameters only: a rate of payment and a reserve robustness factor. We can write the state-variable solutions as

$$OS(t) = Pe^{-k_p t}$$
$$PD(t) = P \cdot RRF \cdot \left(1 - e^{-k_p t}\right) \quad \textbf{(A.4)}$$

Similarly to above, the payment pattern $PD(t)\,\%$ can be derived by dividing $PD(t)$ by ultimate claims, which in this instance is $P \cdot RRF$. Given that we are only considering the claims process from reporting onwards, the resulting pattern can be interpreted as a report to payment pattern:

$$PD(t)\ \% = Report\ to\ Payment(t)\ \% = \ 1 - e^{-k_p t} \quad \textbf{(A.5)}$$

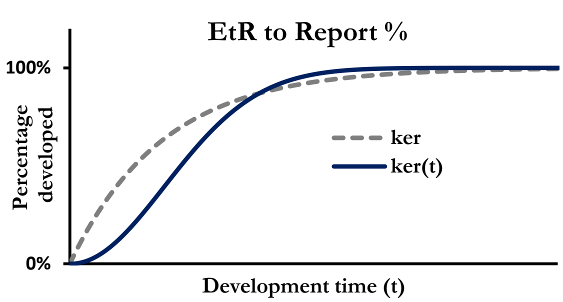
This result can also be obtained by letting $k_{er} \to \infty$ in Eq. (3.2), and dividing $PD(t)$ by $P \cdot RRF \cdot RLR$. For a given set of parameters, the EtR to report and report to payment development patterns can be visualized over development time as follows:

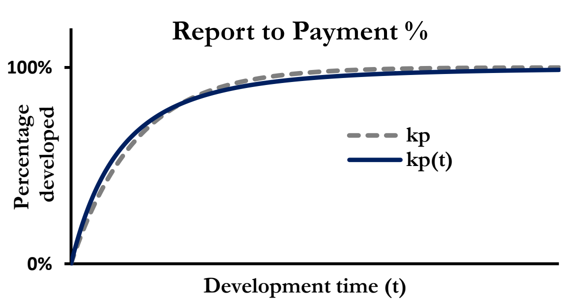


The development patterns are based on rate parameters which are constant over development time. If the rate parameters varied over time however, development patterns would also be expected to vary. Equations (A.3) and (A.5) can be generalized to allow for variable rates by writing

$$EtR\ to\ Report(t)\ \% = 1 - e^{-\int_0^t \mathrm{k}_{er}(t)dt}$$
$$Report\ to\ Payment(t)\ \% = 1 - e^{-\int_0^t \mathrm{k}_{p}(t)dt} \quad \textbf{(A.6)}$$

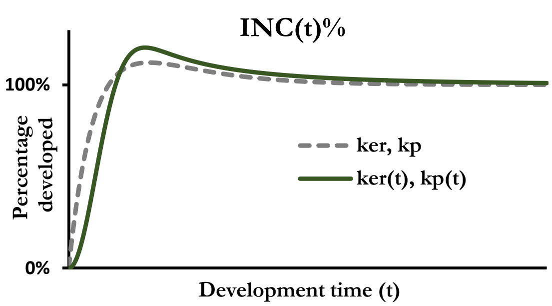
The graphs below show implied EtR to report and report to payment development patterns both for constant rate parameters (dashed lines), and parameters that vary over development time in accordance with the functions outlined in Section 3.2 (solid lines):

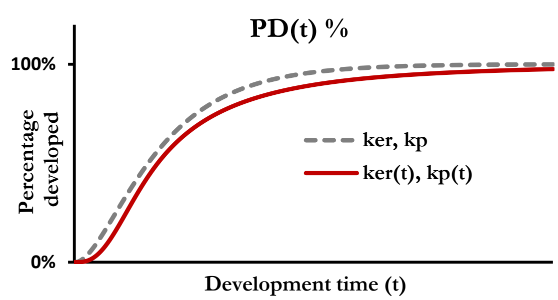




In contrast to a constant rate of reporting, a reporting rate that linearly increases over development time results in a slower pattern of reported claims development initially, which speeds up over time. This could be used to reflect a delay between claim events and claim reports for an accident cohort of claims. Allowing the rate of payment to decrease over development time results in a faster pattern of payment initially, which slows down over time. This is reflective of a slower settlement rate for claims outstanding in later development periods, perhaps due to litigation.

Corresponding incurred and payment patterns for both constant and non-constant rate parameters (obtained using numerical methods) can also be compared as follows:





The impact of altering parameters values/functions on development patterns can be seen in the illustration spreadsheet available at: http://www.casact.org/pubs/forum/16sforum/.

Appendix B: Semi-stochastic compartmental reserving models

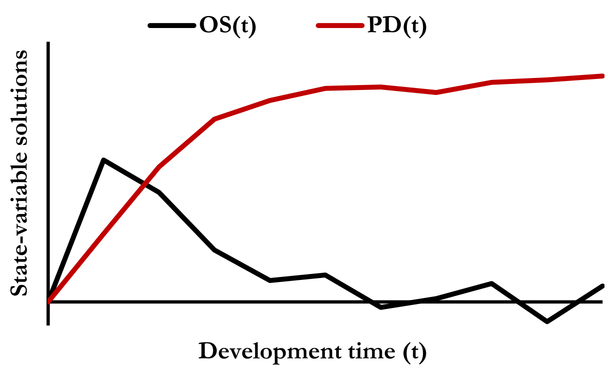
The deterministic compartmental model outlined in Section 3 assumes the same average claims behavior throughout the lifetime of a cohort. However, there are many reasons why there may be additional variability in the process, e.g. erratic case reserve fluctuations, claims payment backlogs etc. It may therefore be appropriate to re-specify the baseline model as a semi-stochastic (or "grey box"; Tornøe *et al.*, 2004b) model by introducing a Wiener process (or multiple processes) into the model's structural form. To do this we must first re-write Eq. (3.1) by moving the time increment (dt) terms to the right hand side of the ODEs, giving

$$\begin{aligned} dEX &= (-k_{er} \cdot EX)dt \\ dOS &= (k_{er} \cdot RLR \cdot EX - k_p \cdot OS)dt \\ dPD &= (k_p \cdot RRF \cdot OS)dt \end{aligned} \qquad \textbf{(B.1)}$$

To incorporate a Wiener process for outstanding claims we can write

$$\begin{aligned} dEX &= (-k_{er} \cdot EX)dt \\ dOS &= \left(k_{er} \cdot RLR \cdot EX - k_p \cdot OS\right)dt + \sigma_{OS} dW \\ dPD &= (k_p \cdot RRF \cdot OS)dt \end{aligned} \qquad \textbf{(B.2)}$$

where W is a standard (and additive) Wiener process such that $W(t_2) - W(t_1) \sim N(0, |t_2 - t_1|)$, and σ_{OS} is the estimable element of the Wiener process standard deviation (the diffusion coefficient), representing volatility in outstanding claims not captured by the deterministic ODEs. For illustration, this allows model solutions (plotted at yearly time steps) to look as follows:



An issue with the model outlined above is that the volatility in outstanding claims is assumed to be constant, and therefore the Wiener process can cause outstanding claims to fall below zero. Although this is plausible for classes of business where salvage/subrogation is material, Eq. (B.2) assumes that large outstanding claims fluctuations can persist at later development times where they would typically be expected to be zero. This can lead to negative paid increments, as shown above. To address this, the above Wiener process can be assumed to be a multiple of the amount in the outstanding claims compartment (i.e. state-dependent), giving

$$\begin{aligned} dEX &= (-k_{er} \cdot EX)dt \\ dOS &= \left(k_{er} \cdot RLR \cdot EX - k_p \cdot OS\right)dt + \sigma_{OS} OS dW \\ dPD &= (k_p \cdot RRF \cdot OS)dt \end{aligned} \qquad \textbf{(B.3)}$$

The volatility introduced to the claims process is therefore proportional to amounts outstanding at each development time, which may be a more realistic assumption. Although this model can be fitted to a single cohort, for the multiple cohort case using hierarchical models (Section 4) it is not straightforward to implement Eq. (B.3) in conventional software (at the time of writing). However, Eq. (B.2) can be implemented in a hierarchical framework using the R package "PSM" (Klim *et al.*, 2009).

A key benefit of using SDEs is that they can account for residual autocorrelation (see Section 4.2) in a flexible manner. Furthermore, SDEs can describe claims process mechanisms that are too complex to include in the structural model (Overgaard *et al.*, 2005). A similar approach could be used to model low-frequency high-severity losses. As an alternative to the semi-stochastic model above, probability transfer mechanisms between compartments could be incorporated (Rescigno and Segre, 1966).

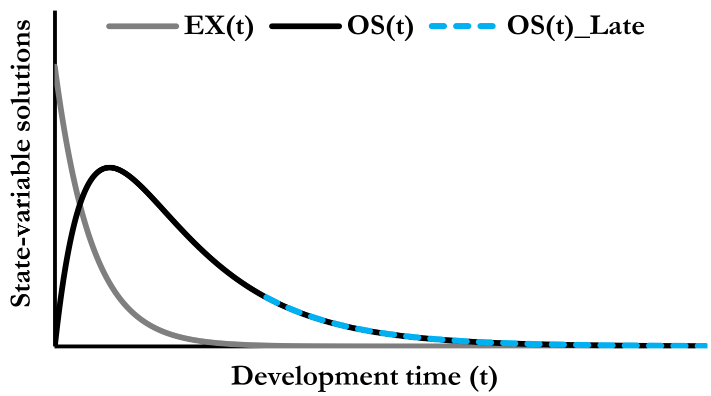
Appendix C: Nonlinear regression self-starting algorithm

Nonlinear regression models require parameter starting values for optimization to take place. Although visual inspection and judgment can be used to select reasonable starting values (see Section 4.1), inappropriate estimates can result in the model converging to a local rather than global likelihood maximum. A starting value algorithm is therefore outlined below for a single-cohort baseline compartmental reserving model, based on the "method of residuals" (Macheras, 1987).

We reexamine the baseline compartmental model defined by Eq. (3.1) and (3.2) and note that by some development time point, most claims will have been reported i.e. $EX(t) \to 0$. From this point onwards, only the claims payment phase of the process will remain. Provided that k_{er} is sufficiently larger than k_p, we can ignore the reporting term $e^{-k_{er}t}$ and obtain the following expression for later development time outstanding claims, $OS(t)^{LATE}$:

$$OS(t)^{LATE} = \frac{P \cdot RLR \cdot k_{er}}{k_{er} - k_p} \cdot e^{-k_p t} \qquad \textbf{(C.1)}$$

This can be viewed graphically as follows:



Denoting $\boldsymbol{\beta} = \{\beta_1, \beta_2, \beta_3, \beta_4\}^T = \{k_{er}, RLR, k_p, RRF\}^T$ and OS_j as the jth outstanding claims observation, we can write down the following regression model:

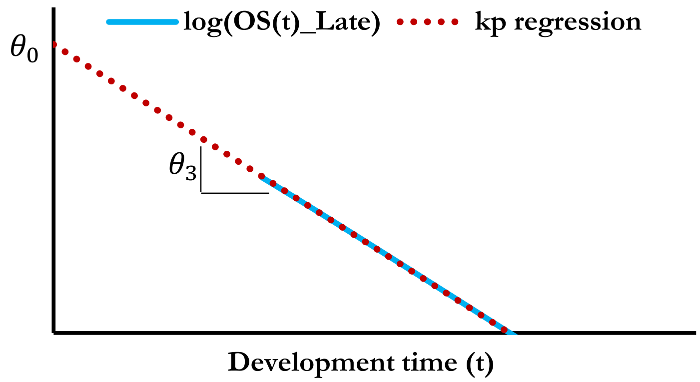
$$OS_j^{LATE} = \frac{P \cdot \beta_2 \cdot \beta_1}{\beta_1 - \beta_3} \cdot e^{-\beta_3 t_j} + \varepsilon_j^{OS} \qquad \textbf{(C.2)}$$

This phase of the solution has only one exponential term, enabling us to take logarithms of both sides to linearize the model:

$$\log(OS_j^{LATE}) = \log\left(\frac{P \cdot \beta_2 \cdot \beta_1}{\beta_1 - \beta_3}\right) - \beta_3 t_j + \epsilon_j^{OS} \qquad \textbf{(C.3)}$$

Denoting $\theta_0 = \log\left(\frac{P \cdot \beta_2 \cdot \beta_1}{\beta_1 - \beta_3}\right)$, $\theta_3 = -\beta_3$, a linear regression can be specified and carried out:

$$\log(OS_j^{LATE}) = \theta_0 + \theta_3 t_j + \epsilon_j^{OS} \qquad \textbf{(C.4)}$$



This regression should be carried out for the logarithm of outstanding claims development values from the point at which the exposure is assumed to be negligible. However, this time point is not likely to be known. Even if it was, there may be practical restrictions to carrying out regression C.4 and subsequent regressions from this time point onwards (discussed at the end of this Appendix).

Once estimates $\hat{\theta}_0$ and $\hat{\theta}_3$ have been found, we establish that $\widehat{\frac{P \cdot \beta_2 \cdot \beta_1}{\beta_1 - \beta_3}} = e^{\hat{\theta}_0}$ and $\hat{\beta}_3 = -\hat{\theta}_3$.

This gives an estimate of the rate of payment, k_p. The next step is to identify that

$$OS_j = OS_j{}^{LATE} - \frac{P \cdot \beta_2 \cdot \beta_1}{\beta_1 - \beta_3} \cdot \left(e^{-\beta_1 t_j}\right) \tag{C.5}$$

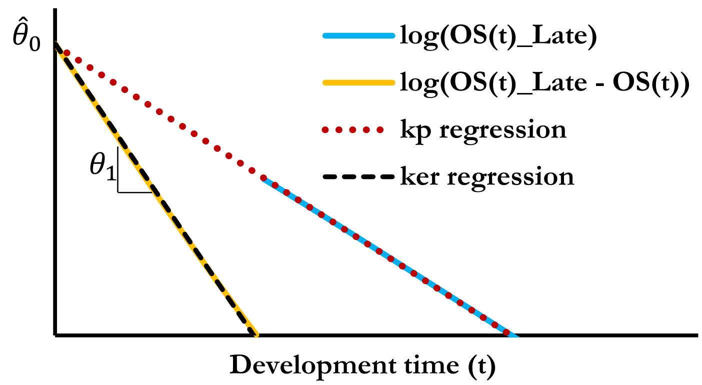
This can be rearranged and linearized as follows for $OS_j{}^{LATE} - OS_j > 0$:

$$\begin{aligned} OS_j - OS_j{}^{LATE} &= -\frac{P \cdot \beta_2 \cdot \beta_1}{\beta_1 - \beta_3} \cdot \left(e^{-\beta_1 t_j}\right) \\ \log(OS_j{}^{LATE} - OS_j)\big|_{OS_j{}^{LATE} - OS_j > 0} &= \log\left(\frac{P \cdot \beta_2 \cdot \beta_1}{\beta_1 - \beta_3}\right) - \beta_1 t_j \end{aligned} \tag{C.6}$$

$OS_j{}^{LATE}$ can be taken as its estimated value in the previous regression, $\widehat{OS}_j{}^{LATE}$, and the intercept $\log\left(\frac{P \cdot \beta_2 \cdot \beta_1}{\beta_1 - \beta_3}\right)$ can be fixed to the previously estimated intercept, $\hat{\theta}_0$.

Denoting $\theta_1 = -\beta_1$ and rearranging, a second linear regression can be specified through the origin (Turner, 1960):

$$\begin{aligned} \log(\widehat{OS}_j{}^{LATE} - OS_j)\big|_{\widehat{OS}_j{}^{LATE} - OS_j > 0} &= \hat{\theta}_0 + \theta_1 t_j + \xi_j \\ \log(\widehat{OS}_j{}^{LATE} - OS_j)\big|_{\widehat{OS}_j{}^{LATE} - OS_j > 0} - \hat{\theta}_0 &= \theta_1 t_j + \xi_j \end{aligned} \tag{C.7}$$



Once an estimate of $\hat{\theta}_1$ of θ_1 has been found, we establish that $\hat{\beta}_1 = -\hat{\theta}_1$, thus providing an estimate of the rate of reporting, k_{er}. Given our estimates of k_{er} and k_p, we can infer an estimate of the RLR. To see how, we recall the definition of θ_0 in Eq. (C.4) and rearrange as follows:

$$\log\left(\frac{P \cdot \beta_2 \cdot \beta_1}{\beta_1 - \beta_3}\right) = \theta_0$$

$$\beta_2 = \frac{e^{\theta_0} \cdot (\beta_1 - \beta_3)}{P \cdot \beta_1} \tag{C.8}$$

$$\beta_2 = \frac{e^{\theta_0} \cdot (-\theta_1 + \theta_3)}{P \cdot -\theta_1}$$

We can therefore substitute in the previously estimated parameters to get an estimate of β_2:

$$\hat{\beta}_2 = \frac{e^{\hat{\theta}_0} \cdot (-\hat{\theta}_1 + \hat{\theta}_3)}{P \cdot -\hat{\theta}_1} \tag{C.9}$$

This is an estimate of the RLR. Finally, to estimate the RRF we note that the above procedure generates parameter estimates for all elements of the paid claims solution in Eq. (3.2) except the RRF:

$$PD(t) = \frac{P \cdot RLR \cdot RRF}{k_{er} - k_p} \cdot \left(k_{er} \cdot \left(1 - e^{-k_p t}\right) - k_p \cdot \left(1 - e^{-k_{er} t}\right)\right)$$
$$PD(t) = \frac{P \cdot RLR}{k_{er} - k_p} \cdot \left(k_{er} \cdot \left(1 - e^{-k_p t}\right) - k_p \cdot \left(1 - e^{-k_{er} t}\right)\right) \cdot RRF \tag{C.10}$$

Rewriting as a regression as per above gives:

$$PD_j = \frac{P \cdot \beta_2}{\beta_1 - \beta_3} \cdot \left(\beta_1 \cdot \left(1 - e^{-\beta_3 t_j}\right) - \beta_3 \cdot \left(1 - e^{-\beta_1 t_j}\right)\right) \cdot \beta_4 + \omega_j \tag{C.11}$$

Substituting in the estimates of each parameter apart from β_4, we can denote $\theta_4 = \beta_4$ and rewrite Eq. (C.11) as follows:

$$PD_j = f\left(P, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, t_j\right) \cdot \theta_4 + \omega_j \tag{C.12}$$

This regression is linear in the parameter that we are interested in estimating, θ_4 (RRF), and therefore a linear regression through the origin can be carried out to derive an estimate of θ_4: $\hat{\theta}_4$.

The vector of parameter starting values, $\boldsymbol{\beta}^0$, can then be set to be:

$$\boldsymbol{\beta}^0 = \{\beta_1^0, \beta_2^0, \beta_3^0, \beta_4^0\}^T = \left\{k_{er}^0, RLR^0, k_p^0, RRF^0\right\}^T = \left\{-\hat{\theta}_1, \hat{\theta}_2, -\hat{\theta}_3, \hat{\theta}_4\right\}^T \qquad \textbf{(C.13)}$$

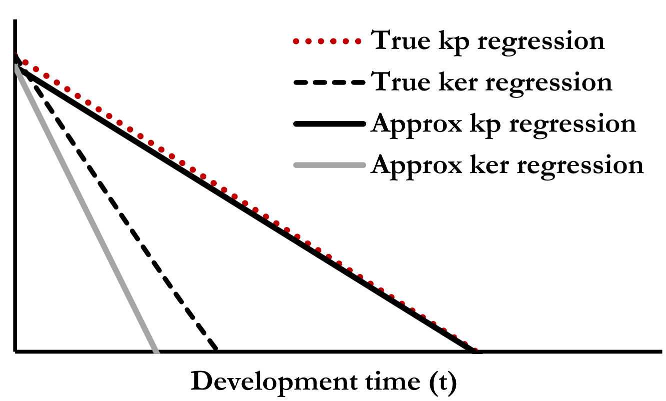
Although this algorithm is based on a single cohort of claims, for the multiple cohort case (e.g. using hierarchical models as outlined in Section 4), the algorithm could be used to derive parameter estimates for each individual cohort. To derive fixed-effect starting values, one could then calculate weighted average parameters based on the number of data points within each cohort, for example.

Selecting $\boldsymbol{t_j}^{\boldsymbol{LATE}}$

As stated above, a practitioner is unlikely to be able to identify when exposure has fallen close to zero for a particular claims cohort. Furthermore, a claims cohort might not have a long enough development history to be able to fit a regression from $t_j{}^{LATE}$ onwards. This issue is more prevalent if the rate of reporting is slow because by definition, exposures will convert to reported claims and tend to zero at a slower rate.

Being that the goal is to specify starting values for the parameters being estimated and not to derive final estimates, it may be acceptable to compromise on the point at which $t_j{}^{LATE}$ is defined at the cost of reducing the accuracy of the initial parameter estimates. One possibility is to calculate peak outstanding claims from the data, MAX_OS, and define the corresponding development time point as $t_j{}^{MAX\_OS}$. For the regressions outlined above, $t_j{}^{LATE}$ could then be defined as $t_j \geq t_j{}^{MAX\_OS}$.

In some instances (e.g. when k_{er} much faster than k_p) this will be a close approximation to when exposure is close to zero. In others however, it is likely to be a less accurate approximation due to a high probability of new non-negligible value claims being subsequently reported. The graph below illustrates the discrepancy in regression slopes, i.e. initial parameter estimates, for $k_{er} = 2.33k_p$:



This approach to defining ${t_j}^{LATE}$ will result in a degree of starting value parameter estimation error (more predominantly for k_{er}), with a magnitude inversely proportional to the underlying rate of reporting. On the other hand, if k_{er} is too fast, there won't be early phase data to derive its estimate in the first place (in which case a large estimate can be selected arbitrarily). Additionally, the less mature the cohort, the less reliable the parameter estimates will be. However, this approach should initialize the nonlinear regression optimization process at a sensible point in the parameter space.

It's worth noting that ${t_j}^{MAX\_OS}$ may appear long before the true payment phase if outstanding claims development is volatile. Therefore in practice, judgment will be necessary to decide from which development time point the observed logarithm of the outstanding claims can be considered linear. The degree of linearity must be balanced with the number of development data points available to carry out the regression for k_p. In cases where there are no observations subsequent to the maximum outstanding claims value, this algorithm cannot be used.

In the case of development time-dependent parameters (Section 3.2), the parameter starting value algorithm could be used to find approximate parameter starting values by setting nonlinear rate *functions* equal to the parameter estimates above. However, identifiability will be an issue for rate functions with more than one parameter (unless at least one of the parameters is arbitrarily fixed).

Appendix D: Frequentist case study data

Data key

"Cohort" = accident year
"t" = development year
"Claims" = outstanding claims for "Type"=1 and cumulative paid claims for "Type"=2
"Dose" = exposure/earned premium
"Cmt" = exposure compartment number

```
> Data <- groupedData(Claims ~ t | Cohort/Type, data = Data)

Grouped Data: Claims ~ t | Cohort/Type
    Cohort  t Claims Type   Dose Cmt
1     1988  0      0    1 104437   1
2     1988  0      0    2      0   1
3     1988  1  53121    1      0   1
4     1988  1   9558    2      0   1
5     1988  2  41222    1      0   1
6     1988  2  22778    2      0   1
7     1988  3  32309    1      0   1
8     1988  3  33298    2      0   1
9     1988  4  24944    1      0   1
10    1988  4  40348    2      0   1
11    1988  5  17104    1      0   1
12    1988  5  45146    2      0   1
13    1988  6  13137    1      0   1
14    1988  6  48048    2      0   1
15    1988  7   9605    1      0   1
16    1988  7  49782    2      0   1
17    1988  8   6515    1      0   1
18    1988  8  50623    2      0   1
19    1988  9   1661    1      0   1
20    1988  9  51812    2      0   1
21    1988 10   1322    1      0   1
22    1988 10  51939    2      0   1
23    1989  0      0    1  88883   1
24    1989  0      0    2      0   1
25    1989  1  54145    1      0   1
26    1989  1   7913    2      0   1
27    1989  2  37188    1      0   1
28    1989  2  19472    2      0   1
29    1989  3  26976    1      0   1
30    1989  3  29622    2      0   1
31    1989  4  20015    1      0   1
32    1989  4  36816    2      0   1
33    1989  5  14319    1      0   1
34    1989  5  40975    2      0   1
35    1989  6  10179    1      0   1
36    1989  6  43302    2      0   1
37    1989  7   6672    1      0   1
38    1989  7  44707    2      0   1
39    1989  8   2575    1      0   1
40    1989  8  45871    2      0   1
41    1989  9   2071    1      0   1
```

42	1989	9	46229	2	0	1
43	1990	0	0	1	85956	1
44	1990	0	0	2	0	1
45	1990	1	55211	1	0	1
46	1990	1	8744	2	0	1
47	1990	2	37221	1	0	1
48	1990	2	24302	2	0	1
49	1990	3	27760	1	0	1
50	1990	3	35406	2	0	1
51	1990	4	17990	1	0	1
52	1990	4	43412	2	0	1
53	1990	5	11417	1	0	1
54	1990	5	48057	2	0	1
55	1990	6	6716	1	0	1
56	1990	6	50897	2	0	1
57	1990	7	4282	1	0	1
58	1990	7	52879	2	0	1
59	1990	8	3015	1	0	1
60	1990	8	53956	2	0	1
61	1991	0	0	1	99339	1
62	1991	0	0	2	0	1
63	1991	1	60617	1	0	1
64	1991	1	13301	2	0	1
65	1991	2	42144	1	0	1
66	1991	2	32950	2	0	1
67	1991	3	25987	1	0	1
68	1991	3	47201	2	0	1
69	1991	4	14805	1	0	1
70	1991	4	56394	2	0	1
71	1991	5	9406	1	0	1
72	1991	5	61650	2	0	1
73	1991	6	5792	1	0	1
74	1991	6	65039	2	0	1
75	1991	7	3966	1	0	1
76	1991	7	66566	2	0	1
77	1992	0	0	1	104897	1
78	1992	0	0	2	0	1
79	1992	1	65719	1	0	1
80	1992	1	11424	2	0	1
81	1992	2	46047	1	0	1
82	1992	2	29086	2	0	1
83	1992	3	31250	1	0	1
84	1992	3	42034	2	0	1
85	1992	4	22245	1	0	1
86	1992	4	50910	2	0	1
87	1992	5	11878	1	0	1
88	1992	5	56406	2	0	1
89	1992	6	8408	1	0	1
90	1992	6	59437	2	0	1
91	1993	0	0	1	119427	1
92	1993	0	0	2	0	1
93	1993	1	68133	1	0	1
94	1993	1	11792	2	0	1
95	1993	2	51102	1	0	1
96	1993	2	27161	2	0	1
97	1993	3	39934	1	0	1

98	1993	3	38229	2	0	1
99	1993	4	21824	1	0	1
100	1993	4	46722	2	0	1
101	1993	5	16955	1	0	1
102	1993	5	50742	2	0	1
103	1994	0	0	1	110784	1
104	1994	0	0	2	0	1
105	1994	1	62434	1	0	1
106	1994	1	11194	2	0	1
107	1994	2	46661	1	0	1
108	1994	2	26893	2	0	1
109	1994	3	32248	1	0	1
110	1994	3	38488	2	0	1
111	1994	4	24140	1	0	1
112	1994	4	45580	2	0	1
113	1995	0	0	1	77731	1
114	1995	0	0	2	0	1
115	1995	1	56971	1	0	1
116	1995	1	12550	2	0	1
117	1995	2	48677	1	0	1
118	1995	2	31604	2	0	1
119	1995	3	35336	1	0	1
120	1995	3	44045	2	0	1
121	1996	0	0	1	63646	1
122	1996	0	0	2	0	1
123	1996	1	56526	1	0	1
124	1996	1	13194	2	0	1
125	1996	2	41707	1	0	1
126	1996	2	31474	2	0	1
127	1997	0	0	1	48052	1
128	1997	0	0	2	0	1
129	1997	1	40799	1	0	1
130	1997	1	9372	2	0	1

Appendix E: Frequentist modeling R code

Baseline structural model (Section 4.1)

```
> DEmodel <- list(
+                 DiffEq=list(
+                     dy1dt = ~ -lker*y1,
+                     dy2dt = ~ lker*lRLR*y1 - lkp*y2,
+                     dy3dt = ~ lkp*lRRF*y2),
+                 ObsEq=list(
+                     EX   = ~ 0,
+                     OS   = ~ y2,
+                     PA   = ~ y3),
+ States=c("y1","y2","y3"),
+ Parms=c("lker","lRLR","lkp","lRRF"),
+ Init=list(0,0,0))
```

Model 1 (Section 4.1)

```
> ReservingModel <- nlmeODE(DEmodel,Data)   ### "Data" = data in Appendix D

> nlmeModel <- nlme(Claims ~
ReservingModel(lker,lRLR,lkp,lRRF,t,Cohort,Type),

+ data = Data,
+ fixed = lker+lRLR+lkp+lRRF ~ 1,           ### fixed-effect parameters
+ random = pdDiag(lRLR + lRRF ~ 1),         ### parameters with random-effects
+ groups = ~Cohort,                         ### data grouping (accident years)
+ weights = varIdent(form = ~1 | Type),     ### residual error functions: OS&PD
+ start = c(lker = log(1.5), lRLR = log(1),
     lkp = log(0.75), lRRF = log(0.75)),    ### parameter starting values
+ control=list(returnObject=TRUE,msVerbose=TRUE,
+ msMaxIter=10000,pnlsMaxIter=10000,
+ pnlsTol=0.4),                             ### tolerance for PNLS convergence
+ verbose=TRUE)

> nlmeModel
Nonlinear mixed-effects model fit by maximum likelihood
 Model: Claims ~ ReservingModel(lker, lRLR, lkp, lRRF, t, Cohort, Type)
 Data: Data
 Log-likelihood: -1164.386
 Fixed: lker + lRLR + lkp + lRRF ~ 1
      lker        lRLR         lkp        lRRF
 0.40824328  0.02575157 -0.79246675 -0.40644353
                                            ### estimated fixed-effects: β̂
Random effects:
 Formula: list(lRLR ~ 1, lRRF ~ 1)
 Level: Cohort
 Structure: Diagonal
            lRLR      lRRF Residual
StdDev: 0.1870103 0.1318661 3171.213
                                            ### estimated random-effect &
Variance function:                          ### residual std dev terms: {ψ̂_ik, σ̂}
 Structure: Different standard deviations per stratum
 Formula: ~1 | Type
 Parameter estimates:
       1         2
1.0000000 0.1790677                         ### OS&PD residual std deviation
Number of Observations: 130                 ### multipliers: {1, λ̂}
Number of Groups: 10
```

Extended structural model – development time-dependent reporting rate (Section 4.1.1)

```
> DEmodel2 <- list(
+                 DiffEq=list(
+                     dy1dt = ~ -lBer*t*y1,
+                     dy2dt = ~ lBer*t*lRLR*y1 - lkp*y2,
+                     dy3dt = ~ lkp*lRRF*y2),
+                 ObsEq=list(
+                     EX   = ~ 0,
+                     OS   = ~ y2,
+                     PA   = ~ y3),
+ States=c("y1","y2","y3"),
+ Parms=c("lBer","lRLR","lkp","lRRF"),
+ Init=list(0,0,0))
```

Model 2 (Section 4.1.1)

```
> ReservingModel2 <- nlmeODE(DEmodel2,Data)

> nlmeModel2 <- nlme(Claims ~
ReservingModel2(lBer,lRLR,lkp,lRRF,t,Cohort,Type),
+ data = Data,
+ fixed = lBer+lRLR+lkp+lRRF ~ 1,
+ random = pdDiag(lRLR + lRRF ~ 1),
+ groups = ~Cohort,
+ weights = varIdent(form = ~1 | Type),
+ start=c(lBer = log(5), lRLR = log(1.03),
     lkp = log(0.45), lRRF = log(0.67)),
+ control=list(returnObject=TRUE,msVerbose=TRUE,
+ msMaxIter=10000,pnlsMaxIter=10000,
+ pnlsTol=0.4),
+ verbose=TRUE)

> nlmeModel2
Nonlinear mixed-effects model fit by maximum likelihood
 Model: Claims ~ ReservingModel2(lBer, lRLR, lkp, lRRF, t, Cohort, Type)
 Data: Data
 Log-likelihood: -1156.344
 Fixed: lBer + lRLR + lkp + lRRF ~ 1
      lBer       lRLR        lkp       lRRF
 1.7637739 -0.1608870 -0.9339032 -0.1886841

Random effects:
 Formula: list(lRLR ~ 1, lRRF ~ 1)
 Level: Cohort
 Structure: Diagonal
             lRLR      lRRF Residual
StdDev: 0.1684008 0.1469151 2491.433

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | Type
 Parameter estimates:
        1         2
1.0000000 0.2509692
Number of Observations: 130
Number of Groups: 10
```

```
> anova(nlmeModel,nlmeModel2)
           Model df      AIC      BIC    logLik
nlmeModel      1  8 2344.771 2367.711 -1164.386
nlmeModel2     2  8 2328.688 2351.628 -1156.344

> intervals(nlmeModel2)
Approximate 95% confidence intervals

 Fixed effects:
         lower       est.       upper
lBer  1.6599011  1.7637739  1.86764663
lRLR -0.2696058 -0.1608870 -0.05216819
lkp  -0.9671036 -0.9339032 -0.90070283
lRRF -0.2873897 -0.1886841 -0.08997844

> summary(nlmeModel2)

Correlation:
     lBer   lRLR   lkp
lRLR -0.110
lkp  -0.723  0.143
lRRF  0.142 -0.077 -0.253
```

Model 3 – random-effects correlation (Section 4.1.2)

```
> nlmeModel3 <- update(nlmeModel2,random=list(lRLR+lRRF~1))

> intervals(nlmeModel3)
Approximate 95% confidence intervals

Random Effects:
 Level: Cohort
                     lower      est.     upper
sd(lRLR)       0.09864002 0.1571791 0.2504587
sd(lRRF)       0.09475350 0.1517442 0.2430128
cor(lRLR,lRRF) 0.34584978 0.7795638 0.9387946

> anova(nlmeModel,nlmeModel2,nlmeModel3)
           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
nlmeModel      1  8 2344.771 2367.711 -1164.386
nlmeModel2     2  8 2328.688 2351.628 -1156.344
nlmeModel3     3  9 2324.543 2350.351 -1153.272 2 vs 3 6.144368  0.0132
```

Model 4 – block-diagonal random-effects structure (Section 4.1.2)

```
> nlmeModel4 <- update(nlmeModel3,random=pdBlocked(list(lRLR + lRRF~1, lkp
~ 1)))

> anova(nlmeModel,nlmeModel2,nlmeModel3,nlmeModel4)
           Model df      AIC      BIC    logLik   Test   L.Ratio p-value
nlmeModel      1  8 2344.771 2367.711 -1164.386
nlmeModel2     2  8 2328.688 2351.628 -1156.344
nlmeModel3     3  9 2324.543 2350.351 -1153.272 2 vs 3  6.144368  0.0132
nlmeModel4     4 10 2305.500 2334.175 -1142.750 3 vs 4 21.043472  <.0001
```

Supplementary code – for structural model ***x***, hierarchical model ***y***

```
> residuals(nlmeModely, type="normalized")  ### standardized model residuals

> fitted(nlmeModely)                        ### model predictions

> IndCoef <- coef(nlmeModely)               ### individual accident year (log)
                                            compartmental parameter estimates

> ReservingModelx(                          ### model projections to time 10
+ rep(IndCoef[,1],each=2*11),
+ rep(IndCoef[,2],each=2*11),
+ rep(IndCoef[,3],each=2*11),
+ rep(IndCoef[,4],each=2*11),
+ Data_Full$t,Data_Full$Cohort,Data_Full$Type)
```

Appendix F: Bayesian modeling OpenBUGS code

Scenario 1: Fully random structure model (Section 4.2.1)
Replace red code with *blue code* to switch to Scenario 2: Calendar shock sub-model (Section 4.2.2).

```
model {
        for (i in 1:n.ind) {
                for (j in 1:1) {
                        data_O[i, j] ~ dnorm(mean_O[i, j] , tau_O)
                        data_P[i, j] ~ dnorm(mean_P[i, j] , tau_P)
                        data_I[i, j] <- data_O[i, j] + data_P[i, j]

                        mean_O[i, j] <- solution[i,j,2]
                        mean_P[i, j] <- solution[i, j, 3]
                        mean_I[i, j] <- mean_O[i, j] + mean_P[i, j]
                                                                        }
                for (j in 2:n.grid) {
                        data_O[i, j] ~ dnorm(mean_O[i, j] , tau_O2)
                        data_P[i, j] ~ dnorm(mean_P[i, j] , tau_P2)
                        data_I[i, j] <- data_O[i, j] + data_P[i, j]

                        mean_O[i, j] <- solution[i, j, 2] + rho2 * (data_O[i, j-1] - mean_O[i, j-1])

                        #Calendar shock substitution
                        #mean_O[i, j] <- solution[i,j,2] * (1 - C[i,j] * a[i]) + rho2 *(data_O[i, j-1] -
                        #mean_O[i, j-1])

                        mean_P[i, j] <- solution[i, j, 3] + rho3 * (data_P[i, j-1] - mean_P[i, j-1])
                        mean_I[i, j] <- mean_O[i, j] + mean_P[i, j]
                                                                        }
                theta[i, 1:p] ~ dmnorm(mu[1:p], omega.inv[1:p, 1:p])

                param[i, 1] <- theta[i, 1]
                param[i, 2] <- theta[i, 2]
                param[i, 3] <- theta[i, 3]
                param[i, 4] <- theta[i, 4]
                param[i, p+1] <- prem[i]

                Ber[i] <- exp(theta[i, 1])
                RLR[i] <- exp(theta[i, 2])
                kp[i] <- exp(theta[i, 3])
                RRF[i] <- exp(theta[i, 4])

                ULR[i] <- RLR[i] * RRF[i]
                ILR10[i] <- data_I[i, 10] / prem[i]

                solution[i, 1:n.grid, 1:dim] <- ode(inits[i, 1:dim],
                                        grid[1:n.grid], D(A[i, 1:dim], t[i]), origin, tol)

                D(A[i, 1], t[i]) <- -Ber[i] * t[i] * A[i, 1]
                D(A[i, 2], t[i]) <- Ber[i] * t[i] * RLR[i] * A[i, 1] - kp[i] * A[i, 2]
```

```
        D(A[i, 3], t[i]) <- kp[i] * RRF[i] * A[i, 2]

        #Calendar shock substitution
        #D(A[i, 3], t[i]) <- kp[i] * RRF[i] * (1 - V[i]*a[i]) * A[i, 2]
        #V[i] <- step((i + t[i]) - 10)
        #a[i] ~ dunif(-0.99,0.99)

        inits[i, 1] <- prem[i]
        inits[i, 2] <- 0
        inits[i, 3] <- 0
                                    }

mu[1:p] ~ dmnorm(mu.prior.mean[1:p], mu.prior.prec[1:p, 1:p])
omega.inv[1:p, 1:p] ~ dwish(omega.inv.matrix[1:p, 1:p], omega.inv.dof)

omega[1:p, 1:p] <- inverse(omega.inv[1:p, 1:p])
ResC <- omega[2, 4] / (sqrt(omega[2, 2]) * sqrt(omega[4, 4]))

sigma_O ~ dunif(0, 10000)
tau_O <- pow(sigma_O, -2)

sigma_O2 <- sigma_O * sqrt(1 - pow(rho2, 2))
tau_O2 <- pow(sigma_O2, -2)

sigma_P ~ dunif(0, 5000)
tau_P <- pow(sigma_P, -2)

sigma_P2 <- sigma_P * sqrt(1 - pow(rho3, 2))
tau_P2 <- pow(sigma_P2, -2)

rho2 ~ dunif(-1,1)
rho3 ~ dunif(-1,1)


#Standardized residuals
for (i in 1:n.ind) {
        for (j in 1:1) {
                r_O[i,j] <- (data_O[i, j] - mean_O[i, j] ) * sqrt(tau_O)
                r_P[i,j] <- (data_P[i, j] - mean_P[i, j] ) * sqrt(tau_P)

                                            }
        for (j in 2:n.grid) {

                r_O[i,j] <- (data_O[i, j] - mean_O[i, j] ) * sqrt(tau_O2)
                r_P[i,j] <- (data_P[i, j] - mean_P[i, j] ) * sqrt(tau_P2)

                                            }
                                    }
                            }
```

Data and prior parameters

```
list(
p = 4, dim = 3,
origin = 0.0,
tol = 1.0E-6,
n.ind = 10, n.grid = 10,
grid = c(1,2,3,4,5,6,7,8,9,10),
prem = c(104437, 88883, 85956, 99339, 104897, 119427, 110784, 77731, 63646, 48052),
mu.prior.mean = c(1.7, -0.15, -0.9, -0.21),
mu.prior.prec = structure(
.Data = c(
650, 0, 0, 0,
0, 380, 0, 0,
0, 0, 5400, 0,
0, 0, 0, 390),
.Dim = c(4, 4)),
omega.inv.matrix = structure(
.Data = c(
1, 0, 0, 0,
0, 1, 0, 0.8,
0, 0, 1, 0,
0, 0.8, 0, 1),
.Dim = c(4, 4)),
omega.inv.dof = 4,

data_O = structure(.Data = c(
53121, 41222, 32309, 24944, 17104, 13137, 9605, 6515, 1661, 1322,
54145, 37188, 26976, 20015, 14319, 10179, 6672, 2575, 2071, NA,
55211, 37221, 27760, 17990, 11417, 6716, 4282, 3015, NA, NA,
60617, 42144, 25987, 14805, 9406, 5792, 3966, NA, NA, NA,
65719, 46047, 31250, 22245, 11878, 8408, NA, NA, NA, NA,
68133, 51102, 39934, 21824, 16955, NA, NA, NA, NA, NA,
62434, 46661, 32248, 24140, NA, NA, NA, NA, NA, NA,
56971, 48677, 35336, NA, NA, NA, NA, NA, NA, NA,
56526, 41707, NA, NA, NA, NA, NA, NA, NA, NA,
40799, NA, NA, NA, NA, NA, NA, NA, NA, NA),
.Dim = c(10,10)),
data_P = structure(.Data = c(
9558, 22778, 33298, 40348, 45146, 48048, 49782, 50623, 51812, 51939,
7913, 19472, 29622, 36816, 40975, 43302, 44707, 45871, 46229, NA,
8744, 24302, 35406, 43412, 48057, 50897, 52879, 53956, NA, NA,
13301, 32950, 47201, 56394, 61650, 65039, 66566, NA, NA, NA,
11424, 29086, 42034, 50910, 56406, 59437, NA, NA, NA, NA,
11792, 27161, 38229, 46722, 50742, NA, NA, NA, NA, NA,
11194, 26893, 38488, 45580, NA, NA, NA, NA, NA, NA,
12550, 31604, 44045, NA, NA, NA, NA, NA, NA, NA,
13194, 31474, NA, NA, NA, NA, NA, NA, NA, NA,
9372, NA, NA, NA, NA, NA, NA, NA, NA, NA),
.Dim = c(10,10))
)
```

```
#Calendar shock substitution
#,C = structure(
#.Data = c(
#0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
#0, 0, 0, 0, 0, 0, 0, 1, 1, 1,
#0, 0, 0, 0, 0, 0, 1, 1, 1, 1,
#0, 0, 0, 0, 0, 1, 1, 1, 1, 1,
#0, 0, 0, 0, 1, 1, 1, 1, 1, 1,
#0, 0, 0, 1, 1, 1, 1, 1, 1, 1,
#0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
#0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
#1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
#1, 1, 1, 1, 1, 1, 1, 1, 1, 1),
#.Dim = c(10,10))
#)
```

Initial values (1)

```
list(
rho2 = 0.5,
rho3= 0.5,
sigma_O = 5000,
sigma_P = 500,
mu = c(1.7, -0.15, -0.9, -0.21),
omega.inv = structure(
.Data = c(
10, 0, 0, 0,
0, 10, 0, 0.8,
0, 0, 10, 0,
0, 0.8, 0, 10),
.Dim = c(4, 4)),
theta = structure(
.Data = c(
1.7, -0.15, -0.9, -0.21,
1.7, -0.15, -0.9, -0.21,
1.7, -0.15, -0.9, -0.21,
1.7, -0.15, -0.9, -0.21,
1.7, -0.15, -0.9, -0.21,
1.7, -0.15, -0.9, -0.21,
1.7, -0.15, -0.9, -0.21,
1.7, -0.15, -0.9, -0.21,
1.7, -0.15, -0.9, -0.21,
1.7, -0.15, -0.9, -0.21),
.Dim = c(10, 4))
)
```

Initial values (2)

```
list(
rho2 = 0.6,
rho3= 0.2,
sigma_O = 3000,
sigma_P = 700,
mu = c(1.4, -0.07, -0.2, -0.51),
omega.inv = structure(
.Data = c(
15, 0, 0, 0,
0, 15, 0, 0.5,
0, 0, 15, 0,
0, 0.5, 0, 15),
.Dim = c(4, 4)),
theta = structure(
.Data = c(
1.4, -0.07, -0.2, -0.51,
1.4, -0.07, -0.2, -0.51,
1.4, -0.07, -0.2, -0.51,
1.4, -0.07, -0.2, -0.51,
1.4, -0.07, -0.2, -0.51,
1.4, -0.07, -0.2, -0.51,
1.4, -0.07, -0.2, -0.51,
1.4, -0.07, -0.2, -0.51,
1.4, -0.07, -0.2, -0.51,
1.4, -0.07, -0.2, -0.51),
.Dim = c(10, 4))
)
```

Initial values (3)

```
list(
rho2 = 0.2,
rho3= 0.6,
sigma_O = 1500,
sigma_P = 1000,
mu = c(1.1, 0, 0, -0.29),
omega.inv = structure(
.Data = c(
5, 0, 0, 0,
0, 5, 0, 0.3,
0, 0, 5, 0,
0, 0.3, 0, 5),
.Dim = c(4, 4)),
theta = structure(
.Data = c(
1.1, 0, 0, -0.29,
1.1, 0, 0, -0.29,
1.1, 0, 0, -0.29,
1.1, 0, 0, -0.29,
1.1, 0, 0, -0.29,
1.1, 0, 0, -0.29,
1.1, 0, 0, -0.29,
1.1, 0, 0, -0.29,
1.1, 0, 0, -0.29,
1.1, 0, 0, -0.29),
.Dim = c(10, 4))
)
```

7. REFERENCES

[1] Antonio, K. & Plat, R. (2014) "Micro–level stochastic loss reserving in general insurance" in *Scandinavian Actuarial Journal*, **7**, pp. 649-669

[2] Antonio, K. & Zhang, Y. (2014) "Nonlinear Mixed Models" in *Predictive Modeling Applications in Actuarial Science, Volume 1: Predictive Modeling Techniques*, UK: Cambridge University Press, pp. 367-396

[3] Barnard, J., McCulloch, R. E., & Meng, X. L. (2000) "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage" in *Statistica Sinica*, **10**, pp. 1281-1311

[4] Blomhøj, M., Kjeldsen, T. H., Ottesen, J. (2014) "Compartment models". Online resource available at: http://www4.ncsu.edu/~msolufse/Compartmentmodels.pdf. [Accessed May 2016]

[5] Bornhuetter, R. L. & Ferguson, R. E. (1972) "The Actuary and IBNR" in *Proceedings of the Casualty Actuarial Society* 1972, **59**, pp. 181–95

[6] Brauer, F. (2008) "Compartmental Models in Epidemiology" in *Mathematical Epidemiology*, van den Driessche & P., Wu, J. (eds.), Berlin: Springer, pp. 19-79

[7] Clark, D. R. (2003) "LDF Curve Fitting and Stochastic Loss Reserving: A Maximum Likelihood Approach" in *CAS Forum 2003: Fall*, pp. 41-91

[8] Clark, D. R. & Rangelova, D. (2015) "Accident Year / Development Year Interactions" in *CAS E-Forum 2015: Fall*, pp. 1-29

[9] Cleveland, W. S. (1979) "Robust locally weighted regression and smoothing Scatterplots" in *Journal of the American Statistical Association*, **74**(368), pp. 829-836.

[10] Doetsch, G. (1974) Introduction to the Theory and Application of the Laplace Transformation, Berlin: Springer-Verlag

[11] England, P. D. & Verrall, R. J. (1999) "Analytic and bootstrap estimates of prediction errors in claims reserving" in *Insurance: Mathematics and Economics*, **25**, pp. 281-293

[12] England, P. & Verrall, R.J. (2002) "Stochastic claims reserving in general insurance" in *British Actuarial Journal*, **8**, pp. 443-544

[13] Verrall, R. J. & England, P. D. (2005) "Incorporating expert opinion into a stochastic model for the chain-ladder technique" in *Insurance, mathematic and economics*, **37**, pp. 355-370

[14] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013) Bayesian Data Analysis (3rd Edition), London: Chapman & Hall/CRC Press

[15] Guszcza, J. C. (2008) "Hierarchical Growth Curve Models for Loss Reserving" in *CAS Forum 2008: Fall*, pp. 146-173

[16] Hachemeister, C.A. (1980) "A stochastic model for loss reserving" in *Proceedings of the International Congress of Actuaries, Zürich*, pp. 185-194

[17] Hesselager, O. & Markov, A. (1994) "Model for loss reserving" in ASTIN Bulletin **24**(2), pp. 183-93. Online resource available at: http://dx.doi.org/10.2143/ast.24.2.2005064 [Accessed May 2016]

[18] Johnson, N. L. & Kotz, S. (1972) Distributions in Statistics, Vol. 4, New York: Wiley

[19] Klim, S., Mortensen, S., Kristensen, N., Overgaard, R., & Madsen, H. (2009) "Population stochastic modelling (PSM) - an R package for mixed-effects models based on stochastic differential equations" in *Computer Methods and Programs in Biomedicine*, **94**, pp. 279–289

[20] Line, N., Archer-Lock, P., Fisher, S., Hilder, I., Shah, S., Wenzel, K., & White, M. (2003) "The Cycle Survival Kit: An Investigation into the Reserving Cycle and Other Issues" in *Proceeding of the General Insurance Convention, 2003*

[21] Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D., (2000) "WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility" in *Statistics and Computing*, **10**, pp. 325–337

[22] Macheras, P. (1987) "Method of residuals: Estimation of absorption and elimination rate constants having comparable values" in *Biopharmaceutics & Drug Disposition*, **8**, pp. 47–56

[23] Martinez-Miranda M.D., Nielsen, J. P. & Verrall, R. J. (2012) "Double Chain Ladder" in *ASTIN Bulletin*, **42**(1), pp.59-76

[24] Meyers, G. (2007) "Estimating Predictive Distributions for Loss Reserve Models" in *Variance*, **1**(2), Fall 2007, pp. 248-272

[25] Norberg, R. (1993) "Prediction of outstanding liabilities in non-life Insurance" in *ASTIN Bulletin*, **23**(1), pp. 95-115. Online resource available at: http://dx.doi.org/10.2143/ast.23.1.2005103 [Accessed May 2016]

[26] Orr, J. (2007) "A simple multi-state reserving model" in *37th Int. ASTIN Colloquium, Orlando, 2007*. Online resource available at: URL: http://www.actuaries.org/ASTIN/Colloquia/Orlando/Papers/Orr.pdf [Accessed May 2016]

[27] Overgaard, R. V., Jonsson, E. N., Tornøe, C. W., & Madsen, H. (2005) "Non-linear mixed-effects models with stochastic differential equations. Implementation of an estimation algorithm" in *Journal of Pharmacokinetics and Pharmacodynamics*, **32**(1), pp. 85-107

[28] Parodi, P. (2013) "Triangle-Free Reserving: A Non-Traditional Framework for Estimating Reserves and Reserve Uncertainty" *discussion paper for The Institute and Faculty of Actuaries*. Online resource available at: http://www.actuaries.org.uk/sites/all/files/documents/pdf/pietro-parodi-triangle-free-reserving-217-final.pdf [Accessed May 2016]

[29] Pinheiro, J. C. & Bates, D. M. (2000) Mixed-Effects Models in S and S-Plus, New York: Springer-Verlag

[30] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2016) nlme: Linear and Nonlinear Mixed Effects Models. *R package version 3.1-127*. URL: http://CRAN.R-project.org/package=nlme.

[31] Quarg, G. & Mack, T. (2004) "Munich chain ladder" in *Blätter der Deutschen Gesellschaft für Versicherungs- und Finanzmathematik*, **26**(4), pp. 597-630

[32] R Core Team (2016) "R: A language and environment for statistical computing" *R Foundation for Statistical Computing, Vienna, Austria*. URL: http://www.R-project.org/.

[33] Rescigno, A., & Segre, G. (1966) Drug and Tracer Kinetics, Waltham, MA: Blaisdell

[34] Seber, G.A.F. & Wild, C. J. (1989) Nonlinear Regression, New York: Wiley

[35] Sheppard, C. W. (1948) "The theory of the study of transfers within a multi-compartment system using isotopic tracers" in *Journal of Applied Physics*, **19**, pp. 70-76

[36] Tornøe, C. W., Agersø, H., Jonsson, E. N., Madsen, H., Nielsen HA (2004a). "Non-Linear Mixed-Effects Pharmacokinetic/Pharmacodynamic Modelling in NLME Using Differential Equations" in *Computer Methods and Programs in Biomedicine*, **76**(1), pp. 31–40

[37] Tornøe, C. W., Jacobsen, J. L., Pedersen, O., Hansen, T., & Madsen, H. (2004b) "Grey-box modelling of pharmacokinetic/pharmacodynamic systems" in *Journal of Pharmacokinetics and Pharmacodynamics*, **31**(5), pp. 401–417

[38] Turner, M.E. (1960) "Straight line regression through the origin" in *Biometrics*, **16**(3), pp. 483–485.

[39] Verrall, R.J. (2004) "A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving" in *North American Actuarial Journal*, **8**, pp. 67-89

[40] Zhang, Y., Dukic, V., & Guszcza, J. C. (2012). "A Bayesian Nonlinear Model for Forecasting Insurance Loss Payments" in *Journal of the Royal Statistical Society*: Series A, **175**, pp. 637–656

Abbreviations and notations

AY, accident year (row dimension of triangle)
AIC, Akaike information criterion
BBNI, bound but not incepted
BIC, Bayesian information criterion
CY, calendar year
DIC, deviance information criterion
EtR, Exposed to Risk
EX, exposure
ExBNR, exposed but not reported
GLM, generalized linear model
IBNR, incurred but not reported
OS, outstanding claims
MCMC, Markov chain Monte Carlo
NLME, nonlinear mixed-effects
ODE, ordinary differential equation
PD, paid claims (cumulative)
PLR, paid loss ratio
PPI, posterior predictive interval
RBNS, reported but not settled
RBC, reported burning cost
RLR, reported loss ratio
RRF, reserve robustness factor
SDE, stochastic differential equation
UBC, ultimate burning cost
ULR, ultimate loss ratio

Biography of the Author

Jake Morris is an actuary at Liberty Specialty Markets in London UK and a Fellow of the Institute and Faculty of Actuaries. Contact: jake.morris@libertyglobalgroup.com