

Adaptive Reserving using Bayesian Revision for the Exponential Dispersion Family

by Greg Taylor and Gráinne McGuire

ABSTRACT

This paper investigates the practical aspects of applying the second-order Bayesian revision of a generalized linear model (GLM) to form an adaptive filter for claims reserving. It discusses the application of such methods to three typical models used in Australian general insurance circles. Extensions, including the application of bootstrapping to an adaptive filter and the blending of results from the three models, are considered.

KEYWORDS

Adaptive filter, adaptive reserving, Bayesian revision, bootstrap, dynamic generalized linear model, DGLM, loss reserving, model blending, stochastic reserving

1. Introduction

The estimation of outstanding claims liabilities is one of the central problems in general insurance. Many techniques exist to carry out this task—for example, the chain ladder in its various forms, payments per claim models, and case estimate models. A good review of existing methods is given in Taylor (2000).

Taylor (2000) also discusses a state space model, the Kalman filter (Kalman 1960), which was introduced into the actuarial literature by De Jong and Zehnwirth (1983). This is a recursive algorithm, commonly run by accident or calendar year, which, when applied to the modeling of outstanding claims liabilities, permits the model parameters to evolve over time.

Although a very flexible tool, the Kalman filter requires the assumption of normally distributed data. In the context of loss reserving, this is generally dealt with by assuming that the variables of interest (claim sizes, payments in a specific period, etc.) are lognormally distributed. Although a common assumption in the actuarial field [see, e.g., England and Verrall (2002)], this is an approach not without its problems. The main difficulty is the requirement for a bias correction which results from modeling a transformation of the data (in this case $\log(\text{payments})$), the determination of which, particularly when the distribution of the data is only approximately known, can be problematic. A second drawback is the restriction to one possible distribution for the data. Modeling claim counts, for example, would not be easily done with either a normal or lognormal distribution.

An alternative family of distributions that is worth considering is the *exponential dispersion family* (termed “EDF” within this paper). This family was introduced by Nelder and Wedderburn (1972) and is treated in depth by McCullough and Nelder (1989). The family includes common distributions like the normal, Poisson,

gamma, and inverse Gaussian. The gamma distribution, and to a lesser extent, the inverse Gaussian distribution, are both candidates for the modeling of long-tailed, strictly positive variables. The Poisson distribution is a natural choice for claim counts. Applied within the framework of a filter, this would mean updating a generalized linear model (McCullough and Nelder 1989) rather than a normal linear model (the Kalman filter).

In general, updating a GLM is analytically intractable unless some very restrictive assumptions are made [see, e.g., West, Harrison, and Migon (1995)]. However, putting the problem into a Bayesian context, where the parameter estimates and distribution from the previous application of the filter form the prior distribution of the current step of the filter, it may be seen that the problem is amenable to modeling using Markov Chain Monte Carlo techniques [see, for example, Hastings (1970); Smith and Roberts (1993); Tierney (1994)]. Examples of MCMC in the actuarial literature appear in Scollnik (2001, 2002a), Nzoufras and Dellaportas (2002) and de Alba (2002). Indeed, Nzoufras and Dellaportas (2002) discuss the application of MCMC to some state space models, the implementation of which is further discussed in Scollnik (2002b).

However, MCMC simulation is a time-consuming process, with pitfalls for the unwary user (primarily convergence issues). It would be advantageous to have an analytical version for members of the exponential dispersion family corresponding to the Kalman filter. Taylor (2008) derives a second order approximation to the posterior likelihood of a naturally conjugated generalized linear model (GLM). This family of second order approximations is forced to be closed under Bayesian revision. This is a useful finding since it means that the process may be used recursively, i.e., it may be used to form an adaptive filter, which may be used in the modeling of claims data.

Results from an adaptive filter may be used to generate central estimates of an outstanding

claims liability. In general insurance practice, other moments are important, in particular the variance of the reserve as well as certain quantiles (e.g., 75%, 95% points). It is possible to use bootstrapping (e.g., Efron and Tibshirani 1993) to yield such information from an adaptive filter. However, standard forms of bootstrapping as applied to actuarial data [see, for example, Taylor (2000)] assume that the residuals are independent. Residuals that result from an adaptive filtering process are, in fact, not independent. Thus, it is necessary to adapt an approach similar to that outlined in Stoffer and Wall (1991) for normally distributed state space models.

Since adaptive filtering, even with bootstrapping, is much quicker than MCMC, it is practical to apply several different models to one data set. These allow the capture of different aspects of the claims experience. To produce a final answer, it is necessary to produce a blended version of all the models. One possible approach is presented in Taylor (2000).

The purpose of this paper is to demonstrate the use of the application of Taylor's (2008) GLM adaptive filter, incorporating the calculation of central estimates, the application of bootstrapping, and the blending of results from several models to produce a final result. In the remainder of the paper, the data are described in Section 2, while Section 3 introduces the methodology underlying the GLM adaptive filter. This discussion encompasses adaptive filter methodology, an appropriate form of bootstrap for such filters, and means for blending the results from several models. The various models applied to the data are described in Section 4. Section 5 presents some numerical results. The paper closes with some discussion in Section 6.

2. Data

The data used in this paper, from a bodily injury liability portfolio, were summarized as

follows:

- Total amount of claim payments for year of accident i ($i = 1, 2, \dots, 14$) and development quarter j ($j = 1, 2, \dots, 53$)
- Total number of claims reported for year of accident i and development quarter j
- Total number of claim closures for year of accident i and development quarter j
- Total amount of case estimates for year of accident i and development quarter j

The data were split into two groups by legal jurisdiction, and each of the above summaries was available for each of these two groups.

Ultimate numbers of incurred claims by accident year were also available; these had been estimated in a separate modeling exercise.

The payments and case estimate data were adjusted for past economic inflation in line with an appropriate wage earning index.

Some notation is defined here. Let

- C_{ij}^L = claim payments in development quarter j for accident year i for jurisdictional grouping L ($L = 1, 2$);
- N_{ij}^L = number of claims reported in development quarter j for accident year i for jurisdictional grouping L ;
- F_{ij}^L = number of claim closures in development quarter j for accident year i for jurisdictional grouping L ;
- E_{ij}^L = case estimates at end of development quarter j for accident year i for jurisdictional grouping L ;
- U_i^L = estimated ultimate number of claims incurred in accident year i for jurisdictional grouping L ; and
- O_{ij}^L = number of claims open for accident year i , at start of development quarter j for jurisdictional grouping L .

3. Methodology

This section opens with a discussion of the Kalman filter and outlines the generalization to

the analytical GLM filter developed by Taylor (2008). It then describes the adaptation of the Stoffer and Wall (1991) bootstrap required for GLM filters. Finally, the blending methodology is described.

3.1. Kalman filter

Following Taylor's (2000) description, the model underlying the Kalman filter consists of two equations, called the *system* equation and *observation* equation. The former describes the model's parameter evolution, while the latter describes the model of observations conditional on the parameters. The two equations are as follows:

System equation

$$\underset{p \times 1}{\beta(s+1)} = \underset{p \times p}{\Phi(s+1)} \underset{p \times 1}{\beta(s)} + \underset{p \times 1}{r(s+1)}. \quad (3.1)$$

Observation equation

$$\underset{n \times 1}{Y(s+1)} = \underset{n \times p}{X(s+1)} \underset{p \times 1}{\beta(s+1)} + \underset{n \times 1}{v(s+1)}. \quad (3.2)$$

These equations are written in vector and matrix form with dimensions written beneath, and

- $Y(s+1)$ denotes the vector of observations made at time $s+1$ ($= 1, 2, \dots, T$).
- $\beta(s+1)$ denotes the vector of parameters at time $s+1$.
- $X(s+1)$ is the design matrix applying at time $s+1$.
- $\Phi(s+1)$ is a transition matrix governing the evolution of the expected parameter values from one epoch to the next.
- $r(s+1)$ and $v(s+1)$ are stochastic perturbations, each with zero mean, and with

$$\begin{aligned} \text{Var}[r(s+1)] &= R(s+1) \quad \text{and} \\ \text{Var}[v(s+1)] &= V(s+1). \end{aligned} \quad (3.3)$$

The Kalman filter is an algorithm for calculating estimates of the $\beta(s+1)$. It revises the parameter estimates iteratively over time. Each iteration introduces additional data $Y(s+1)$. Specific de-

tails of its application to loss reserving may be found in Taylor (2000).

The formal statement of the filter requires a little extra notation. Let $Y(s | s-k)$ denote the filter's estimate of $Y(s)$ on the basis of information up to and including epoch $s-k$; and similarly for other symbols at $s | s-k$. Also, let $\Theta(s | s-k)$ denote the estimate of $\text{Var}[\beta(s | s-k)]$.

Equations (3.4) through (3.8) present the mathematical detail behind the Kalman filter. The process may be split into four steps.

Step 1

$$\beta(s | s-1) = \Phi(s)\beta(s-1 | s-1) \quad (3.4)$$

$$\Theta(s | s-1) = \Phi(s)\Theta(s-1 | s-1)\Phi^T(s) + R(s). \quad (3.5)$$

Step 2

$$L(s | s-1) = X(s)\Theta(s | s-1)X^T(s) + V(s) \quad (3.6)$$

$$K(s) = \Theta(s | s-1)X^T(s)[L(s | s-1)]^{-1}. \quad (3.7)$$

Step 3

$$\hat{Y}(s | s-1) = X(s)\beta(s | s-1). \quad (3.8)$$

Step 4

$$\beta(s | s) = \beta(s | s-1) + K(s)(Y(s) - \hat{Y}(s | s-1)) \quad (3.9)$$

$$\Theta(s | s) = (1 - K(s)X(s))\Theta(s | s-1). \quad (3.10)$$

Equations (3.4) to (3.8) generate forecasts of the new epoch's parameters and observations based on no new information. The gain matrix (credibility of the new observation) is calculated in Equation (3.7). Finally the parameter estimates are updated in Equations (3.9) and (3.10). The process then moves on to the next epoch, starting again with Equation (3.4).

The filter starts with prior estimates for $\beta(0 | 0)$ and the associated dispersion $\Theta(0 | 0)$.

The key equation for revising parameter estimates is Equation (3.9). The main characteristics of this formula are that:

- It is linear in the data vector $Y(s)$.

- $\beta(s | s)$ is the Bayesian revision of the estimate of $\beta(s-1 | s-1)$ in the case that $r(s)$ and $v(s)$ are normally distributed.
- Consequently, $K(s)$ is equivalent to the credibility of $Y(s)$.

The Kalman filter is a form of an *adaptive filter*—a form of time series estimation in which parameter estimates are constructed so as to track evolving parameters.

3.2. Exponential dispersion family

The exponential dispersion family (EDF) (Nelder and Wedderburn 1972) is a large family of distributions of varying tail length. This paper uses a slightly restricted version of the family whose members have a likelihood of the following form:

$$L(y | \theta, \lambda) = a(\lambda, y) \exp \lambda[y\theta - b(\theta)], \quad (3.11)$$

where θ , λ are parameters and $a(\cdot)$ and $b(\cdot)$ are functions characterizing the members of the family. The more general form of EDF allows λ to be a function $\lambda(\varphi)$ of a dispersion parameter φ rather than a constant.

It may be shown that, for the distribution defined in Equation (3.11),

$$\begin{aligned} E[Y | \theta, \lambda] &= b'(\theta), \quad \text{and} \\ \text{Var}[Y | \theta, \lambda] &= b''(\theta)/\lambda. \end{aligned} \quad (3.12)$$

Denote $b'(\theta)$ by $\mu(\theta)$ whence, provided that $\mu(\cdot)$ is one-one

$$\text{Var}[Y | \theta, \lambda] = V(\mu(\theta))/\lambda \quad (3.13)$$

for some function $V(\cdot)$ called the *variance function*.

Many applications of the EDF restrict the form of the variance function to

$$V(\mu) = \mu^p \quad (3.14)$$

for some constant $p \leq 0$ or $p \geq 1$. This family is referred to as the Tweedie family (Tweedie 1984). The likelihoods corresponding to such a restriction will be referred to below as EDF(p).

Some special cases of the EDF are:

- $p = 0$: normal
- $p = 1$: Poisson
- $p = 2$: gamma
- $p = 3$: inverse Gaussian

A further quantity associated with the EDF is the *scale parameter*. This is given by

$$\varphi = 1/\lambda. \quad (3.15)$$

3.3. Analytical GLM filter

Taylor (2008) examined the problem of finding closed-form solutions to the Bayesian revision problem for GLMs. That paper builds on the univariate case where the exponential dispersion family is closed under Bayesian revision in the presence of natural conjugate priors. Although this is not the case for the general multivariate EDF, Taylor (2008) derives a second order approximation to the posterior likelihood of a naturally conjugated generalized linear model. This family of second order approximations is forced to be closed under Bayesian revision. This is a useful result since it means that the process may be used recursively, i.e., it may be used to form an adaptive filter, thereby allowing a model to adapt to changing experience.

Although Taylor (2008) gives a general form of the filter, he notes that it is only analytically tractable in a limited set of cases, in particular where the link function corresponds to either the canonical or companion canonical link (as defined within the paper). Tractable cases include the Poisson distribution with log link and the gamma distribution with reciprocal link (canonical link) or log link (companion canonical link). This filter is termed the *analytical GLM filter* (or GLM filter for short) in this paper. The filter implements a dynamic generalized linear model, i.e., a GLM in which parameters vary over time.

Note also that the normal distribution with identity link is also a member of this class of filters; Taylor (2008) shows that this is exactly

equal to the Kalman filter. In other words, the second order approximations used in that paper become exact in the case of the normal distribution and an identity link.

Full details of the general form of the filter as well as some specific examples are available in Taylor (2008). However, since the cases of gamma error/log link (companion canonical link) and the Poisson error/log link (canonical link) versions of the filter will be used extensively in this paper, the series of equations defining this filter is given here.

3.3.1. The gamma GLM filter

The equations underlying the analytical GLM filter follow the Kalman filter, in principle, but are considerably more complex. Below, the key definitions and equations for the gamma GLM filter are given, with supplementary formulae in Appendix A.1. Note that in all the equations below (and those in Appendix A and Appendix B), functions are assumed to operate componentwise on vectors.

The system equation for the analytical gamma GLM filter is the same as that in (3.1). However, the observation equation differs due to the log link and is

$$Y(s+1) = \exp\{X(s+1)\beta(s+1)\} + v(s+1). \quad (3.16)$$

Some other quantities of importance within the gamma GLM are defined as follows:

$$\begin{aligned} 1/h^{-1}(s | s-j) &= E[\exp\{-M(s | s-j)\beta(s | s-j)\}] \\ \Gamma(s | s-j) &= \text{Var}[\exp\{-M(s | s-j)\beta(s | s-j)\}] \end{aligned} \quad (3.17)$$

where $M(s | s-j)$ is defined in Appendix A and $s | s-j$ denotes estimates at epoch s on the basis of information up to and including epoch $s-j$.

The gamma GLM filter is applied to these quantities, from which the parameter means and variances are estimated. In the case of the Kalman

filter, the corresponding expressions in Equations (3.17) simplify to $\beta(s | s-j)$ and $\Theta(s | s-j)$ from Section 3.1.

The equations describing the gamma GLM filter are given below. Where they make reference to matrices not defined in this section, the reader should refer to Appendix A.1 for their definitions.

Step 1

$$E[\beta(s | s-1)] = \Phi(s)E[\beta(s-1 | s-1)] \quad (3.18)$$

$$\begin{aligned} \text{Var}[\beta(s | s-1)] &= \Phi(s)\text{Var}[\beta(s-1 | s-1)]\Phi^T(s) \\ &\quad + R(s), \end{aligned} \quad (3.19)$$

where $\Phi(s+1)$ is the identity matrix in the example of filtering in this paper. Equation (3.19) indicates that the parameter vector at any epoch (accident period) is a deterministic linear transformation and stochastic perturbation of its value at the previous epoch.

Also adopt the following approximations to Equations (3.17):

$$\begin{aligned} \frac{1}{h^{-1}(s | s-1)} &= \left[I + \frac{1}{2}Q(s | s-1) \right] \\ &\quad \times \exp\{-E[M(s | s-1)\beta(s | s-1)]\} \end{aligned} \quad (3.20)$$

$$\begin{aligned} \Gamma(s | s-1) &= Q(s | s-1)\text{DIAG} \\ &\quad \times \exp\{-2E[M(s | s-1)\beta(s | s-1)]\} \end{aligned} \quad (3.21)$$

where $\text{DIAG}(\cdot)$ is an operator that converts the subject vector to a diagonal matrix.

Step 2

$$K(s) = B(s | s)^{-1}P(s)J(s)G(s). \quad (3.22)$$

Step 3

$$\hat{Y}(s | s-1) = \exp\{X(s)\beta(s | s-1)\}. \quad (3.23)$$

Step 4

$$\begin{aligned} 1/h^{-1}(s | s) &= \text{DIAG} \exp[-M(s | s)\beta(s | s-1)] \\ &\quad \times [1 - K(s)(Y(s) - \hat{Y}(s | s-1))] \end{aligned} \quad (3.24)$$

where $\mathbf{1}$ represents a vector with all entries 1, rather than the unit matrix

$$\Gamma(s | s) = -D(s)^{-1}B(s | s)^2 \quad (3.25)$$

$$E[\beta(s | s)] = M(s | s)^T \left\{ -\log \frac{1}{h^{-1}(s | s)} + \frac{1}{2} \Gamma(s | s) H_{(-1)}(s | s)^2 \mathbf{1} \right\} \quad (3.26)$$

$$\text{Var}[\beta(s | s)] = M(s | s)^T H_{(-1)}(s | s) \Gamma(s | s) H_{(-1)}(s | s) M(s | s). \quad (3.27)$$

In (3.26) and (3.27),

$$H_{(-1)}(s | s) = \text{DIAG } h^{-1}(s | s) \quad (3.28)$$

where $h^{-1}(s | s)$ is the reciprocal of $1/h^{-1}(s | s)$.

The filter is initialized by setting initial values (i.e., prior values) for $E[\beta(0 | 0)]$ and $\text{Var}[\beta(0 | 0)]$. Further, input values are required for $\Lambda(s)$ —for the gamma distribution, the coefficients of variation of the data $Y(s)$ are the appropriate choice. The $R(s)$ may all be set to a constant matrix.

3.3.2. The Poisson adaptive filter

The Poisson adaptive filter is also used in this paper. It is very similar to the gamma filter. Taylor (2008) discusses the relationship between the two, one based on the canonical link (Poisson/log link) and the other based on the companion canonical link (gamma/log link). The key equations are detailed below with supplementary detail in Appendix A.2.

Step 1

$$E[\beta(s | s-1)] = \Phi(s)E[\beta(s-1 | s-1)] \quad (3.29)$$

$$\text{Var}[\beta(s | s-1)] = \Phi(s)\text{Var}[\beta(s-1 | s-1)]\Phi(s) + R(s). \quad (3.30)$$

Also define:

$$h^{-1}(s | s-1) = [I + \frac{1}{2}Q(s | s-1)] \times \exp\{M(s | s-1)E[\beta(s | s-1)]\} \quad (3.31)$$

$$\Gamma(s | s-1) = Q(s | s-1)\text{DIAG} \times \exp\{2E[M(s | s-1)\beta(s | s-1)]\}. \quad (3.32)$$

Step 2

$$\hat{Y}(s | s-1) = \exp\{X(s)\beta(s | s-1)\}. \quad (3.33)$$

Step 3

$$h^{-1}(s | s) = \exp\{M(s | s)\beta(s | s-1)\} + P(s)J(s)[Y(s) - \hat{Y}(s | s-1)] \quad (3.34)$$

$$\Gamma(s | s) = D(s)^{-1}B(s | s)^2. \quad (3.35)$$

$$E[\beta(s | s)] = M(s | s)^T [\log(h^{-1}(s | s)) - \frac{1}{2}\Gamma(s | s)\{H_{(-1)}(s | s)^2\}\mathbf{1}] \quad (3.36)$$

$$\text{Var}[\beta(s | s)] = M(s | s)^T H_{(-1)}(s | s) \Gamma(s | s) \times H_{(-1)}(s | s) M(s | s). \quad (3.37)$$

In Equation (3.36), $\mathbf{1}$ represents a vector of ones, rather than a unit matrix, while

$$H_{(-1)}(s | s) = [\text{DIAG } h(s | s)]^{-1}. \quad (3.38)$$

3.4. Bootstrap

The GLM filter yields a central estimate of liability. For many applications, an indication of the uncertainty that is associated with the estimate is also useful.

Variability in a forecast results from a number of sources:

- **Parameter error:** sampling error in the parameter estimates;
- **Process error:** even if the parameter estimates match their true values, future experience will be subject to disturbance
- **Specification error:** the error caused by the model being incorrect

The first two of these may be measured using the bootstrap (Efron and Tibshirani 1993). Broadly speaking, a typical bootstrap application

involves repeatedly resampling the data to obtain many pseudo data sets. The filter is run through each and the resulting collection of forecast values (and parameter estimates if desired) forms an empirical estimate of the true distribution of that quantity.

Taylor (2000) and Pinheiro, Andrade e Silva, and Centeno (2003) review the application of bootstrapping to loss reserving problems; since the data in a loss reserving triangle are not, in general, identically distributed, an adjusted version of the bootstrap, based on standardized residuals must be used rather than the traditional bootstrap method of simply resampling data points.

When bootstrapping filtered data, a further complication arises in that the resulting residuals are not independent. The correct way of dealing with such lack of independence was discussed in Stoffer and Wall (1991) in relation to the case of normal distributions, identity links (i.e., the Kalman filter). A similar approach should be applied to GLM filters.

The key concept of the Stoffer and Wall (1991) paper is that dependencies must be allowed for in any bootstrapping process. Following the notation in Section 3.1, they first construct the *innovations*, defined as follows:

$$\varepsilon(s) = Y(s) - X(s)\beta(s | s-1). \quad (3.39)$$

The *innovations covariance matrix* is defined as

$$X(s)\Gamma(s | s-1)X^T(s) + V(s) = L(s | s-1). \quad (3.40)$$

By (3.39) and (3.40), the standardized innovations may be calculated as

$$e(s) = L^{-1/2}(s | s-1)\varepsilon(s). \quad (3.41)$$

To carry out one iteration of the bootstrap, the standardized innovations, $e(s)$, $s = 1, \dots, T$, are sampled T times with replacement (where T is the total number of epochs to which the filter is applied). Note that all components of these e_s are

stochastically independent and equivariant, and, for $\varepsilon(s)$ normal, equidistributed as is required for the bootstrap resampling.

To create a bootstrap data set, the standardization of the innovations is reversed using Equation (3.41), leading to pseudo innovations, $\varepsilon^*(s)$, $s = 1, \dots, T$. These may then be used in the Kalman filter, in Equation (3.9) in place of $Y(s+1) - \hat{Y}(s+1 | s)$. A key point to note is that, since the filter depends on the data only through this equation, there is no need to reconstruct a pseudo data set, as would be the case in typical applications of bootstrapping [see, e.g., Efron and Tibshirani (1993); Pinheiro, Andrade e Silva, and Centeno (2003)].

Apart from the modification to Equation (3.9), the remainder of the Kalman filter equations flow through unchanged for a bootstrap iteration, yielding one set of bootstrapped results. The process may then be repeated a large number of times to yield the bootstrapped estimate of the distribution of the results.

For the analytical GLM filter, Equation (3.39) is replaced with Equation (3.42), below, for the gamma/log link case:

$$\varepsilon(s) = 1/h^{-1}(s | s) - \exp\{-M(s | s)\beta(s | s-1)\}, \quad (3.42)$$

and Equation (3.43) for the Poisson/log link scenario:

$$\varepsilon(s) = h^{-1}(s | s) - \exp\{M(s | s)\beta(s | s-1)\}. \quad (3.43)$$

There is no simple expression for the innovations covariance matrix corresponding to Equation (3.40) for either case. Instead, these must be calculated using Equations (3.24) and (3.34) to find alternative specifications for Equations (3.42) and (3.43) and then calculating the approximate variance of these terms. Appendix B gives these calculations.

As in Stoffer and Wall (1991), the standardized innovations are resampled T times (where T is

the number of epochs in the data) for each bootstrap iteration, yielding $\tilde{e}(s)$, $s = 1, \dots, T$. Similarly, a pseudo innovation is calculated for each epoch and used in place of Equation (3.24) for gamma/log link and Equation (3.34) for Poisson log link. In the gamma/log case the substitution is

$$1/h^{-1}(s | s) = \tilde{e}(s) + \text{DIAG} \exp[-M(s | s)\beta(s | s - 1)], \quad (3.44)$$

where $\tilde{e}(s) = L^{1/2}(s | s - 1)\tilde{e}(s)$ represents the sampled innovation. A similar substitution is made in the Poisson/log link case.

The gamma/log link case requires some additional special handling. This is due to the appearance of the data, $Y(s)$, in Equations (A.7) and (A.9), meaning that, in the bootstrap, pseudo data appears within the calculation of the pseudo sample parameters. This renders the computation of pseudo data self-referencing.

The approach that has been taken here is to:

- Work through the Equations (A.7) and (A.9).
- Calculate (3.24) using the pseudo innovations. This yields an initial estimate of $1/h^{-1}(s | s)$.
- Calculate the implied value, $\tilde{\beta}(s | s - 1)$ from the relation $1/h^{-1}(s | s) = \exp\{-M(s | s) \cdot \tilde{\beta}(s | s - 1)\}$ (refer to (3.17)).
- Calculate the implied value, $\tilde{Y}(s)$ as $\tilde{Y}(s) = \exp\{X(s)\tilde{\beta}(s | s - 1)\}$ (refer (3.23)).
- Recalculate (A.7) and (A.9) with $\tilde{Y}(s)$ and then carry these effects through the remainder of the filter equations.

Note that, as for the Kalman filter, the pseudo innovations (3.42) and (3.43), after diagonalization and standardization, are stochastically independent and equivariant. However, they may not be equidistributed. For example, Poisson variates with differing means, when standardized, are not equidistributed.

Strictly, this breaches the bootstrap requirement of exchangeability of standardized (and diagonalized) residuals, and so introduces an error

into the application of the bootstrap to a GLM. We suppose this error to be slight.

Process error

Process error may be incorporated into the bootstrap results by means of the distribution assumed for the data. Thus, the methodology outlined above leads to a mean value for each cell; a value may then be sampled from the appropriate error distribution (e.g., gamma or Poisson) to incorporate process error.

Output

The ultimate output of a bootstrapping procedure, for an outstanding claims estimation process, is a distribution of the outstanding claims liability. However, the mean of the bootstrap distribution will not, in general, equal the mean estimated from the original data set. The disparity will be caused at least by sampling error arising from the finiteness of the sample forming the bootstrap distribution, and it may arise from other errors such as discussed immediately above.

To correct for this, the user may scale the bootstrap values to match the original estimates. This ensures consistency between the bootstrapped results and the mean (such as, for example, the location of various percentile points) while maintaining the same level of variability.

3.5. Blending of model results

It is common practice in actuarial reserving to apply more than one model to the data with the final estimates being a blend of those from all models. In principle, weights assigned to each model should be high where that model is reliable and low where it is not. Further, it may be desirable to take into account the smoothness of certain quantities as they progress over accident periods—for example, the smoothness of the weights themselves and the smoothness of the ratio of the estimates of liability to case es-

timates of that liability (call these the “case estimate ratios”).

Note that the notation used in this subsection differs in a couple of respects from that introduced in Section 2. For example C and U have different meanings, within the context of the blending algorithm, from those previously used.

The blending procedure followed here shares a lot of common ground with that described in Chapter 12 of Taylor (2000), which in turn is a modification of a procedure proposed in an earlier paper (Taylor 1985). The principle difference is that smoothness of the case estimate ratios is defined here in terms of their logged values, whereas it is defined in terms of their raw values in the cited work.

The reasoning supporting this is as follows. If R_i denotes the case estimate ratio for accident period i , the sequence $\{R_i, i = 1, 2, \dots, I\}$ will typically display relatively large values for the highest values of i , but tend to converge to a constant as $i \rightarrow 1$.

This will resemble an exponential rather than a linear sequence. If it were precisely exponential, then the sequence $\{\ln R_i\}$ would be linear, and second differences $\Delta^2 \ln R_i$ of this sequence would be zero. In this sense, smallness of the $\Delta^2 \ln R_i$ is a reasonable criterion for smoothness of the sequence $\{R_i\}$.

An outline of the blending procedure is presented below.

Let $\hat{L}_{ih}, i = 1, 2, \dots, I, h = 1, 2, \dots, H$ denote the estimate of outstanding claims liabilities for model h , for accident period i .

Consider blended estimates from the family of linear combinations of the models:

$$\hat{L}_i = \sum_{h=1}^H w_{ih} \hat{L}_{ih} \quad (3.45)$$

for deterministic coefficients w_{ih} .

Let

$$\hat{L}_i = (\hat{L}_{i1}, \hat{L}_{i2}, \dots, \hat{L}_{iH})^T \quad (3.46)$$

$$w_i = (w_{i1}, w_{i2}, \dots, w_{iH})^T \quad (3.47)$$

$$\hat{L}_{I \times HI} = \begin{bmatrix} \hat{L}_1^T & & & \\ & \hat{L}_2^T & & \\ & & \ddots & \\ & & & \hat{L}_I^T \end{bmatrix} \quad (3.48)$$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_I \end{bmatrix}. \quad (3.49)$$

Note that $\hat{L}$ is a *block diagonal* matrix, i.e., a block matrix with, in this case, a $1 \times H$ block for each entry on the diagonal and zero blocks elsewhere. Let

$$C_{HI \times HI} = \text{Var} \begin{bmatrix} \hat{L}_1 \\ \hat{L}_2 \\ \vdots \\ \hat{L}_I \end{bmatrix}. \quad (3.50)$$

Let q_i be the most recent case estimates for accident year i . Let

$$Q = \text{diag}(q_1^{-1}, q_2^{-1}, \dots, q_I^{-1}). \quad (3.51)$$

Define the r th order differencing matrix, D , as the $(I - r) \times I$ matrix with (i, j) element

$$d_{ij} = (-1)^{i+j} \binom{r}{j-i}, \quad j \geq i, \quad j-i \leq r$$

$$= 0, \quad \text{otherwise.} \quad (3.52)$$

The procedure requires differencing matrices associated with both the (logged) case estimate ratios (D_Q , an $(I - r) \times I$ matrix, defined as in Equation (3.52)) and $\tilde{D}_W$ which is defined as

$$\tilde{D}_W = \begin{pmatrix} D^T D & & 0 \\ & \ddots & \\ 0 & & D^T D \\ \text{(H diagonal blocks)} \end{pmatrix} \quad (3.53)$$

where D is as defined in Equation (3.52).

A reasonable choice for the level of differencing is second order ($r = 2$), for both the case estimate ratios and the weights.

Let E be the $HI \times HI$ matrix which permutes the components of w as follows:

$$\begin{bmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{I1} \\ \vdots \\ w_{IH} \end{bmatrix} \rightarrow \begin{bmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{1H} \\ \vdots \\ w_{IH} \end{bmatrix}. \quad (3.54)$$

Let

$$R = \log(QL) \quad (3.55)$$

and

$$B = R^T D_Q^T D_Q R. \quad (3.56)$$

Now let

$$T = E^T \tilde{D}_w E \quad (3.57)$$

and

$$G = C + k_1 B + k_2 T, \quad (3.58)$$

which is an objective function that assigns relative weights 1, k_1 and k_2 to

- variance of total liability according to the blend of models;
- smoothness of case estimate ratios (more precisely, ratios of logged estimates to logged case estimates); and
- smoothness of the weights;

respectively. Values of $k_1, k_2 \geq 0$ are to be chosen by the user, and the choice may be made with the assistance of Taylor (1992) and Verrall (1993).

The weights that minimize this objective function, subject to the constraint that $\sum_h w_{ih} = 1$ for each i , are

$$w = G^{-1} U^T (U G^{-1} U^T)^{-1} \mathbf{1} \quad (3.59)$$

where

$$U = \begin{bmatrix} \mathbf{1}^T & & & \\ 1 \times H & & & \\ & \mathbf{1}^T & & \\ & 1 \times H & & \\ & & \ddots & \\ & & & \mathbf{1}^T \\ & & & 1 \times H \end{bmatrix}. \quad (3.60)$$

The matrix U arises here from the constraint that the weights for each accident period must sum to unity.

There is no constraint for the weights to be non-negative in the algorithm above; all that is required is for the weights to sum to one for each accident period. It is possible for the algorithm to lead to large, negative weights balanced by large positive weights. This may be seen as undesirable. Therefore, it is proposed that negative weights be forced to zero, with the remaining positive weights being scaled so that they sum to one.

Further, the definition of C in Equation (3.50) may be based upon variances only rather than the full covariance matrices for each model. Restricting C to be a strictly diagonal matrix (rather than block diagonal—where covariances between different accident periods may be nonzero within a model, but covariances between models are assumed zero) can lead to quite different results. In some cases, these may be judged better by the practitioner.

In practice, both the log and linear versions of the blending algorithm should be considered. The linear version of the algorithm follows the same process as outlined above except that Equations (3.55) and (3.56) are replaced by

$$A = Q^T D_Q^T D_Q Q \quad (3.61)$$

and

$$B = \hat{L}^T A \hat{L}. \quad (3.62)$$

The version that yields the best results for the problem under consideration should be used.

4. Models applied to data

It has been previously indicated that it is customary to consider fitting more than one type of model to claims data when calculating outstanding claims reserves. Three models are considered in this paper, representing to a certain extent the standard repertoire in Australian practice:

- Payments per claims incurred (PPCI);
- Payments per claims finalized (PPCF); and
- Projected case estimates (PCE).

The rationale behind this is that the first two models are payments based. These should perform well where there is a substantial volume of payments (e.g., in early to middle development periods). The PPCF model is particularly suited to claims for which the majority of the payments are made at finalization (and conversely less suited to claims with substantial recurring payments). The PCE model would be expected to perform best in the later development periods. Here, much is known about outstanding claims, so the case estimates would be expected to be reliable.

Definition and discussion of the three models appears in Taylor (2000). Some more detail on each of them follows below. Note that the notation used here is that defined in Sections 2 and 3 (with the exception of Section 3.5).

4.1. Payments per claim incurred model

Let

$$\text{PPCI}_{ij}^L = C_{ij}^L / U_i^L. \quad (4.1)$$

A commonly used model for PPCI_{ij} is the Hoerl Curve (De Jong and Zehnwirth 1983; Wright 1990). This has the following form:

$$\mu_{ij} = \exp\{\beta_0 + \beta_1 j + \beta_2 \log j\}, \quad j = 1, 2, \dots \quad (4.2)$$

where μ_{ij} represents the mean value of PPCI_{ij} .

For the particular data set described in Section 2, the Hoerl Curve is used as a starting point and then modified slightly. Further, each jurisdictional group is permitted to have its own shape. The applied model has the form

$$\begin{aligned} \text{PPCI}_{ij}^L &\sim \text{Gamma}(\mu_{ij}^L, v_j) \\ \mu_{ij}^L &= \exp\{\beta_{i0} + \beta_{i1} \log j + \beta_{i2} j + \beta_{i3} \min(j, 16)\} \\ &\quad + I(L=1)\{\beta_{i4} + \beta_{i5} \log j + \beta_{i6} j + \beta_{i7} \min(j, 16)\} \end{aligned} \quad (4.3)$$

where

v_j = coefficient of variation in development quarter j ; and

$I(\text{condition})$ is an indicator function which is 1 if condition is true and 0 otherwise.

The quantity v_j may be estimated on the basis of data and actuarial judgment.

Examples of the actual data and fitted model are shown below in Figure 1, which plots PPCIs for one of the jurisdictional groups (labeled as group 0). Note that both graphs are presented on the same vertical scale. Thus, movement in average claim payments from accident year 1 to 5 is apparent in these plots.

4.2. Payments per claims finalized model

The payments per claims finalized model (PPCF) actually consists of two submodels:

- Average payments per claim finalized;
- The probability that a claim finalizes in a particular quarter.

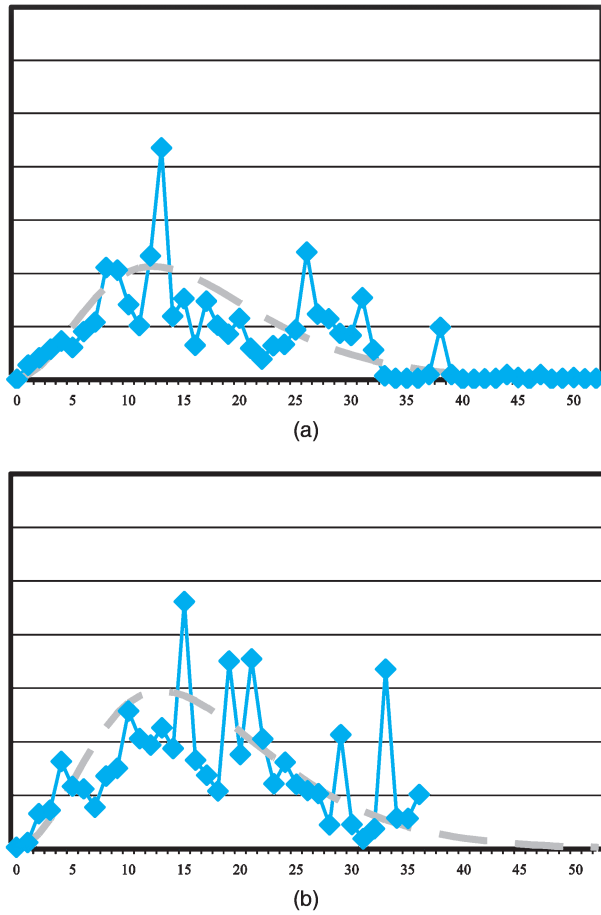
Model of average payments per claim finalized

Let

$$\text{PPCF}_{ij}^L = C_{ij}^L / F_{ij}^L. \quad (4.4)$$

Average payments per finalized claim tend to increase with age of claim since the more com-

Figure 1. Fitted PPCI model



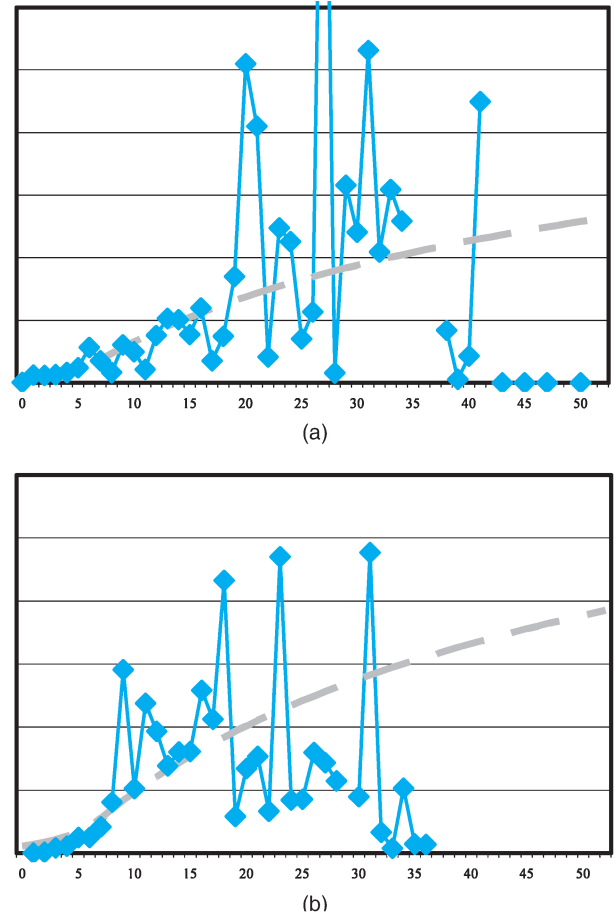
plicated and serious claims are more likely to take longer to settle. For the data in this paper, the following model is used:

$$\begin{aligned} \text{PPCF}_{ij}^L &\sim \text{Gamma}(\mu_{ij}^L, v_j) \\ \mu_{ij}^L &= \exp \left\{ \beta_{i0} + \mathbf{I}(L=0) \left[\beta_{i1} \left(\frac{j-1}{4} \right)^{-0.4} \right. \right. \\ &\quad \left. \left. + \beta_{i2} \min(6, j-1) \right] \right. \\ &\quad \left. + \mathbf{I}(L=1) \left[\beta_{i3} + \beta_{i4} \left(\frac{j-1}{4} \right)^{-0.75} \right. \right. \\ &\quad \left. \left. + \beta_{i5} \min(6, j-1) \right] \right\} \end{aligned} \quad (4.5)$$

where

μ_{ij}^L = the mean PPCF_{ij}^L in cell (i, j) ;

Figure 2. Fitted PPCF model



v_j = coefficient of variation in development quarter j ; and

modeling begins in the second development quarter (i.e., $j = 2$).

Examples of the actual data and fitted model are shown in Figure 2. Again, both graphs are presented on the same vertical scale.

Model of probability of finalization

Let

$$\text{PRF}_{ij}^L = F_{ij}^L / (O_{ij}^L + N_{ij}^L / 3). \quad (4.6)$$

The denominator in (4.6) is an exposure measure of the number of claims that may be finalized in cell (i, j) . The probability of finalization for the data in this paper is quite a volatile quantity; therefore simple models have been used.

$\text{PRF}_{ij}^L \sim \text{Overdispersed Poisson}(\mu_{ij}^L, v_j)$

$$\begin{aligned} \mu_{ij}^L = \exp\{ & \beta_{i0} + \mathbf{I}(L = 0)[\beta_{i1} \min(j - 1, 15) \\ & + \beta_{i2} \max(1, j - 13)^{-0.25}] \\ & + \mathbf{I}(L = 1)[\beta_{i3} + \beta_{i4} \min(j - 1, 15) \\ & + \beta_{i5} \max(1, j - 13)^{-0.25}] \} \end{aligned} \quad (4.7)$$

where

μ_{ij}^L = the mean PRF_{ij}^L in cell (i, j) ; and

v_j = ratio of the variance to the mean, which may differ from 1 (standard Poisson) through the use of an overdispersed Poisson.

Examples of the actual data and fitted model are shown in Figure 3. Again, both graphs are presented on the same vertical scale.

4.3. Projected case estimates model

The projected case estimates model examines the further development required by case estimates at a given point in time to be sufficient to settle claims in full. In any given development period and accident period, the hindsight estimate of case estimates at the end of the quarter (the sum of payments in that quarter and closing case estimates) may be compared against opening case estimates. If the opening case estimates were sufficient, then the ratio would be one; a value greater than one indicates that the opening estimates are now considered insufficient, while a value less than one indicates they are now considered excessive.

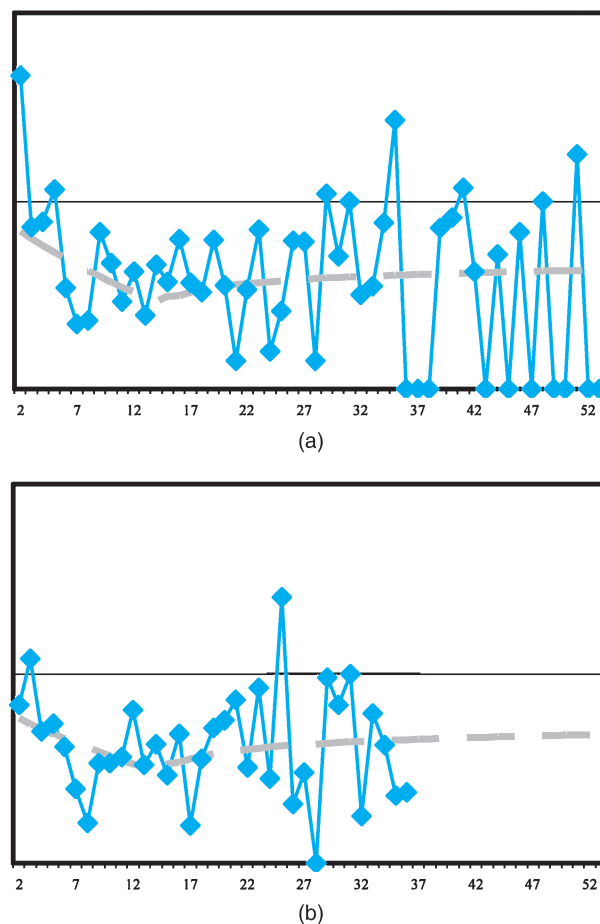
Therefore, there are two quantities to be modeled—case estimates and payments. Both may be expressed as factors relative to the opening case estimates.

Payment factors

Payment factors are defined, for $j = 2, 3, \dots$, as

$$\text{PF}_{ij}^L = C_{ij}^L / E_{i,j-1}^L. \quad (4.8)$$

Figure 3. Modeled finalization probabilities



The PF_{ij}^L are modeled as follows:

$\text{PF}_{ij}^L \sim \text{Gamma}(\mu_{ij}^L, v_j)$

$$\begin{aligned} \mu_{ij}^L = \exp\{ & \beta_{i0} + \mathbf{I}(L = 0)[\beta_{i1} \max(1, j - 9)^{-0.5} + \beta_{i2} \mathbf{I}(j \leq 9)] \\ & + \mathbf{I}(L = 1)[\beta_{i3} + \beta_{i4} (j - 1)^{-0.5} \\ & + \beta_{i5} \max(0, 7 - j)^2] \} \end{aligned} \quad (4.9)$$

where

μ_{ij}^L = the mean PF_{ij}^L in cell (i, j) ; and

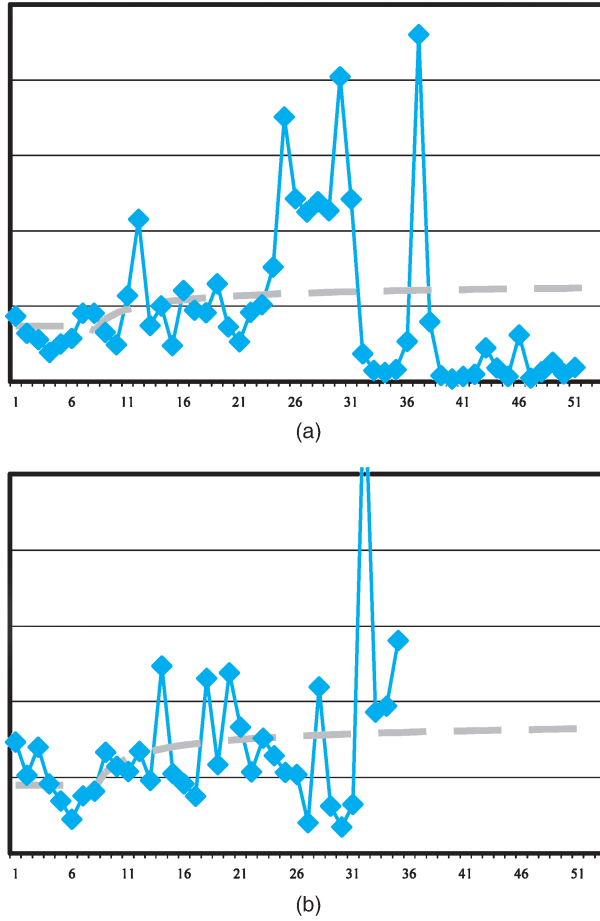
v_j = coefficient of variation of the payment factors in development quarter j .

Examples of the actual data and fitted model are shown in Figure 4.

Case estimate development factors

Traditionally, the case estimate development factors model is based on the development of

Figure 4. Modeled payment factors



the hindsight case estimates. Thus, the modeled quantities are defined, for $j = 2, 3, \dots$ as

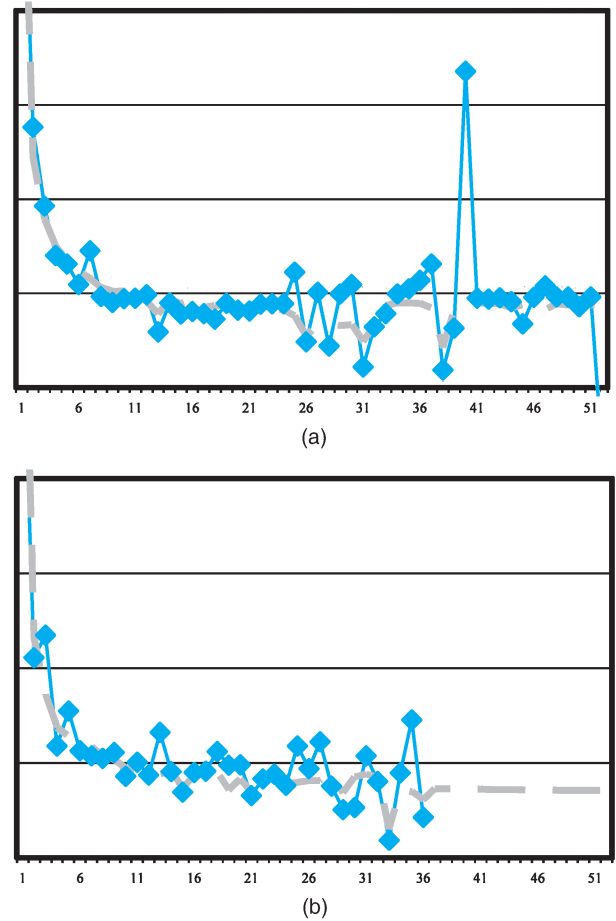
$$\text{CEDF}_{ij}^L = (E_{ij}^L + C_{ij}^L) / E_{i,j-1}^L. \quad (4.10)$$

However, comparison with Equation (4.8) indicates that, based on this definition, the $\{\text{CEDF}\}$ and $\{\text{PF}\}$ would not be independent. Although this is inconsequential for the traditional deterministic application of the PCE model, it does matter for the stochastic application. Therefore, in this paper, the following definition of a case estimate development factor is used:

$$\text{CEDF}_{ij}^L = E_{ij}^L / E_{i,j-1}^L. \quad (4.11)$$

To preserve the relationship with amount paid in a cell (i, j) , the payment factors are includ-

Figure 5. Modeled case estimate development factors



ed within the model for the $\{\text{CEDF}\}$. Thus, the model is

$$\begin{aligned} \text{CEDF}_{ij}^L &\sim \text{Gamma}(\mu_{ij}^L, v_j) \\ \mu_{ij}^L &= \exp\{\beta_{i0} + \beta_{i1} \text{PF}_{ij}^L \\ &\quad + \mathbf{I}(L=0)[\beta_{i1} j^{-1} + \beta_{i2} \mathbf{I}(j=1)] \\ &\quad + \mathbf{I}(L=1)[\beta_{i3} + \beta_{i4} j^{-1} \\ &\quad + \beta_{i5} \mathbf{I}(j=1)]\} \end{aligned} \quad (4.12)$$

where

μ_{ij}^L = the mean CEDF_{ij}^L in cell (i, j) ; and
 v_j = coefficient of variation in development quarter j .

Examples of the actual data and fitted model are shown in Figure 5.

Bootstrapping the case estimate development factor submodel

The model for the {CEDF} is different from the other four models described in that it depends on a modeled quantity, the {PF}. Allowance for this must be made within the bootstrapping process to allow correctly for variability within the {CEDF}. The following bootstrap process is proposed:

For each bootstrap iteration, k ,

- Take the past pseudo payment factors from one bootstrapping iteration of payment factors;
- First, model the actual {CEDF} values using these pseudo payment factors for the actual payment factors. This yields a model for the {CEDF} that is different from the original fitted model through its dependency on the pseudo payment factors;
- Now bootstrap the modeled {CEDF} as described in Section 3.4, yielding one set of pseudo {CEDF}. Together with the pseudo {PF}, these produce the liability from one pseudo data set;
- Repeat this process, starting by taking the next set of past pseudo payment factors from the payment factor bootstrap until the desired number of pseudo results has been obtained.

5. Results

The models described in Section 4 were applied to the data using the GLM adaptive filtering methodology described in Section 3.3. The bootstrap process outlined in Section 3.4 was applied. The results of this were blended to yield an overall estimate of outstanding claims liability following the algorithm in Section 3.5.

Table 1 displays the results for each of the three models for legislative jurisdiction 0, together with the bootstrapped estimates of coefficient of variation. Note that the results have been scaled for confidentiality reasons. The PPCF and PCE models perform best as measured by coeffi-

Table 1. Results for legislative jurisdiction 0

Accident year	PPCI		PPCF		PCE	
	Mean	CV	Mean	CV	Mean	CV
1	8	240%	132	55%	22	105%
2	20	213%	242	47%	56	108%
3	58	169%	165	58%	23	98%
4	110	132%	268	47%	70	90%
5	242	107%	861	30%	317	62%
6	291	74%	1,216	27%	671	64%
7	678	57%	1,257	27%	799	44%
8	817	52%	1,672	27%	1,319	40%
9	2,259	48%	3,366	25%	2,040	32%
10	3,544	48%	3,510	22%	2,368	31%
11	6,366	48%	6,041	21%	5,480	31%
12	7,182	44%	6,742	20%	6,700	31%
13	8,544	43%	8,664	21%	7,234	33%
14	9,001	43%	9,015	21%	3,749	98%
Total ex 14	30,119		34,136		27,099	
Total	39,120	42%	43,151	18%	30,849	22%

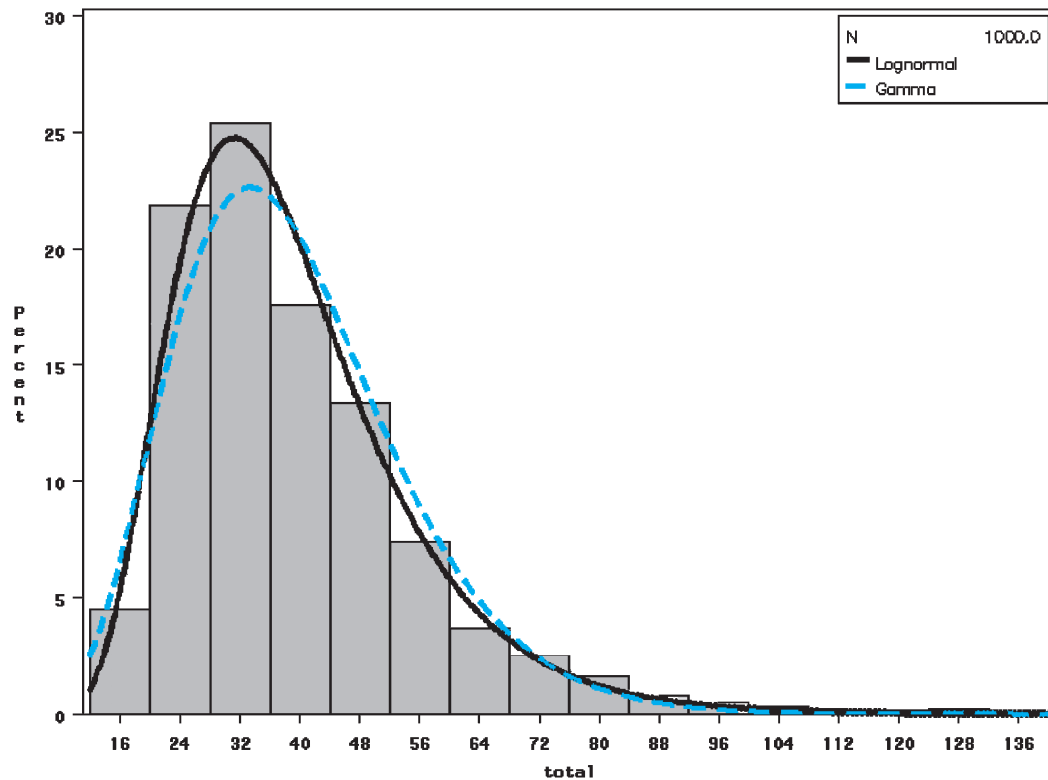
Table 2. Results for legislative jurisdiction 1

Accident year	PPCI		PPCF		PCE	
	Mean	CV	Mean	CV	Mean	CV
1	45	213%	6	232%	0	0%
2	118	214%	146	58%	215	148%
3	160	177%	36	116%	0	0%
4	142	194%	236	45%	72	112%
5	239	153%	226	48%	52	73%
6	334	116%	365	40%	417	60%
7	446	106%	686	34%	565	54%
8	645	89%	780	34%	519	53%
9	1,859	84%	1,430	27%	1,303	48%
10	5,795	83%	6,470	24%	6,225	55%
11	8,517	78%	6,219	21%	6,923	59%
12	9,477	78%	6,580	20%	5,536	65%
13	9,828	76%	7,716	21%	5,713	74%
14	13,385	76%	10,826	21%	1,827	128%
Total ex 14	37,602		30,895		27,539	
Total	50,988	73%	41,721	19%	29,366	45%

cient of variation. For these data, the three models give broadly consistent results when the most recent accident year is excluded.

Table 2 displays the results for the other legislative grouping of data. It is seen that the PPCF and PCE models perform best for these data. The PPCI model with universally high coefficients of variation performs poorly. These data show changes in the rates of finalization in recent years, which changes the profile of average

Figure 6. Distribution of PPCI results—jurisdiction 0



payments paid per development period, which in turn, impacts on the reliability of the PPCI model.

Further, the results of the PPCF and PCE models are quite similar (ignoring the most recent year—PCE models generally perform poorly on recent accident periods) while the PPCI model gives substantially higher results.

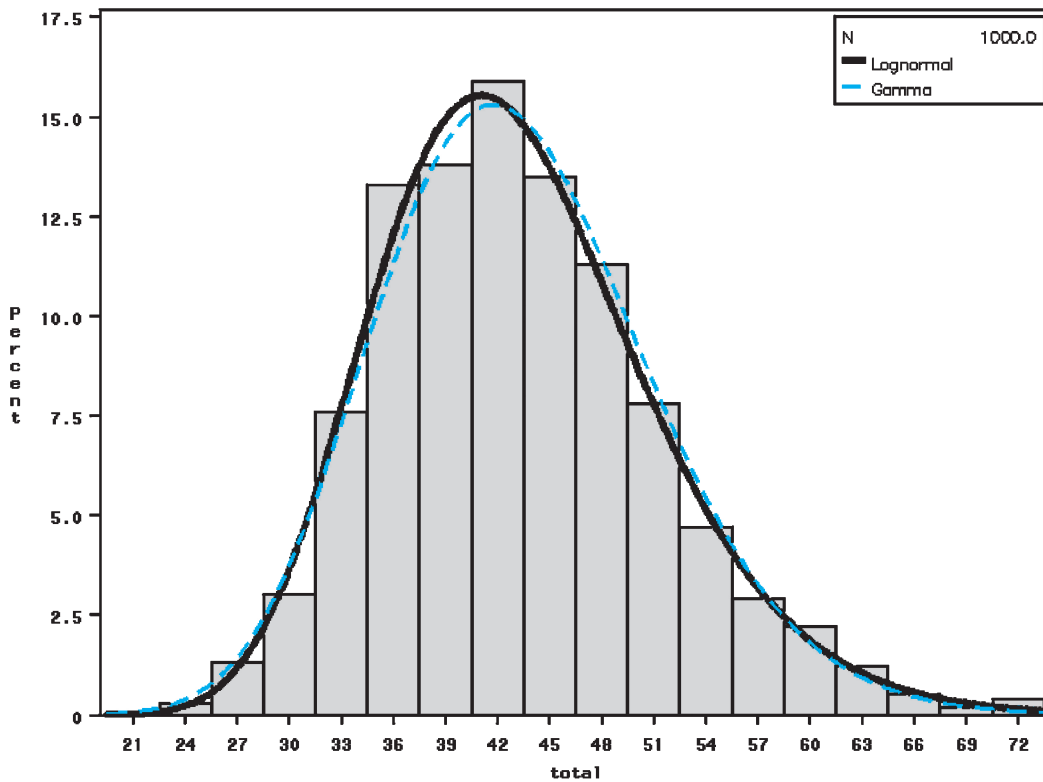
Histograms of the bootstrapped distributions for each of the models, each with 1,000 replications, are presented in Figures 6, 7, and 8. These graphs all relate to jurisdiction 0.

In the application of the blending algorithm in Section 3.5, the following decisions were made:

- Both log and linear blending were considered; for these data the log blending was found to perform the better;
- Only variances were used in the blending algorithm (i.e., the covariance matrix in Equation (3.50) was assumed to be strictly diagonal);

- The progression of weights, and the ratio of blended liability to case estimates, were smoothed over accident years 1 to 13 only. For year 14, the minimum variance unbiased estimates were used since the case estimates for the most recent year are considerably lower than those in previous years, due to the immaturity of that experience (only one quarter into the accident year). Using them in the smoothing algorithm would have led to distorted results;
- Negative weights were not allowed (see the discussion at the end of Section 3.5);
- Values of $k_1 = k_2 = 10^{14}$ were selected. These values were selected by first choosing values of similar order to the matrices B and T in (3.58) and then examining plots of case estimate ratios and weights for a suitably smooth progression (e.g., Figures 9 to 11).

The weights that result from the algorithm are shown in Figure 9 (jurisdiction 0) and Figure 10

Figure 7. Distribution of PPCF results—jurisdiction 0

(jurisdiction 1). The omission of smoothing in the last year is evident from the graphs. As would be expected from the results in Table 1 and 2, the PPCI model is assigned low weights in both models.

A final point to note is the relatively high weight assigned to the PCE model for accident year 14, particularly for jurisdictional grouping 1. Although the PCE results for year 14 have a high coefficient of variation (refer to Tables 1 and 2), the quantum of liability is small relative to the other models, meaning, in turn, that the standard deviation in absolute terms is low.

The practitioner may wish to intervene to increase the standard deviation to a level perceived as realistic relative to the other models. This has not been done here, however.

The ratio of the blended liability (that results from the weights above) to case estimates is shown in Figure 11, in which the horizontal axis relates to accident years. Given the immaturity of

the case estimates in year 14 (the most recent), this comparison has been omitted. It is seen that the progression is satisfactory. Missing ratios exist in Years 1 and 3 in jurisdiction 1, as case estimates are zero in these years.

The final results are given in Table 3. It is observed that the coefficients of variation are generally lower than the individual coefficients of variation from each component model, particularly for the coefficient of variation for the total liability.

6. Discussion

The practicalities of the application of the analytical GLM filter (Taylor 2008) to claims data have been described in this paper. Also discussed are appropriate methods for bootstrapping filtered data as well as algorithms for blending the results from two or more models.

The filter described here is applied on a row-by-row basis. In insurance terms, this means that

Figure 8. Distribution of PCE results—jurisdiction 0

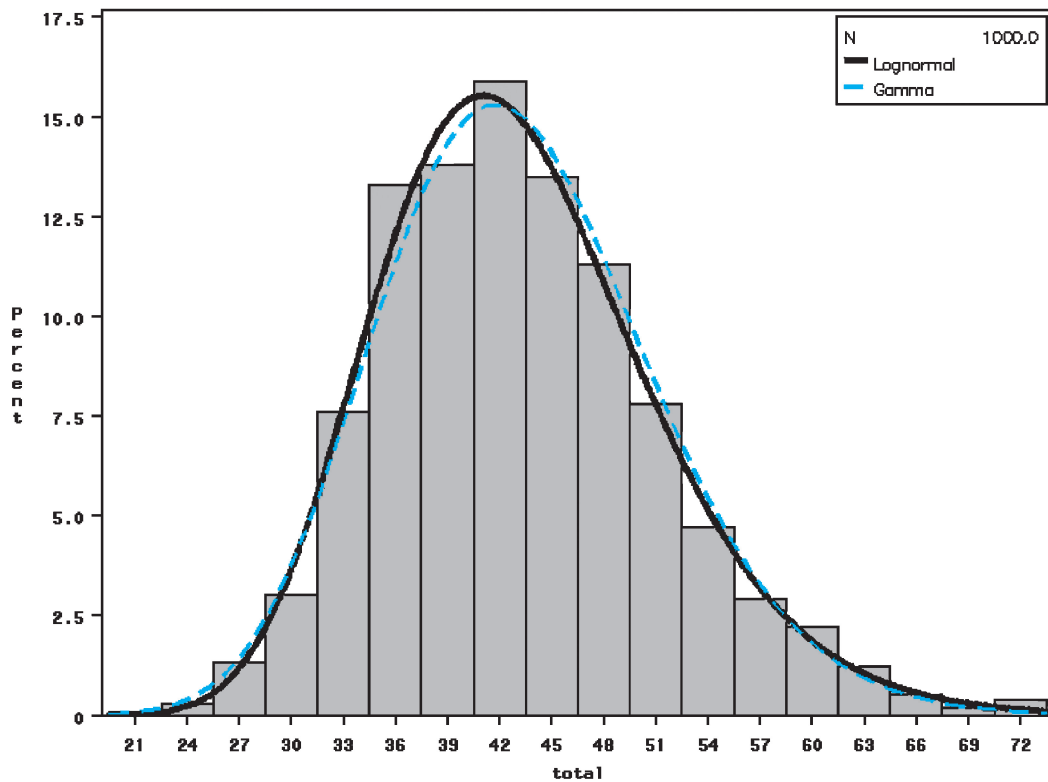
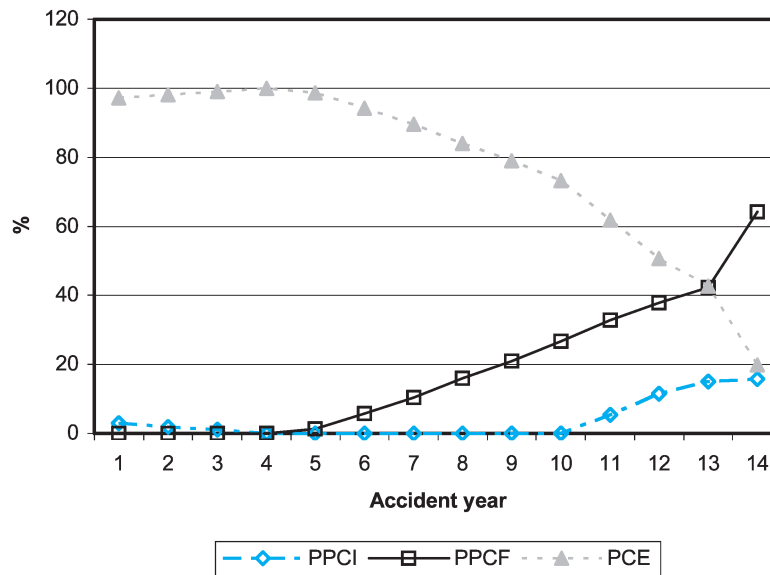
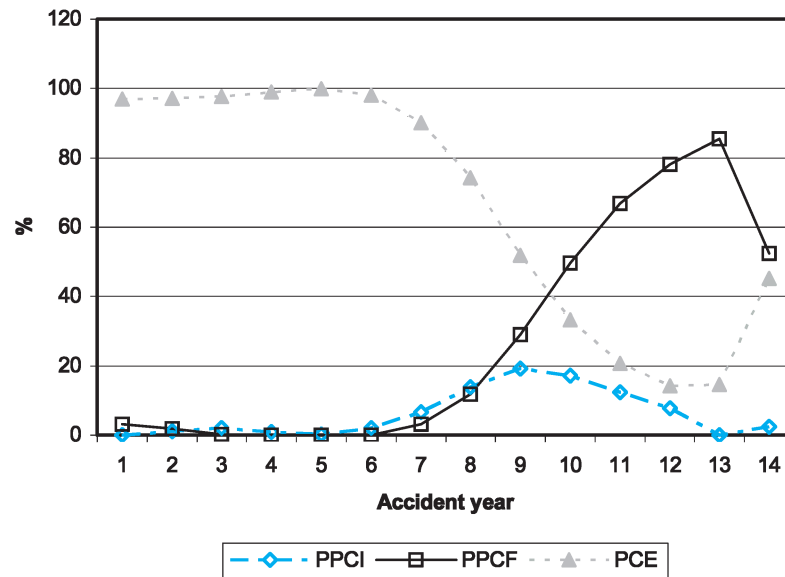


Figure 9. Selected weights for legislative jurisdiction 0



accident period changes may be incorporated easily into the model, but there is no easy way to incorporate diagonal effects (e.g., calendar period effects such as superimposed inflation). Fur-

ther, data naturally arise on a calendar-year basis. Therefore, an implementation of the filter by calendar period would be useful. In principle, this should be relatively straightforward [e.g., see the

Figure 10. Selected weights for legislative jurisdiction 1**Table 3. Blended results**

Accident year	$L = 0$		$L = 1$	
	Mean	CV	Mean	CV
1	22	104%	0	20051%
2	56	107%	213	145%
3	24	96%	3	1413%
4	70	90%	73	110%
5	324	60%	52	72%
6	702	58%	415	59%
7	847	38%	561	49%
8	1,375	32%	567	39%
9	2,317	24%	1,447	32%
10	2,672	21%	6,273	26%
11	5,712	20%	6,649	22%
12	6,771	18%	6,655	20%
13	8,035	17%	7,425	20%
14	7,963	20%	6,811	23%
Total	36,891	12.7%	37,144	16.6%

description of the Kalman filter by calendar period in Taylor (2000)]. However, in practice, a number of issues may need to be addressed, such as stability of parameter estimates. This is the subject of future work.

Such methodology has potential use as part of an automatic reserving process. Within the general insurance industry, many valuations take place on a “revolving door” basis, i.e., one valuation finishes only for the next one to begin.

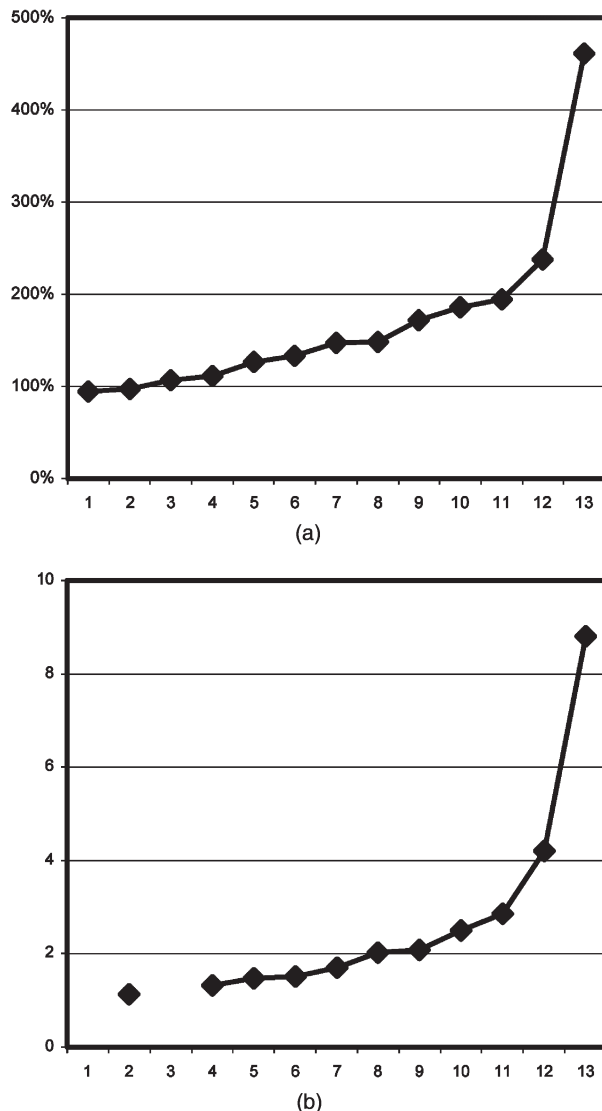
Since filters adapt to changing experience, an automated process could be set up which would apply filters, run the bootstrap and blending process, and produce final results, without human intervention. This would lead to considerable savings in the valuation process while still producing fully stochastic models.

An automatic approach would require a range of diagnostic tests to alert users to problems which require intervention. Developing such tests is the subject of current work.

The analytical GLM filter implements an approximate solution to a Bayesian updating problem, and uses bootstrapping to estimate the distribution of predicted liabilities. More exact results may be found using a method like MCMC which can generate Bayesian posterior distributions using simulation. A comparison between the filter presented here and more exact approaches, including MCMC, will be presented in a separate paper.

The models implemented here are dynamic models, rather than static models. This means that their parameters evolve through time and that the estimates of those parameters can change through time in response, should this be war-

Figure 11. Ratio of blended results to case estimates for $L = 0$ (left graph) and $L = 1$ (right graph)



ranted by experience. An example of this movement is given in Figure 12, where the PPCI model fitted to the first accident year is compared with the models for a number of following accident years.

The amount of movement permitted in the fitted curves is dependent on the “parameter variances,” defined in Equation (3.19) (gamma case of the Analytical GLM filter) and Equation (3.30) (Poisson case). The choice of parameter variances may be thought of as picking a point along a continuum of levels from variances of zero

(which corresponds to a static model) to high variances (each accident period modeled on just its own data, without any reference to other periods).

The selection of the variance levels is currently based on judgment; developing a more objective way to select the variances is currently being investigated.

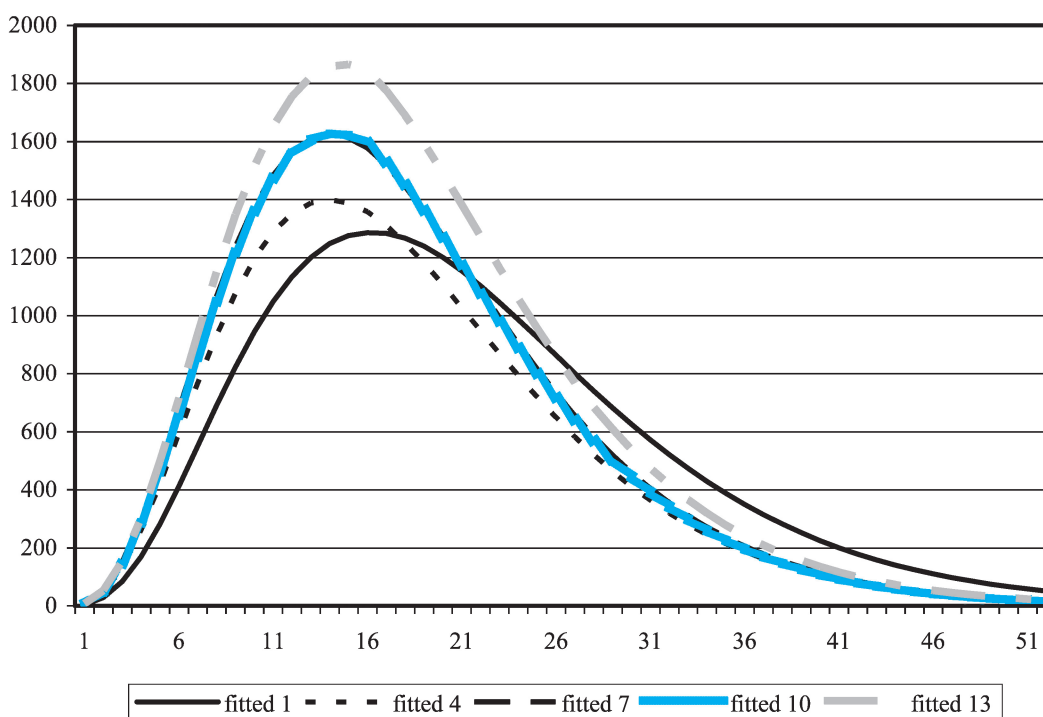
Acknowledgments

We would like to thank Insurance Australia Group for their support of this work.

References

- de Alba, E., “Bayesian Estimation of Outstanding Claim Reserves,” *North American Actuarial Journal* 6:1, 2002, pp. 1–20.
- De Jong, P., and B. Zehnwirth, “Claims Reserving State Space Models and the Kalman Filter,” *Journal of the Institute of Actuaries* 110, 1983, pp. 157–181.
- Efron, B., and R. J. Tibshirani, *An Introduction to the Bootstrap*, New York: Chapman and Hall, 1993.
- England, P. D., and R. J. Verrall, “Stochastic Claims Reserving in General Insurance,” *British Actuarial Journal* 8, 2002, pp. 443–518.
- Hastings, W. K., “Monte Carlo Sampling Methods Using Markov Chains and their Application,” *Biometrika* 57, 1970, pp. 97–109.
- Kalman, R. E., “A New Approach to Linear Filtering and Prediction Problems,” *Journal of Basic Engineering* 82, 1960, pp. 340–345.
- McCullagh, P., and J. A. Nelder, *Generalised Linear Models* (3rd ed.), New York: Chapman and Hall, 1989.
- Nelder, J. A., and R. W. M. Wedderburn, “Generalized Linear Models,” *Journal of the Royal Statistical Society, Series A* 135, 1972, pp. 370–384.
- Ntzoufras, I., and P. Dellaportas, “Bayesian Modelling of Outstanding Liabilities Incorporating Claim Count Uncertainty,” *North American Actuarial Journal* 6:1, 2002, pp. 113–136.
- Pinheiro, P. J. R., J. M. Andrade e Silva, and M. de Lourdes Centeno, “Bootstrap Methodology in Claim Reserving,” *Journal of Risk and Insurance* 70, 2003, pp. 701–714.
- Scollnik, D. P. M., “Actuarial Modelling with MCMC and BUGS,” *North American Actuarial Journal* 5:2, 2001, pp. 96–125.
- Scollnik, D. P. M., “Regression Models for Bivariate Loss Data,” *North American Actuarial Journal* 6:4, 2002a, pp. 67–80.

Figure 12. Dynamic GLM



Scollnik, D. P. M., "Implementation of Four Models for Outstanding Liabilities in WinBUGS: A discussion of a paper by Ntzoufras and Dellaportas." *North American Actuarial Journal* 6:1, 2002b, pp. 128–136.

Smith, A. F. M., and G. O. Roberts, "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Method," *Journal of the Royal Statistical Society, Series B* 55, 1993, pp. 3–24.

Stoffer, D. S., and K. D. Wall, "Bootstrapping State=Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter," *Journal of the American Statistical Association* 86:416, 1991, pp. 1024:1033.

Taylor, G., "Combination of Estimates of Outstanding Claims in Non-Life Insurance," *Insurance: Mathematics and Economics* 4, 1985, pp. 81–91.

Taylor, G., "A Bayesian Interpretation of Whittaker-Henderson Graduation," *Insurance: Mathematics and Economics* 11, 1992, pp. 7–16.

Taylor, G., *Loss Reserving: An Actuarial Perspective*, New York: Kluwer Academic Publishers, 2000.

Taylor, G., "Second Order Bayesian Revision of a Generalised Linear Model," *Scandinavian Actuarial Journal*, 2008, pp. 202–242.

Tierney, L., "Markov Chains for Exploring Posterior Distributions" (with discussion), *The Annals of Statistics* 22, 1994, pp. 1701–1762.

Tweedie, M. C. K., "An Index which Distinguishes Between Some Important Exponential Families" Appears in *Statistics: applications and new directions. Proceedings of the Indian Statistical Golden Jubilee International Conference*, (eds. J. K. Ghosh and J. Roy), Indian Statistical Institute, 1984, pp. 579–604.

Verrall, R. J., "A State Space Formulation of Whittaker Graduation with Extensions," *Insurance: Mathematics and Economics* 13, 1993, pp. 7–14.

West, M. P., P. J. Harrison, and H. S. Migon, "Dynamic Generalised Models and Bayesian Forecasting" (with discussion), *Journal of the American Statistical Association* 80, 1995, pp. 73–97.

Wright, T. S., "A Stochastic Method for Claims Reserving in General Insurance," *Journal of the Institute of Actuaries* 117, 1990, pp. 677–731.

Appendix A. GLM filter

A.1. Gamma GLM filter

Let $1/h^{-1}(s | s - j)$ and $\Gamma(s | s - j)$ be as defined in Equation (3.17). Further, let $h^{-1}(s | s - j)$ be the reciprocal of $1/h^{-1}(s | s - j)$. The filter begins with equations (3.18) and (3.19).

Define $M(s | s - 1)$ and $Q(s | s - 1)$ as the orthogonal and diagonal matrices, respectively, sat-

isfying

$$\begin{aligned} M(s | s-1)^T Q(s | s-1) M(s | s-1) \\ = \text{Var}[\beta(s | s-1)]. \end{aligned} \quad (\text{A.1})$$

It is then possible to calculate Equations (3.20) and (3.21).

Define $N(s | s-1)$, $W(s | s-1)$ as

$$\begin{aligned} N(s | s-1) &= [\Gamma(s | s-1)]^{-1} [I + \frac{1}{2} Q(s | s-1)] \text{DIAG} \\ &\times \exp\{-M(s | s-1) E[\beta(s | s-1)]\}, \end{aligned} \quad (\text{A.2})$$

$$W(s | s-1) = N(s | s-1) \left[\frac{1}{h^{-1}(s | s-1)} \right]. \quad (\text{A.3})$$

Then

$$\begin{aligned} \beta(s | s-1) \\ = -M(s | s-1)^T \log(N(s | s-1)^{-1} W(s | s-1)). \end{aligned} \quad (\text{A.4})$$

Further define:

$$\begin{aligned} B(s | s-j) &= -\text{DIAG} \exp\{-M(s | s-j) \beta(s | s-1)\}, \\ j &= 0, 1, \end{aligned} \quad (\text{A.5})$$

$$G(s) = -\text{DIAG} \exp\{-X(s) \beta(s | s-1)\}. \quad (\text{A.6})$$

and $P(s)$ and $D(s)$ as the orthogonal and diagonal matrices, respectively, satisfying

$$\begin{aligned} P(s)^T D(s) P(s) \\ = N(s | s-1) B(s | s-1) \\ + M(s | s-1) X(s)^T \text{DIAG}[G(s) \Lambda(s) Y(s)] \\ \times X(s) M(s | s-1)^T \end{aligned} \quad (\text{A.7})$$

Then

$$M(s | s) = P(s) M(s | s-1), \quad (\text{A.8})$$

and

$$\begin{aligned} J(s) &= P(s)^T B(s | s) P(s) \\ &\times \left[\begin{aligned} &N(s | s-1) B(s | s-1) + M(s | s-1) X(s)^T \\ &\text{DIAG}\{G(s) \Lambda(s) Y(s)\} X(s) M(s | s-1)^T \end{aligned} \right]^{-1} \\ &\times M(s | s-1) X(s)^T \Lambda(s). \end{aligned} \quad (\text{A.9})$$

Equation (3.22) may then be calculated, followed by Equation (3.23). The iteration of the filter then concludes with Equations (3.24) to (3.27).

A.2. Poisson GLM filter

Let $h^{-1}(s | s-j)$ and $\Gamma(s | s-j)$ be as defined in Equation (3.17). The filter begins with Equations (3.29) and (3.30).

Define $M(s | s-1)$ and $Q(s | s-1)$ as the orthogonal and diagonal matrices respectively satisfying

$$\begin{aligned} M(s | s-1)^T Q(s | s-1) M(s | s-1) \\ = \text{Var}[\beta(s | s-1)]. \end{aligned} \quad (\text{A.10})$$

It is then possible to calculate Equations (3.31) and (3.32).

Define $N(s | s-1)$, $W(s | s-1)$ as

$$\begin{aligned} N(s | s-1) &= [\Gamma(s | s-1)]^{-1} [I + \frac{1}{2} Q(s | s-1)] \text{DIAG} \\ &\times \exp\{M(s | s-1) E[\beta(s | s-1)]\}, \end{aligned} \quad (\text{A.11})$$

$$W(s | s-1) = N(s | s-1) [h^{-1}(s | s-1)]. \quad (\text{A.12})$$

Then

$$\begin{aligned} \beta(s | s-1) \\ = M(s | s-1)^T \log(N(s | s-1)^{-1} W(s | s-1)). \end{aligned} \quad (\text{A.13})$$

Further define:

$$\begin{aligned} B(s | s-j) &= \text{DIAG} \exp\{M(s | s-j) \beta(s | s-1)\} \\ j &= 0, 1, \end{aligned} \quad (\text{A.14})$$

$$\begin{aligned} G(s) &= \text{DIAG} \exp\{X(s) \beta(s | s-1)\}. \\ \end{aligned} \quad (\text{A.15})$$

and $P(s)$ and $D(s)$ as the orthogonal and diagonal matrices respectively satisfying

$$\begin{aligned} P(s)^T D(s) P(s) \\ = N(s | s-1) B(s | s-1) \\ + M(s | s-1) X(s)^T \Lambda(s) G(s) X(s) M(s | s-1)^T. \end{aligned} \quad (\text{A.16})$$

Then

$$M(s | s) = P(s)M(s | s - 1). \quad (\text{A.17})$$

And

$$\begin{aligned} J(s) &= P(s)^T B(s | s) P(s) [N(s | s - 1) \\ &\quad \times B(s | s - 1) + M(s | s - 1) X(s)^T \\ &\quad \Lambda(s) G(s) X(s) M(s | s - 1)^T]^{-1} \\ &\quad \times M(s | s - 1) X(s)^T \Lambda(s). \end{aligned} \quad (\text{A.18})$$

Equation (3.33) may then be calculated. The iteration of the filter then concludes with Equations (3.34) and (3.37).

Appendix B. Calculation of variances of GLM innovations

B.1. Definition of some common terms

Let $\hat{\beta} = \beta + \delta$ where $E[\delta] = 0$ and $\text{Var}(\delta) = V_\delta (= R(s))$ where $R(s)$ is that from Equation (3.3) and let $E(\beta) = \tilde{\beta}$, $\text{Var}(\beta) = V_\beta$. At each step, $\tilde{\beta}$ and V_β are assumed to be the current values from Equations (3.18) and (3.19). Further, let $\mu_{Y_i} = E[e^{X_i \beta}]$ where Y_i represents the i th element of $Y(s)$ and X_i is the i th row of the matrix X .

Using the approximation

$$E[g(X)] \approx g\left(\mu_X + \frac{1}{2}\sigma_X^2 \frac{g''(\mu_X)}{g'(\mu_X)}\right), \quad (\text{B.1})$$

where $\mu_X = E[X]$ and $\sigma_X^2 = \text{Var}[X]$, then

$$\mu_Y = \exp\{X_i \tilde{\beta} + \frac{1}{2} X_i V_\beta X_i^T\}. \quad (\text{B.2})$$

Further let

$$V_{\beta\delta} = V_\beta + V_\delta. \quad (\text{B.3})$$

B.2. Variance of gamma/log innovations

Equation (3.24) may be rewritten as

$$\begin{aligned} 1/h^{-1}(s | s) &= \exp\{-M(s | s)\beta(s | s - 1)\} \\ &\quad - P(s)J(s)[H_{(-1)}(s | s - 1)^{-1}Y(s) - 1], \end{aligned} \quad (\text{B.4})$$

and

$$H_{(-1)}(s | s - 1) = \text{DIAG} h^{-1}(X(s)\beta(s | s - 1)) \quad (\text{B.5})$$

(refer to Equations (6.51) and (6.52) of Taylor (2008)).

Rearranging (B.4), it is seen that the covariance of $1/h^{-1}(s | s) - \exp\{-M(s | s)\beta(s | s - 1)\}$ depends on the covariance of the quantity $H_{(-1)}(s | s - 1)^{-1}Y(s)$.

Note that

$$\begin{aligned} \text{Cov}(AB, CD) &= \text{Cov}[(\mu_A + \varepsilon_A)(\mu_B + \varepsilon_B), (\mu_C + \varepsilon_C)(\mu_D + \varepsilon_D)] \\ &= \mu_A \mu_C C_{BD} + \mu_A \mu_D C_{BC} + \mu_B \mu_C C_{AD} + \mu_B \mu_D C_{AC} \\ &\quad + \text{third and higher order terms in } \varepsilon_i \end{aligned} \quad (\text{B.6})$$

where $E[\varepsilon_I] = 0$, $I = A, B, C, D$ and $\mu_I = E[I]$ for $I = A, B, C, D$, and $C_{IJ} = \text{Cov}[I, J] = E[IJ]$.

Then, letting $H^{-1} = H_{(-1)}(s | s - 1)^{-1}$ and H_i^{-1} be the i th diagonal element of H^{-1} ,

$$\begin{aligned} \text{Cov}(H_i^{-1}Y_i, H_j^{-1}Y_j) &= \mu_{H_i^{-1}} \mu_{H_j^{-1}} C_{Y_i Y_j} + \mu_{H_i^{-1}} \mu_{Y_j} C_{H_i^{-1} Y_j} \\ &\quad + \mu_{Y_i} \mu_{H_j^{-1}} C_{H_i^{-1} Y_j} + \mu_{Y_i} \mu_{Y_j} C_{H_i^{-1} H_j^{-1}}. \end{aligned} \quad (\text{B.7})$$

Thus the following must be calculated: $\mu_{H_i^{-1}}$, μ_{Y_i} , $C_{Y_i Y_j}$, $C_{H_i^{-1} Y_j}$, $C_{H_i^{-1} H_j^{-1}}$. It will be necessary to use Equation (B.1) in these calculations.

$$\begin{aligned} \mu_{H_i^{-1}} &= E[e^{-X_i \hat{\beta}}] \\ &\approx \exp\left\{-X_i \tilde{\beta} + \frac{1}{2} X_i \text{Var}(\beta + \delta) X_i^T \frac{e^{-X_i \tilde{\beta}}}{-e^{-X_i \tilde{\beta}}}\right\} \\ &= \exp\{-X_i \tilde{\beta} - \frac{1}{2} X_i V_{\beta\delta} X_i^T\} \end{aligned} \quad (\text{B.8})$$

and μ_{Y_i} is as defined in Equation (B.2).

An approximation parallel to (B.1) is

$$\text{Cov}[f(X), Y] \approx f''(\mu_X) \text{Cov}[X, Y]. \quad (\text{B.9})$$

Then

$$\text{Cov}(H_i^{-1}, Y_j) = \frac{-1}{\mu_{H_i}^2} \text{Cov}(H_i, Y_j). \quad (\text{B.10})$$

Now

$$\begin{aligned}
 \text{Cov}(H_i, Y_j) &= \text{Cov}(e^{X_i\hat{\beta}}, e^{X_j\beta}) \\
 &= \text{Cov}(e^{X_i\beta + X_i\delta}, e^{X_j\beta}) \\
 &= E[e^{X_i\delta} e^{(X_i + X_j)\beta}] - E[e^{X_i\delta} e^{X_i\beta}] E[e^{X_j\beta}] \\
 &= E[e^{X_i\delta}] (E[e^{(X_i + X_j)\beta}] - E[e^{X_i\beta}] E[e^{X_j\beta}]) \\
 &= E[e^{X_i\delta}] \text{Cov}(e^{X_i\beta}, e^{X_j\beta}) \quad (\text{B.11})
 \end{aligned}$$

$$\begin{aligned}
 e^{X_i\delta} &\approx \exp \left\{ X_i(0) + \frac{1}{2} X_i V_\delta X_i^T \frac{e^{X_i\delta}}{e^{X_i\delta}} \right\} \\
 &= \exp \left\{ \frac{1}{2} X_i V_\delta X_i^T \right\} \quad (\text{B.12})
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}(e^{X_i\beta}, e^{X_j\beta}) &= E[e^{(X_i + X_j)\beta}] - E[e^{X_i\beta}] E[e^{X_j\beta}] \\
 &= \exp[(X_i + X_j)\tilde{\beta}] + \frac{1}{2} (X_i + X_j) V_\beta (X_i + X_j)^T \\
 &\quad - \mu_{Y_i} \mu_{Y_j} \quad (\text{B.13})
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}(Y_i, Y_j) &= \text{Cov}(e^{X_i\beta} + \varepsilon_i, e^{X_j\beta} + \varepsilon_j) \\
 &= E[e^{(X_i + X_j)\beta} + \varepsilon_i e^{X_j\beta} + \varepsilon_j e^{X_i\beta} + \varepsilon_i \varepsilon_j] \\
 &\quad - E[e^{X_i\beta}] E[e^{X_j\beta}] \\
 &= E[e^{(X_i + X_j)\beta}] + E[\varepsilon_i \varepsilon_j] - \mu_{Y_i} \mu_{Y_j} \\
 &= \exp[(X_i + X_j)\tilde{\beta}] + \frac{1}{2} (X_i + X_j) V_\beta (X_i + X_j)^T \\
 &\quad + \Delta_{i=j} \text{Var}(\varepsilon_j) - \mu_{Y_i} \mu_{Y_j} \quad (\text{B.14})
 \end{aligned}$$

$\Delta_{i=j}$ is 1 if $i = j$ and 0 otherwise. ε_i represents the data error where

$$Y_i = \exp\{X_i\beta\} + \varepsilon_i. \quad (\text{B.15})$$

$$\begin{aligned}
 \text{Var}[\varepsilon_i] &= \text{Var}[Y_i - \exp\{X_i\beta\}] \\
 &= \text{Var}[Y_i] + \text{Var}[\exp\{X_i\beta\}] \\
 &\quad - 2\text{Cov}[Y_i, \exp\{X_i\beta\}] \\
 &= \mu_Y^2 / \Lambda^2 + (E[\exp\{2X_i\beta\}] - E[\exp\{X_i\beta\}]^2) \\
 &\quad - 2(E[\exp\{2X_i\beta\}] - E[Y_i] E[\exp\{X_i\beta\}]) \\
 &= \mu_Y^2 / \Lambda^2 - E[\exp\{2X_i\beta\}] + \mu_Y^2 \\
 &= \mu_Y^2 / \Delta^2 - \exp\{2X_i\beta\} \{1 + 2X_i V_\beta X_i^T\} + \mu_Y^2 \quad (\text{B.16})
 \end{aligned}$$

using the approximation $E[f(x)] \approx f(\mu_x) + \frac{1}{2} \sigma_x^2 f''(\mu_x)$ and where Λ is the reciprocal of the

square of the coefficient of variation (refer back to Section 3.3.1).

Finally,

$$\begin{aligned}
 \text{Cov}(H_i^{-1}, H_j^{-1}) &= \text{Cov}(e^{-X_i\hat{\beta}}, e^{-X_j\hat{\beta}}) = \text{Cov}(e^{-X_i(\beta + \delta)}, e^{-X_j(\beta + \delta)}) \\
 &= E[e^{-(X_i + X_j)(\beta + \delta)}] - E[e^{-X_i(\beta + \delta)}] E[e^{-X_j(\beta + \delta)}] \\
 &= \exp\{-(X_i + X_j)\tilde{\beta} - \frac{1}{2} (X_i + X_j) V_{\beta\delta} (X_i + X_j)^T\} \\
 &\quad - \mu_{H_i^{-1}} \mu_{H_j^{-1}}. \quad (\text{B.17})
 \end{aligned}$$

It is now possible to calculate $\text{Cov}[H_{(-1)}(s | s - 1)^{-1} Y(s)]$. The covariance of $1/h^{-1}(s | s) - \exp\{-M(s | s)\beta(s | s - 1)\}$ may then be calculated using a rearranged Equation (B.4).

B.3. Variance of Poisson/log innovations

Substituting Equation (3.33) into Equation (3.34), it may be seen that

$$\begin{aligned}
 \text{Cov}[h^{-1}(s | s) - \exp\{M(s | s)\beta(s | s - 1)\}] \\
 &= \text{Cov}[P(s)J(s)[Y(s) - \exp\{X(s)\beta(s | s - 1)\}]] \quad (\text{B.18})
 \end{aligned}$$

meaning that it is necessary to calculate $\text{Cov}[Y(s) - \exp\{X(s)\beta(s | s - 1)\}]$.

With some abuse of notation, let $h^{-1} = \exp\{X(s)\beta(s | s - 1)\}$ and let h_i^{-1} be the i th component of h^{-1} . Then

$$\begin{aligned}
 \text{Cov}[Y_i - h_i^{-1}, Y_j - h_j^{-1}] &= E[(Y_i - h_i^{-1})(Y_j - h_j^{-1})] - E[(Y_i - h_i^{-1})] E[(Y_j - h_j^{-1})] \\
 &= \text{Cov}[Y_i, Y_j] + \text{Cov}[h_i^{-1}, h_j^{-1}] - \text{Cov}[h_i^{-1}, Y_j] - \text{Cov}[h_j^{-1}, Y_i] \quad (\text{B.19})
 \end{aligned}$$

$\text{Cov}[Y_i, Y_j]$ was previously defined in (B.14). The remainder of the terms are calculated as follows:

$$\begin{aligned}
 \mu_{h_i^{-1}} &= E[\exp\{X_i\hat{\beta}\}] \\
 &= E[\exp\{X_i(\beta + \delta)\}] \\
 &\approx \exp\{X_i\tilde{\beta} + \frac{1}{2} X_i V_{\beta\delta} X_i^T\}, \quad (\text{B.20})
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}[h_i^{-1}, h_j^{-1}] &= E[\exp\{X_i \hat{\beta}\} \exp\{X_j \hat{\beta}\}] \\
 &\quad - E[\exp\{X_i \hat{\beta}\}] E[\exp\{X_j \hat{\beta}\}] \\
 &= \exp\{(X_i + X_j) \tilde{\beta} + \frac{1}{2}(X_i + X_j) V_{\beta\delta} (X_i + X_j)^T\} \\
 &\quad - \mu_{h_i^{-1}} \mu_{h_j^{-1}}, \tag{B.21}
 \end{aligned}$$

and

$$\begin{aligned}
 \text{Cov}[h_i^{-1}, Y_j] &= E[\exp\{X_i \hat{\beta}\} \exp\{X_j \beta\}] \\
 &\quad - E[\exp\{X_i \hat{\beta}\}] E[\exp\{X_j \beta\}] \\
 &= E[\exp\{X_i \delta\}] \{E[\exp\{(X_i + X_j) \beta\}] \\
 &\quad - E[\exp\{X_i \beta\}] E[\exp\{X_j \beta\}]\} \\
 &\approx \exp\{\frac{1}{2} X_i V_{\delta} X_i^T\} \\
 &\quad \times \left[\begin{array}{c} \exp\{(X_i + X_j) \tilde{\beta} + \frac{1}{2}(X_i + X_j) V_{\beta} (X_i + X_j)^T\} \\ - \mu_{Y_i} \mu_{Y_h} \end{array} \right] \tag{B.22}
 \end{aligned}$$

It is now possible to calculate Equation (B.19) and then $\text{Cov}[h^{-1}(s | s) - \exp\{M(s | s)\beta(s | s - 1)\}]$.