Volume LXXXVIII, Part 1 **No. 168**

PROCEEDINGS
May 6, 7, 8, 9, 2001

A FLEXIBLE FRAMEWORK FOR STOCHASTIC CLAIMS RESERVING

PETER D. ENGLAND AND RICHARD J. VERRALL

Abstract

In this paper, a flexible framework for stochastic claims reserving is considered which includes several models proposed to date as special cases. The methodology is embedded within the generalized additive class of models (Hastie and Tibshirani [7]). The methodology is particularly useful since it allows smoothing of chain ladder development factors and estimation of tail factors automatically and easily as part of the model-fitting process, traditionally performed as an additional stage in the claims reserving process. The framework also provides estimates of reserve variability, which could prove useful in formulating and calibrating dynamic financial analysis (DFA) models.

1. INTRODUCTION

The setting and monitoring of claims reserves is a vital task required of the general insurance actuary. To aid in the setting

of reserves, the actuary can make use of a variety of techniques, the most familiar of which is the chain ladder model or variation thereof (e.g., inflation-adjusted chain ladder, *n*-year average volume-weighted chain ladder, etc.). The principal aim of a reserving exercise is to provide an *estimate* of the amount of money a company should set aside now to meet claims arising in the future on the policies already written. The actuary cannot predict with certainty and knows that there is a distribution of possible outcomes, but uses the techniques at his or her disposal to arrive at the best estimate of the reserve (even if the *best* estimate is not that which is carried in the accounts). Knowledge of the *precision* of that estimate is also desirable. Traditional reserving techniques can help provide a best estimate (a measure of *location* in the distribution of possible outcomes), but cannot help with measures of precision. Of course, the actuary knows that the reserve estimate associated with a well-behaved class of business will be more precise than that of a poorly-behaved class, and that the reserve estimate associated with a short-tailed class is likely to be more precise than that of a long-tailed class, but measuring that precision is difficult.

Stochastic claims reserving models aim to provide measures of location (best estimates) and measures of precision (measures of variability) by treating the reserving process as a data analysis exercise and building a reserving model within a statistical framework. Once within a statistical framework, diagnostic checks of the fitted models are possible, such as goodness-of-fit tests and analysis of residuals (which highlight systematic and isolated departures from the fitted model). Various stochastic reserving models have been proposed over the last two decades, and work progresses as new techniques in the field of statistical modeling become available.

Considerable attention has been given to the relationship between various stochastic models and the chain ladder technique. Stochastic models have been constructed with the aim of producing exactly the same reserve estimates as the traditional de-

terministic chain ladder model. This might seem like a futile exercise, but has the advantages that measures of precision are readily available, and the assumptions underlying the chain ladder model are clarified. More importantly, it provides a bridge between traditional methods and stochastic methods, which is useful for the practitioner who is familiar with traditional methods and needs a starting point for exploring stochastic methods.

Other stochastic reserving models which have been proposed attempt to overcome shortcomings of the chain ladder model by incorporating smoothing, or a parametric form which reduces the number of underlying parameters used to fit the model. The aim of this paper is to present a flexible framework for stochastic claims reserving which allows the practitioner to choose whether to use the basic chain ladder model, or to apply some smoothing, or in the limit to use a parametric curve for the runoff. Several of the models proposed to date fit within this framework, and further extensions are possible which have not yet been tried.

For technical reasons, we consider the modeling of paid losses only. Furthermore, information regarding claim numbers is not taken into account; we consider the modeling of claim amounts only. In this respect we take the basic chain ladder model with paid losses as our starting point. Typically data provided for a simple reserving exercise is in the form of a triangle of paid losses (see Section 6) in which the rows i denote accident years and the columns j delay or development years. Although we consider annual development here only, the methods can be extended easily to semiannual, quarterly or monthly development. The triangle is augmented each period by the addition of a new diagonal. The aim in reserving is to predict likely claim amounts in the missing southeast corner of the claims rectangle, the total reserve (ignoring the tail for the moment) being the sum of these amounts. For monitoring purposes, we might also be interested in the reserve for each accident year.

A review of some existing stochastic reserving models appears in Section 2. This is not exhaustive but provides the neces-

sary background from which the flexible framework in Section 3 can be derived. In Section 4, variability of reserve estimates is considered, and formal goodness-of-fit is considered briefly in Section 5. A worked example is then provided, considering the systematic structure of the model in Section 6 and the error structure in Section 7, before concluding in Section 8.

2. A BRIEF REVIEW OF EXISTING STOCHASTIC CLAIMS RESERVING MODELS

Let C_{ij} denote the (incremental) claims amount arising from accident year i paid in development year j. Early work in this field focused on the logarithm of the incremental claims amounts $Y_{ij} = \ln(C_{ij})$ and the lognormal class of models $Y_{ij} = m_{ij} + \varepsilon_{ij}$ with

$$\varepsilon_{ij} \sim IN(0, \sigma^2) \qquad \text{and} \qquad Y_{ij} \sim IN(m_{ij}, \sigma^2), \tag{2.1}$$

where the expression "$\sim IN(\mu, \sigma^2)$" is interpreted as "distributed as independent normal with mean μ and variance σ^2."

The use of the logarithmic transform immediately imposes a limitation on this class of models in that claim amounts must be positive. The normal responses Y_{ij} are assumed to decompose (additively) into a deterministic nonrandom component with mean $m_{ij} = \eta_{ij}$ and homoscedastic normally distributed random error components about a zero mean. Two model structures are of specific interest:

CASE 1

$$\eta_{ij} = c + \alpha_i + \beta_j; \tag{2.2}$$

CASE 2

$$\eta_{ij} = c + \alpha_i + \beta_i \ln(j) + \gamma_i j \qquad (j > 0). \tag{2.3}$$

A third case, which is a mixture of Cases 1 and 2, uses Equation 2.2 for $j \leq q$ and Equation 2.3 for $j > q$ for some integer q specified by the modeler.

Equations 2.1 and 2.2 define the model introduced by Kremer [8] and used by Renshaw [13], Verrall [21], Zehnwirth [27] and Christofides [2], amongst others. Accident year and development year are treated as factors, with a parameter α for each accident year i and a parameter β for each development year j. This representation is analogous to the chain ladder model, which implies the same development pattern for all accident years, where that pattern is defined by the parameters β_j. Use of this model produces predicted values close, but not identical, to those from the simple chain ladder technique.

Equations 2.1 and 2.3 broadly define the model used by Zehnwirth [28]. A special case is created by setting $\beta_i = \beta$ for all i and $\gamma_i = \gamma$ for all i, where the decay pattern is the same for all accident years and represented by only two parameters. Unlike Case 1, this imposes a strict parametric form on the shape of the runoff. Although this sacrifices goodness-of-fit, it has the advantage that payments can be predicted by extrapolation beyond the range of j observed. This representation is known as the Hoerl curve.

Parameters in the predictor structure η_{ij} are estimated by maximum likelihood, which in the case of normally distributed data is equivalent to minimizing the residual sum of squares. Obtaining this "least squares" solution is straightforward, and is a major reason for the importance of log-linear models in the history of stochastic claims reserving. Although it was possible to use other error distributions (using generalized linear models) at the time these models were propounded, their use was not common and suitable statistical software was in its infancy. De Jong and Zehnwirth [4] adopted the Kalman filter to pass information between accident years and provide smoothed estimates of the parameters β_i and γ_i in Equation 2.3. This idea was adopted by Verrall [21] who used the Kalman filter to smooth over the parameters α_i and β_j in Equation 2.2.

The unknown variance σ^2 is estimated by the residual sum of squares divided by the degrees of freedom (the number of obser-

vations minus the number of parameters estimated). Zehnwirth [29] also considers allowing a different variance estimator for each development period.

Given the parameter estimates, predicted values on a log scale can be obtained by introducing those estimates back into the appropriate equation. Exponentiating then provides an estimate of the median on the untransformed scale, and an estimate of the mean is given by incorporating a variance component to give predicted values on the untransformed scale. Specific details can be found in Verrall [22].

Significant advances were made in stochastic claims reserving with the publication of a paper by Wright [26], which was interesting in two main respects:

- The systematic and random components of the underlying model for the data are based on a risk theoretic model of the claims generating process;
- The error distribution implied by the model is no longer (log) normal.

Wright considered the incremental paid claims C_{ij} to be the sum of N_{ij} (independent) claims of amount X_{ij}. Standard results from risk theory give:

$$\mathrm{E}[C_{ij}] = \mathrm{E}[N_{ij}]\mathrm{E}[X_{ij}], \tag{2.4}$$

and

$$\mathrm{Var}[C_{ij}] = \mathrm{E}[N_{ij}]\mathrm{Var}[X_{ij}] + \{\mathrm{E}[X_{ij}]\}^2\mathrm{Var}[N_{ij}]. \tag{2.5}$$

The formulation is completed by specifying a model for each of $\mathrm{E}[N_{ij}]$ and $\mathrm{E}[X_{ij}]$, a relationship between the mean and variance of the claim numbers N_{ij}, and a relationship between the mean and variance of the claim severities X_{ij}.

Wright considered the claim numbers N_{ij} to be Poisson random variables where

$$\mathrm{E}[N_{ij}] = e_i a_j \kappa_i j^{A_i} e^{-b_i j}, \tag{2.6}$$

and

$$\mathrm{Var}[N_{ij}] = \mathrm{E}[N_{ij}] \tag{2.7}$$

where κ, A and b are unknown constants to be estimated, e_i is a measure of exposure, and a is a *known* adjustment term needed on technical grounds. The values a are specified in Appendix 1 of Wright [26] for each value of j. (Note: Wright also recommended a technical adjustment to development time j, which has been ignored here for simplicity.)

Claim amounts X_{ij} were considered to be Gamma type random variables where

$$\mathrm{E}[X_{ij}] = e^{\delta t} k j^{\lambda}, \tag{2.8}$$

and

$$\mathrm{Var}[X_{ij}] = \upsilon \{\mathrm{E}[X_{ij}]\}^2, \tag{2.9}$$

where k and λ are unknown constants. The optional term $e^{\delta t}$ is included to allow for possible claims inflation, where $t = i + j$ represents calendar time and δ is the estimated constant force of claims inflation. Wright chose not to assume that the claim amounts are actually Gamma distributed, only that the variance exists and is proportional to the mean squared with constant of proportionality υ. This is a subtle technicality which makes no practical difference when claim amounts are all positive.

Equations 2.6 and 2.8 are designed to model the mean claim numbers and mean individual claim severities as functions of delay j.

This formulation is interesting because it uses the same model specification in the claims reserving context as in pricing; that is, claim numbers are modeled as Poisson random variables and claim severities are modeled as Gamma random variables.

Combining Equations 2.4 to 2.9 gives

$$\mathrm{E}[C_{ij}] = m_{ij} = e_i a_j \kappa_i j^{A_i} e^{-b_i j} e^{\delta t} k j^{\lambda}, \tag{2.10}$$

and

$$\mathrm{Var}[C_{ij}] = (1 + \upsilon)kj^{\lambda}e^{\delta t}\mathrm{E}[C_{ij}]. \tag{2.11}$$

Wright showed that with a suitable reparameterization, Equations 2.10 and 2.11 represent a generalized linear model (GLM). Standard statistical methods can be used to estimate the parameters involved.

This model formulation can be viewed as a way of allowing the incremental paid claims C_{ij} to be modeled directly, without the necessity of modeling claim numbers and claim severities separately and then combining. The only information needed to fit the model is the standard triangle of incremental paid claims.

Wright went on to use the Kalman filter to pass information between accident years to produce smoothed parameter estimates, thus avoiding problems associated with the excessive parameterization.

The formulation of the problem as a GLM and the fitting method adopted by Wright are not easy to follow, so the simpler derivation by Renshaw [14] is presented here. Writing:

$$\begin{aligned} u_{ij} &= \ln(e_i a_j), \\ c &= \ln(k), \\ \alpha_i &= \ln(\kappa_i) \qquad \text{with} \quad \kappa_1 = 1, \\ \beta_i &= \lambda + A_i, \qquad \text{and} \\ \gamma_i &= -b_i, \end{aligned}$$

gives

$$\mathrm{E}[C_{ij}] = e^{(u_{ij} + c + \alpha_i + \beta_i \ln(j) + \gamma_i j + \delta t)}.$$

We can then write

$$\eta_{ij} = u_{ij} + c + \alpha_i + \beta_i \ln(j) + \gamma_i j + \delta t, \tag{2.12}$$

giving

$$\mathrm{E}[C_{ij}] = m_{ij}, \tag{2.13}$$

where

$$\ln(m_{ij}) = \eta_{ij}. \tag{2.14}$$

Ignoring the known offset (u_{ij}) and the optional term for claims inflation (δt), Equation 2.12 represents the familiar Hoerl curve which appeared in Equation 2.3.

Using Equations 2.7 and 2.9 in 2.5 gives:

$$\mathrm{Var}[C_{ij}] = \mathrm{E}[N_{ij}]\ \upsilon\{\mathrm{E}[X_{ij}]\}^2 + \{\mathrm{E}[X_{ij}]\}^2\mathrm{E}[N_{ij}];$$

then using Equation 2.4 gives

$$\mathrm{Var}[C_{ij}] = (1 + \upsilon)\mathrm{E}[X_{ij}]\mathrm{E}[C_{ij}].$$

Writing

$$\phi_{ij} = (1 + \upsilon)\mathrm{E}[X_{ij}]$$

gives

$$\mathrm{Var}[C_{ij}] = \phi_{ij}\mathrm{E}[C_{ij}] = \phi_{ij}m_{ij}. \tag{2.15}$$

Equations 2.13, 2.14 and 2.15 define a GLM (see Section 3) in which the response C_{ij} is modeled with a logarithmic link function, the variance is proportional to the mean, and the linear predictor is given by Equation 2.12. The ϕ_{ij} are unknown scale parameters to be estimated by the model.

With GLMs, the unknown scale parameter is usually constant for all observations (i.e., $\phi_{ij} = \phi$ for all i, j) and is estimated by the deviance (or alternatively the Pearson χ^2 statistic) divided by the degrees of freedom. However, in this formulation, it is possible to estimate the scale parameters as part of an extended fitting procedure, known as joint modeling (see Renshaw [14]).

It should be noted that in Renshaw's formulation, the assumption that claim numbers are Poisson distributed was relaxed slightly, the only requirement being that the variance of the number of claims exists and is proportional to the mean.

Therefore

$$\text{Var}[N_{ij}] = \varphi \text{E}[N_{ij}]. \tag{2.16}$$

This is in the spirit of the relaxed assumptions made by Wright [26] about the distribution of claim severities. Claim numbers are said to be distributed as "overdispersed" Poisson random variables. Using Equation 2.16 instead of 2.7 gives:

$$\phi_{ij} = (\varphi + \upsilon)\text{E}[X_{ij}]$$

without changing the specification as a GLM.

Comparing Equation 2.12 with 2.3, it can be seen that Wright is effectively using the same linear predictor as Zehnwirth [28], with the inclusion of an optional term to model possible claims inflation. The u_{ij} terms are known and represent small technical adjustments. They are declared as *offsets* when fitting the model using standard statistical software packages. The important differences between the model used by Zehnwirth and the model proposed by Wright are that:

- Zehnwirth uses the logarithm of the incremental claims as the response, and links the predictor (2.3) to the expected value of the response through the identity link function, therefore requiring the introduction of a variance component when focusing on the mean on the untransformed scale. Wright treats the incremental claims themselves as the responses, and links (essentially) the same predictor to the expected value of the response through the logarithmic link function, thereby avoiding the necessity of the inclusion of a variance component when focusing on the predicted mean.
- In the model proposed by Zehnwirth, the variance is constant for all observations (or constant for each development period), whereas in the model proposed by Wright, the variance is proportional to the mean. A critique of these assumptions can be found in Appendix 4 of Wright [26].
- The log transformation used by Zehnwirth excludes the mass point at zero (although it is possible to make minor adjustments

to zero payments in the data), which Wright's model includes naturally. In fact Wright's model can also be used for data sets that include some negative payments.

It should be noted that in the software package ICRFS, Zehnwirth [29] includes a range of predictor structures, not just the one alluded to above, which could provide an improved fit to the data. However, all are based on log-incremental claims in his "Probabilistic Trend Family."

Equations 2.12 to 2.15 define the model proposed by Wright, and suggest possible alternatives. For example, Renshaw and Verrall [16, 17] replace the linear predictor used by Wright (Equation 2.12) by the linear predictor suggested by Kremer [8], and use a constant scale parameter by setting $\phi_{ij} = \phi$ for all i, j. Therefore,

$$\mathrm{E}[C_{ij}] = m_{ij} \qquad \text{and} \qquad \mathrm{Var}[C_{ij}] = \phi m_{ij}, \tag{2.17}$$

where

$$\ln(m_{ij}) = \eta_{ij} = c + \alpha_i + \beta_j. \tag{2.18}$$

Equations 2.17 and 2.18 define a GLM in which incremental claims are modeled as overdispersed Poisson random variables. This model is particularly interesting since the predicted values given by the model are *exactly* the same as those given by the simple chain ladder model, thus providing a stochastic version of the chain ladder model.

Renshaw and Verrall were not the first to notice the link between the chain ladder model and the Poisson distribution, but were the first to implement the model using standard methodology in statistical modeling and to provide a link with the analysis of contingency tables. Wright [26] also describes a similar model, including a term to model claims inflation, but did not consider the model in detail. Mack [9] also points out that the chain ladder estimates can be obtained by maximizing a Poisson likelihood by appealing to the so-called "method of marginal totals."

Mack [9] suggested using the same linear predictor as Kremer [8] (and therefore the same as Renshaw and Verrall [17]) but proposed using a Gamma distribution for claim amounts. However, Mack developed his own fitting procedure for obtaining maximum likelihood parameter estimates. As Renshaw and Verrall [17] note, the same model can be fitted using the GLM described by Equations 2.17 and 2.18, but replacing $\text{Var}[C_{ij}] = \phi m_{ij}$ by $\text{Var}[C_{ij}] = \phi m_{ij}^2$. Standard statistical software packages can then be used to obtain maximum likelihood parameter estimates.

In Verrall [24], the stochastic chain ladder model of Renshaw and Verrall [16] was extended to incorporate smoothing of parameter estimates over accident years (the α_is in Equation 2.18), while leaving the model describing the runoff pattern (the β_js) alone. Nonparametric smoothers were used and fitted using generalized *additive* models (GAMs). GAMs differ from GLMs in the way in which the relationship between the response variable and the covariates is modeled. In GLMs the relationship is parametric; in GAMs the response is assumed to vary smoothly with the covariates through the introduction of a smoothing procedure. In this paper, the idea is extended to allow smoothing over development years, which is of considerable practical benefit and provides a flexible framework for stochastic claims reserving.

3. A FLEXIBLE FRAMEWORK FOR STOCHASTIC CLAIMS RESERVING

A GLM is defined by focusing on a set of independent response variables $\{Y_u : u = 1, 2, \ldots, n\}$. The objective is to model the expected value of the response as a function of one or more covariates. We assume that the Y_u are distributed according to a member of the one-parameter exponential family of distributions, which includes the normal, Poisson and Gamma distributions, amongst others. Denoting the expected value of Y_u by m_u, the first two moments take the general form

$$\text{E}[Y_u] = m_u \qquad \text{and} \qquad \text{Var}[Y_u] = \frac{\phi V(m_u)}{w_u},$$

TABLE 3.1

SCALE PARAMETERS AND VARIANCE FUNCTIONS FOR SOME STANDARD DISTRIBUTIONS

Distribution	Scale Parameter ϕ	Variance function $V(m_u)$
Normal	σ^2	1
Poisson	1	m_u
Gamma	>0	m_u^2
Inverse Gaussian	>0	m_u^3

where ϕ denotes a scale parameter, w_u are prior weights (often set to 1 for all observations), and $V()$ is the so-called *variance function* (a function of the mean). The choice of distribution dictates the values of ϕ and $V()$. The values of the scale parameter and variance function for various standard distributions are shown in Table 3.1. The definition of a GLM is completed by specifying the deterministic structure, which is achieved through a linear predictor η_u where

$$\eta_u = \sum_{v=1}^{p} x_{uv}\beta_v \tag{3.1}$$

with known covariates x_v associated with each observation u, and unknown parameters β_v. The expected value of the response is linked to the linear predictor through a link function $g()$ such that

$$g(m_u) = \eta_u.$$

It is helpful to think of GAMs as extensions of GLMs. A GAM is defined by replacing Equation 3.1 by

$$\eta_u = \sum_{v=1}^{p} s_v(x_u),$$

where $s(x)$ represents a nonparametric smoother on x. It is possible to choose from several different types of smoothers, such as locally weighted regression smoothers (loess), cubic smooth-

ing splines and kernel smoothers. Other features of GAMs, such as the choice of error distribution, link function, goodness-of-fit measures and residual definitions are common to GLMs with the main difference between GAMs and GLMs being the specification of the predictor η.

A complete exposition of the statistical background of generalized linear models and generalized additive models can be found in McCullagh and Nelder [12] and Hastie and Tibshirani [7] respectively.

It should be noted in passing that we are not restricted to using a smoother for all covariates; the predictor may comprise a mixture of parametric and nonparametric components. The predictor then becomes

$$\eta_u \sum_{v=1}^{p-r} x_{uv}\beta_v + \sum_{v=p-r+1}^{p} s_v(x_u).$$

In claims reserving, the cubic smoothing spline has been found to be particularly useful. When data are normally distributed, the (univariate) cubic smoothing spline $s(x)$ is found by minimizing the penalized residual sum of squares

$$\sum_{u=1}^{n} (y_u - s(x_u))^2 + \theta \int (s''(t))^2 dt. \tag{3.2}$$

The second part of Equation 3.2 defines a smoothness penalty based on curvature of the spline function $s(x)$. The level of smoothing is controlled by the single parameter $\theta(> 0)$. When θ tends to zero, there is no smoothness penalty and the model provides a perfect fit: the fitted values are the data points themselves. When θ is large (tends to infinity), the fit is perfectly smooth and the fitted values fall along a straight line, effectively forcing the relationship to be linear in x. The parameter θ is set between these extremes to produce the desired level of smoothness, and controls the trade-off between goodness-of-fit and smoothness. Although the cubic smoothing spline has received considerable attention recently in statistical modeling, it is usually attributed with

appearing first in the actuarial literature in a paper on graduating mortality rates by Whittaker [25]. In fact Whittaker graduation is used widely for graduating mortality rates in the US.

Within the context of non-normal error distributions from the exponential family, a weighted version of Equation 3.2 is fitted by inserting an extra iterative algorithm within the optimization procedure. Details of this can be found in Hastie and Tibshirani [7], and Green and Silverman [6].

To construct a flexible framework for stochastic claims reserving, within which several of the models described in Section 2 can be regarded as special cases, we focus on the incremental paid claims C_{ij} and define

$$\mathrm{E}[C_{ij}] = m_{ij}, \tag{3.3}$$

$$\mathrm{Var}[C_{ij}] = \phi m_{ij}^{\rho}, \tag{3.4}$$

and

$$\ln(m_{ij}) = \eta_{ij} = u_{ij} + \delta t + c + s_{\theta_i}(i) + s_{\theta_j}(j) + s_{\theta_j}(\ln(j)). \tag{3.5}$$

Equations 3.3, 3.4 and 3.5 specify a generalized additive model with power variance function and constant scale parameter. The power ρ dictates the choice of error distribution, with normal, Poisson, Gamma and Inverse Gaussian specified by $\rho = 0$, 1, 2, and 3, respectively. The predictor is linked to the expected value of the response through the logarithmic link function. The offsets u_{ij} and inflation term δt are optional (where $t = i + j$), and may be suggested by a particular context. The function $s(i)$ represents a smooth of accident year i, obtained using a smoothing spline with smoothing parameter θ_i. Similarly, the functions $s(j)$ and $s(\ln(j))$ represent smoothing splines specifying the shape of the runoff pattern, with smoothing parameter θ_j chosen (for simplicity) to be the same for both functions. In practice, it may not be necessary to include smooths in both j and $\ln(j)$. It should be noted that both accident year i and development year j are considered as continuous covariates. It can

TABLE 3.2

GENERALIZED ADDITIVE MODEL REPRESENTATION OF SOME PUBLISHED STOCHASTIC RESERVING MODELS

	Variance power ρ	Row smoothing parameter θ_i	Column smoothing parameter θ_j
Wright (1990)*	1	0	∞
Mack (1991)	2	0	0
Renshaw and Verrall (1994, 1998)	1	0	0
Renshaw (1994)	1,2	0	0
Verrall (1996)	1	>0	0

*We consider here only the special case in which the same runoff pattern is used for all accident years, the Kalman filter is not used, and the scale parameter is constant.

be seen that use of Equation 3.5 implicitly assumes the same runoff pattern for all accident years, although the model can be extended using carefully chosen interaction terms. It is trivial to extend Equation 3.5 further, for example, to allow for a step change in a particular calendar year introduced by a change in legislation.

The extremes of the smoothing parameters are interesting and provide the link between Equation 3.5 and Equations 2.12 and 2.18 (ignoring the optional terms u_{ij} and δt). When θ_i is zero, there is no smoothing and the model is forced to pass through each value of i, which treats accident year i as though it is a factor (as in 2.12 and 2.18). The same is true of θ_j; when θ_j is zero, the model is forced to pass through each value of j, and development time is treated as though it is a factor (as in 2.18). When θ_j tends to infinity, the part of the model relating to development time is linear in j and $\ln(j)$, giving the Hoerl curve (as in 2.12 and 2.3). It is also necessary to choose the power function ρ to complete the model specification.

Table 3.2 shows how several previous stochastic reserving models can be seen as special cases of the model specified by

Equations 3.3, 3.4 and 3.5. The optional terms u_{ij} and δt are ignored without loss of generality.

The early log-linear models do not fit so neatly into the same framework because those models used log-incremental claims as the response, and required incorporation of a variance component in the mean of the predicted values. However, the framework could easily be extended to allow for this.

Notice that we consider only models in which the scale parameter in Equation 3.4 is assumed constant. This is for ease of exposition, although the model can be generalized further by relaxing this assumption and estimating the unknown scale parameters by joint modeling.

Having chosen the model specification, the model can be fitted using maximum quasi likelihood to obtain parameter estimates (and their approximate standard errors). At this point we make use of standard statistical software packages which have the facility to fit generalized additive models. Currently the choice is limited, although greater choice is likely in the future as the popularity of generalized additive models increases. The authors used S-PLUS [19] for the example (see also Chambers and Hastie [1]).

Having fitted the model, we obtained reserve estimates by summing the appropriate predicted values in the southeast region of the claims rectangle. All that remains is the estimation of variability in the reserve estimates, considered in the next section.

4. PRECISION OF RESERVE ESTIMATES

One of the principal advantages of stochastic reserving models is the availability of estimates of precision. Commonly used in prediction problems (as we have here) is the standard error of prediction, also known as the prediction error, or root mean square error of prediction. For claim payments in development year j for accident year i (yet to be observed), the mean square

error of prediction is given by

$$\mathrm{E}[(C_{ij} - \hat{C}_{ij})^2] \approx \mathrm{Var}[C_{ij}] + \mathrm{Var}[\hat{C}_{ij}]. \tag{4.1}$$

Note that the mean square error of prediction can be considered as the sum of two components: variability in the data (process variance) and variability due to estimation (estimation variance). The precise form of the two components of variance is dictated by the specification of the model fitted. For a detailed justification of Equation 4.1, see Renshaw [15].

For the general model defined above, the process variance is given by Equation 3.4. For the estimation variance, we note that

$$\hat{C}_{ij} = \hat{m}_{ij} = e^{\hat{\eta}_{ij}}.$$

Then, using a Taylor series expansion,

$$\mathrm{Var}[\hat{C}_{ij}] \approx \left|\frac{\partial m_{ij}}{\partial \eta_{ij}}\right|^2 \mathrm{Var}[\hat{\eta}_{ij}],$$

giving

$$\mathrm{E}[(C_{ij} - \hat{C}_{ij})^2] \approx \phi \hat{m}_{ij}^{\rho} + \hat{m}_{ij}^2 \mathrm{Var}[\hat{\eta}_{ij}]. \tag{4.2}$$

The final component of Equation 4.2, the variance of the (linear) predictor, is usually available directly from statistical software packages, enabling the mean square error to be calculated without difficulty. The standard error of prediction is the square root of the mean square error of prediction.

The standard error of prediction for origin year reserve estimates and the total reserve estimates can also be calculated. Denoting the missing southeast region of the claims rectangle by Δ, then the reserve estimate in origin year i is given by summing the predicted values in row i of Δ; that is,

$$\hat{C}_{i+} = \sum_{j \in \Delta_i} \hat{C}_{ij}.$$

The mean square error of prediction of the origin year reserve is given by

$$\mathrm{E}[(C_{i+} - \hat{C}_{i+})^2] = \sum_{j \in \Delta_i} \phi \hat{m}_{ij}^{\rho} + \sum_{j \in \Delta_i} \hat{m}_{ij}^2 \mathrm{Var}[\hat{\eta}_{ij}] + 2 \sum_{\substack{j_1, j_2 \in \Delta_i \\ j_2 > j_1}} \hat{m_{ij_1}} \hat{m}_{ij_2} \mathrm{Cov}. \tag{4.3}$$

The total reserve estimate is given by

$$\hat{C}_{++} = \sum_{i,j \in \Delta} \hat{C}_{ij},$$

and the mean square error of prediction of the total reserve is given by

$$\mathrm{E}[(C_{++} - \hat{C}_{++})^2] = \sum_{i,j \in \Delta} \phi \hat{m}_{ij}^{\rho} + \sum_{i,j \in \Delta} \hat{m}_{ij}^2 \mathrm{Var}[\hat{\eta}_{ij}] + 2 \sum_{\substack{i_1 j_1 \in \Delta \\ i_2 j_2 \in \Delta \\ i_1 j_1 \neq i_2 j_2}} \hat{m}_{i_1 j_1} \hat{m}_{i_2 j_2} \mathrm{Cov}[\hat{\eta}_{i_1 j_1}, \hat{\eta}_{i_2 j_2}]. \tag{4.4}$$

Although Equations 4.3 and 4.4 look fairly complex, they are relatively easy to calculate by summing the appropriate elements. The only components not readily available from statistical software packages are the covariance terms. Provided the *design matrix* and *variance-covariance matrix* of the parameter estimates can be extracted from the statistical software package used, a full matrix of the covariance terms can be calculated without difficulty for any specification of the predictor η. Indeed, the variances of the (linear) predictors are simply the diagonal of such a matrix.

It is also possible to obtain estimates of payments to be made in future settlement years by summing over diagonals in Δ, and also to obtain the associated standard error of prediction. Further

details of this and a detailed derivation of Equations 4.3 and 4.4 can be found in Renshaw [15].

5. ASSESSING THE GOODNESS-OF-FIT

For a given error distribution (chosen by the power ρ), specific models are chosen by the smoothing parameters θ_i and θ_j, and different models are fitted by varying the smoothing parameters until a satisfactory fit is achieved. Assessing whether a model is satisfactory in practice is part art and part science. Usually, informal checks will suffice in practice, although model comparison can proceed formally in the usual way by comparing the difference in *deviances* of the fitted models (for fixed ρ) to the appropriate percentage point on the χ^2 or F distributions. However, because the smoothers are nonparametric, it is not obvious how many degrees of freedom should be used in the model comparison. According to the theory of cubic smoothing splines, it is possible to assess the equivalent degrees of freedom used in fitting the spline. This has an inverse relationship to the smoothing parameter: as the smoothing parameter increases, the equivalent degrees of freedom decrease. After fitting a cubic smoothing spline, statistical software packages provide the equivalent degrees of freedom as part of the model output. One problem is that the smoothing parameter is a continuous measure, which can result in noninteger degrees of freedom. For this reason, software packages tend to allow the amount of smoothness to be defined alternatively by the equivalent degrees of freedom, which is provided by the user. The smoothness parameter to be used is then calculated from the given degrees of freedom.

The choice of error distribution is not easy to justify but may be suggested on theoretical grounds. Formally, given identical specifications of the predictor, the optimum value of ρ (which specifies the choice of error distribution) is that which produces the highest likelihood.

Residual plots are also used to assess the adequacy of any fitted model. Two types of residual used commonly are the Pearson and deviance residuals. The scaled Pearson residuals are defined by

$$r_{ij} = \frac{C_{ij} - \hat{m}_{ij}}{\sqrt{\phi \hat{m}_{ij}^{\rho}}},$$

and the scaled deviance residuals are defined by

$$\sim r_{ij} = \text{sign}(C_{ij} - \hat{m}_{ij})\sqrt{\frac{d_{ij}}{\phi}},$$

where d_{ij} is the contribution to the deviance made by observation C_{ij}.

For a reasonable model, a histogram of scaled residuals is expected to be approximately normal (i.e., bell shaped) with 95% of the residuals between the values plus two and minus two. Residuals can also be plotted against the predictor, against origin year and against development year. The plots are expected to be pattern free, where an obvious pattern in the residuals would indicate a systematic departure from the fitted model. Isolated departures from the model would be indicated by residuals whose values are far from zero. Other residual plots are also possible. It is usual to assess residual plots visually, any serious model deficiencies being immediately obvious.

A further visual check which is useful when comparing models is to plot that part of the predictor that explains the runoff pattern against development time. From Equation 3.5, this translates into plotting $c + s_{\theta_j}(j) + s_{\theta_j}(\ln(j))$ against j for various values of θ_j. The constant c is needed to ensure the plots start at equivalent levels. A plot such as this might result in the choice of a model which is not optimal in the statistical sense, but which may have convenient properties (for example, the way it behaves when extrapolating into the tail).

TABLE 6.1

INCREMENTAL PAID LOSSES FORMED BY AGGREGATING ACROSS DIFFERENT CLASSES

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$
$i=1$	45630	23350	2924	1798	2007	1204	1298	563	777	621
$i=2$	53025	26466	2829	1748	732	1424	399	537	340	
$i=3$	67318	42333	−1854	3178	3045	3281	2909	2613		
$i=4$	93489	37473	7431	6648	4207	5762	1890			
$i=5$	80517	33061	6863	4328	4003	2350				
$i=6$	68690	33931	5645	6178	3479					
$i=7$	63091	32198	8938	6879						
$i=8$	64430	32491	8414							
$i=9$	68548	35366								
$i=10$	76013									

6. EXAMPLE: PART 1—A COMPARISON OF PREDICTOR STRUCTURES

Incremental paid losses from an aggregation of classes of business are shown in Table 6.1 and are used to illustrate the methodology. The incremental claims fall fairly rapidly, but are not completely runoff by the end of the tenth development year, implying the necessity for a tail factor greater than 1 when using the traditional chain ladder model. Notice the negative incremental claim at position (3,3), which is not a problem when implementing the models.

Initially, to illustrate the methodology, we fit three models, using an overdispersed Poisson model ($\rho = 1$ in Equation 3.4) with a logarithmic link function. For all three models

$$\mathrm{E}[C_{ij}] = m_{ij}, \qquad \mathrm{Var}[C_{ij}] = \phi m_{ij}, \qquad \text{and} \qquad \ln(m_{ij}) = \eta_{ij}.$$

The models differ only in the choice of the predictor. The predictor structures are:

- Model 1: The stochastic model of Renshaw and Verrall [17], which gives the same reserve estimates as the chain ladder

model:

$$\eta_{ij} = c + \alpha_i + \beta_j.$$

This model can be specified as a generalized additive model with $\theta_i = 0$ and $\theta_j = 0$ (no smoothing), giving

$$\eta_{ij} = c + s_0(i) + s_0(j) + s_0(\ln(j)).$$

- Model 2: The Hoerl curve, ignoring inflation:

$$\eta_{ij} = u_j + c + \alpha_i + \beta \ln(j) + \gamma j.$$

This is in the spirit of the model proposed by Wright [26]. Here we adopt the technical adjustments to development time recommended by Wright, and the associated offset (ignoring exposure information). However, we are using the same runoff pattern for each accident year (since β and γ do not depend on i), we ignore the Kalman filter, and we are using a constant scale parameter.

Again, this model can be specified as a generalized additive model with $\theta_i = 0$ and $\theta_j = \infty$, giving

$$\eta_{ij} = u_j + c + s_0(i) + s_\infty(j) + s_\infty(\ln(j)).$$

- Model 3: A generalized additive model with a parameter for each accident year, but with the pattern over development year represented by a smooth in log development time. We have chosen not to include additionally a smooth in development time, which in this case is unnecessary. Therefore $\theta_i = 0$ and θ_j is chosen to provide a suitable level of smoothing, giving

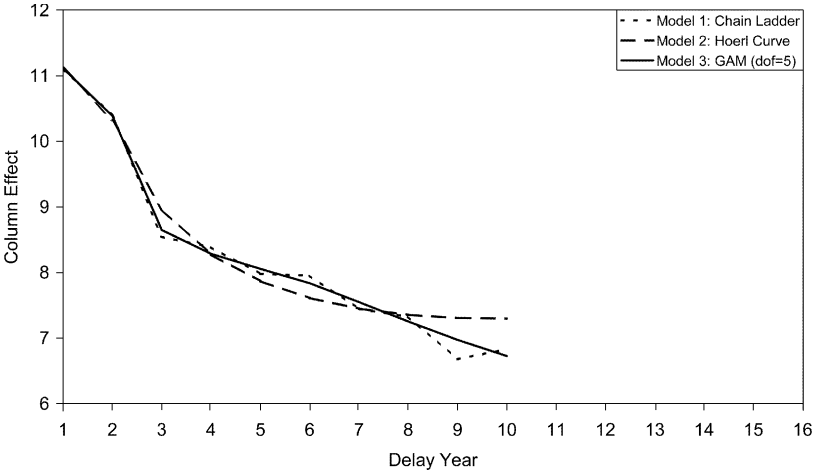
$$\eta_{ij} = u_j + c + s_0(i) + s_{\theta_j}(\ln(j)),$$

or equivalently

$$\eta_{ij} = u_j + c + \alpha_i + s_{\theta_j}(\ln(j)).$$

FIGURE 6.1

COLUMN EFFECTS VS. DELAY YEAR

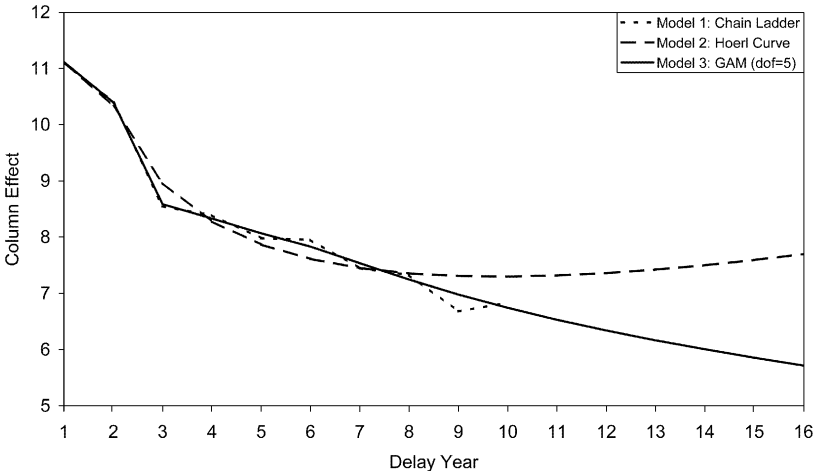


This can be seen as a smooth model in between the chain ladder and Hoerl curve models. For this example, the smoothing parameter was dictated by setting the equivalent degrees of freedom (dof) used in the fit (in this case dof = 5).

First, consider the part of each predictor that describes the shape of the decay of the incremental claims (the sum of the components not dependent on i). We shall call this the "column effects." Figure 6.1 shows the column effects for all three models, and there we can see the jagged shape of the decay in the incremental claims assumed by the chain ladder model and the smooth shape of the model using the Hoerl curve. The Hoerl curve passes through the chain ladder model, fitting closely in the early stages of development (where we have the most data) but fails to fall rapidly enough in the later stages of development. This is the result of the strict parametric form imposed by the Hoerl curve. (A practitioner would probably reject the model at this point, but we will continue to highlight the characteristics

FIGURE 6.2

COLUMN EFFECTS (EXTRAPOLATED) VS. DELAY YEAR



of the Hoerl curve and to enable a comparison with the generalized additive model methodology.) Model 3 is in between the extremes of Models 1 and 2, and exhibits a satisfactory mix of smoothness and adherence to the data. If the smoothing parameter of Model 3 is reduced, it will tend towards Model 1. Conversely, if the smoothing parameter of Model 3 is increased, it will tend towards Model 2.

In Figure 6.2, Models 2 and 3 have been extrapolated a further six years. With this example, an inherent danger of extrapolating using rigid parametric curves like the Hoerl curve is highlighted since the curve bends upwards beyond the range of data observed. One advantage of Model 3 is that it continues in a more desirable direction when extrapolating.

Although natural in stochastic claims reserving, it is unusual to focus on the shape of the decay of incremental claims using traditional actuarial methods, in which it is common to focus on

the relative increase in cumulative claims through *development factors*, the traditional "parameters" in a standard chain ladder exercise. After fitting a stochastic claims reserving model, it is straightforward to obtain *equivalent development factors* by applying the standard chain ladder model to the fitted values of the stochastic model. If the model is fully parametric, it may be possible to obtain a relationship between the model parameters and the chain ladder development factors (e.g., Verrall [23]).

Equivalent development factors are shown in Table 6.2 for Models 1 to 3, together with the actual development factors obtained by applying the standard chain ladder model to the data in Table 6.1. It can be seen that the development factors implied by the stochastic chain ladder model (Model 1) are identical to those obtained using standard chain ladder methodology (therefore reserve estimates obtained using the two models will also be identical). A comparison of the development factors implied by the Hoerl curve (Model 2) and the chain ladder models reveals where these two models differ. In particular, the Hoerl curve does not fully capture the fall in the development factors in the later stages of development. The development factors implied by Model 3 can be seen as a smoothed version of the chain ladder development factors.

Also shown in Table 6.2 are the equivalent development factors obtained when extrapolating beyond development year 10. It can be seen clearly that the development factors implied by the Hoerl curve increase in value, whereas the development factors implied by Model 3 continue to decrease.

The reserve estimates implied by Models 1, 2 and 3 are shown in Table 6.3, together with their prediction errors (as a percentage of the reserves). For ease of comparison with the chain ladder model, we have not extrapolated into the tail. The reserve estimates given by the Hoerl curve are higher for the older years than those given by the chain ladder model, reflecting the higher development factors at the later stages of development. The

TABLE 6.2

EQUIVALENT DEVELOPMENT FACTORS: OVERDISPERSED POISSON MODEL

Delay Year	Standard Chain Ladder	Model 1 Stochastic Chain Ladder	Model 2 Hoerl Curve	Model 3 GAM (dof = 5)
2	1.4906	1.4906	1.4496	1.4891
3	1.0516	1.0516	1.0796	1.0537
4	1.0419	1.0419	1.0372	1.0395
5	1.0268	1.0268	1.0238	1.0292
6	1.0254	1.0254	1.0180	1.0224
7	1.0149	1.0149	1.0150	1.0163
8	1.0130	1.0130	1.0135	1.0120
9	1.0067	1.0067	1.0127	1.0091
10	1.0078	1.0078	1.0124	1.0071
11			1.0125	1.0057
12			1.0129	1.0047
13			1.0135	1.0039
14			1.0144	1.0033
15			1.0156	1.0029
16			1.0171	1.0025

reserve estimates given by Model 3 are close to those provided by the chain ladder model for all years individually and in total, with any differences arising due to the amount of smoothing.

The reduced number of parameters in the Hoerl curve compared to the stochastic chain ladder model should drive down the prediction error, but this is offset by the increased variability imposed by the poor fit, resulting in prediction errors for the Hoerl curve which are close to those provided by the stochastic chain ladder model. The equivalent degrees of freedom used up in fitting Model 3 is lower than the degrees of freedom used up in fitting the stochastic chain ladder model, which will drive down the prediction errors. Furthermore, the fit is good relative to the chain ladder model, which has the desirable effect of lower prediction errors for Model 3 compared to the stochastic chain ladder model.

TABLE 6.3

RESERVE ESTIMATES AND PREDICTION ERRORS: OVERDISPERSED POISSON MODEL

	Reserve Estimates			Prediction Error		
Accident Year	Model 1 Stochastic Chain Ladder	Model 2 Hoerl Curve	Model 3 GAM (dof = 5)	Model 1 Stochastic Chain Ladder	Model 2 Hoerl Curve	Model 3 GAM (dof = 5)
1	0	0	0	—	—	—
2	683	1,085	622	159%	95%	110%
3	1,792	3,101	1,998	100%	61%	62%
4	4,363	6,129	4,470	63%	46%	43%
5	5,657	7,173	5,940	50%	43%	38%
6	8,209	8,689	8,106	40%	39%	33%
7	10,914	11,031	11,106	34%	34%	29%
8	15,199	14,765	15,112	28%	30%	25%
9	21,135	24,002	21,293	24%	23%	22%
10	60,335	59,625	60,377	17%	17%	16%
Total	128,286	135,600	129,024	15%	15%	12%

Models 1 and 2 can be fitted in any statistical software package that fits generalized linear models. Model 3 can only be fitted in statistical software packages that fit generalized additive models.

The comparison of Model 3 with Models 1 and 2 begins to show how our modeling framework can be considered generic, since the chain ladder model and Hoerl curve model can be fitted as special cases, using extremes of the smoothing parameters. A model that has the desirable characteristic of being able to smooth development factors can be fitted by choosing smoothing parameters between these extremes.

7. EXAMPLE: PART 2—A COMPARISON OF ERROR STRUCTURES

Continuing the example, the same three model predictors are used, but with a Gamma error structure ($\rho = 2$) giving:

$$\mathrm{E}[C_{ij}] = m_{ij}, \qquad \mathrm{Var}[C_{ij}] = \phi m_{ij}^2, \qquad \text{and} \qquad \ln(m_{ij}) = \eta_{ij},$$

TABLE 7.1

EQUIVALENT DEVELOPMENT FACTORS: GAMMA MODEL

Delay Year	Standard Chain Ladder	Model 4 Stochastic Chain Ladder	Model 5 Hoerl Curve	Model 6 GAM (dof = 5)
2	1.4906	1.4969	1.4515	1.4771
3	1.0516	1.0470	1.0799	1.0512
4	1.0419	1.0381	1.0372	1.0357
5	1.0268	1.0259	1.0237	1.0280
6	1.0254	1.0251	1.0178	1.0221
7	1.0149	1.0154	1.0148	1.0165
8	1.0130	1.0131	1.0131	1.0125
9	1.0067	1.0084	1.0123	1.0098
10	1.0078	1.0086	1.0119	1.0079
11			1.0119	1.0066
12			1.0122	1.0055
13			1.0127	1.0048
14			1.0135	1.0041
15			1.0145	1.0036
16			1.0157	1.0032

and the following three models:

- Model 4:
$$\eta_{ij} = c + \alpha_i + \beta_j;$$
- Model 5:
$$\eta_{ij} = u_j + c + \alpha_i + \beta \ln(j) + \gamma j;$$
- Model 6:
$$\eta_{ij} = u_j + c + \alpha_i + s_{\theta_j}(\ln(j)).$$

Equivalent development factors are shown in Table 7.1, and reserve estimates and prediction errors are shown in Table 7.2 (ignoring tail factors).

Comparison of the equivalent development factors from the Gamma model with those from the overdispersed Poisson model is uninformative on the whole. It is perhaps surprising at first

TABLE 7.2

RESERVE ESTIMATES AND PREDICTION ERRORS: GAMMA MODEL

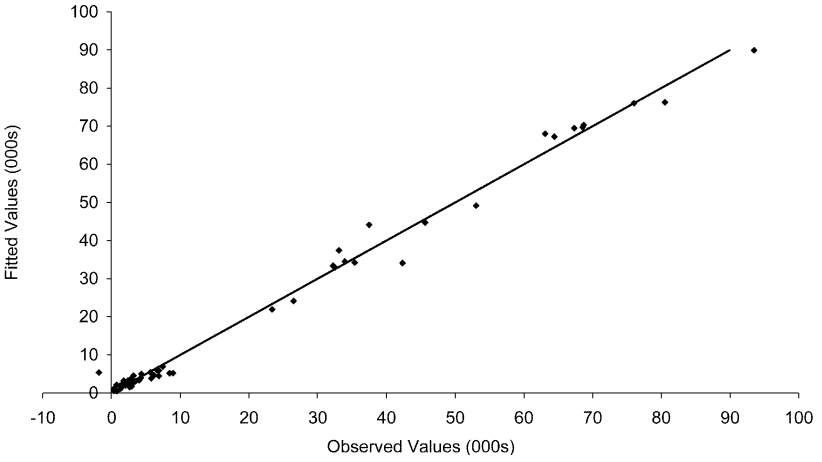
	Reserve Estimates			Prediction Error		
Accident Year	Model 4 Stochastic Chain Ladder	Model 5 Hoerl Curve	Model 6 GAM (dof = 5)	Model 4 Stochastic Chain Ladder	Model 5 Hoerl Curve	Model 6 GAM (dof = 5)
1	0	0	0	—	—	—
2	488	675	450	62%	46%	43%
3	2,086	3,296	2,205	43%	36%	33%
4	5,240	6,818	5,300	36%	32%	29%
5	6,169	7,061	6,313	32%	30%	28%
6	9,750	9,305	9,427	31%	29%	28%
7	15,080	13,029	15,097	31%	29%	29%
8	18,498	15,069	17,671	32%	30%	31%
9	20,470	24,400	20,896	36%	35%	35%
10	60,043	59,576	58,519	52%	48%	48%
Total	137,824	139,229	135,878	25%	23%	24%

sight that the final development factor for the Gamma "chain ladder" model (1.0086) is greater than the equivalent factor from the overdispersed Poisson model (1.0078), but at the same time the reserve estimate is lower (488 vs. 683). This is because the cumulative fitted values for the final observed diagonal of the two models are not the same, resulting in the observed effect. In fact, the cumulative fitted values for the final observed diagonal are identical to the cumulative paid to date for the overdispersed Poisson chain ladder model only.

The main difference between the overdispersed Poisson and Gamma models in this example is in the prediction errors as a percentage of the total reserve estimates, which for the Gamma model are around twice those of the Poisson model. Inspection of the prediction errors of the row reserves gives a hint as to why this is so. For the Gamma model, the prediction errors for the earlier years are lower than those for the Poisson model.

FIGURE 7.1

FITTED VALUES (POISSON MODEL) VS. OBSERVED VALUES

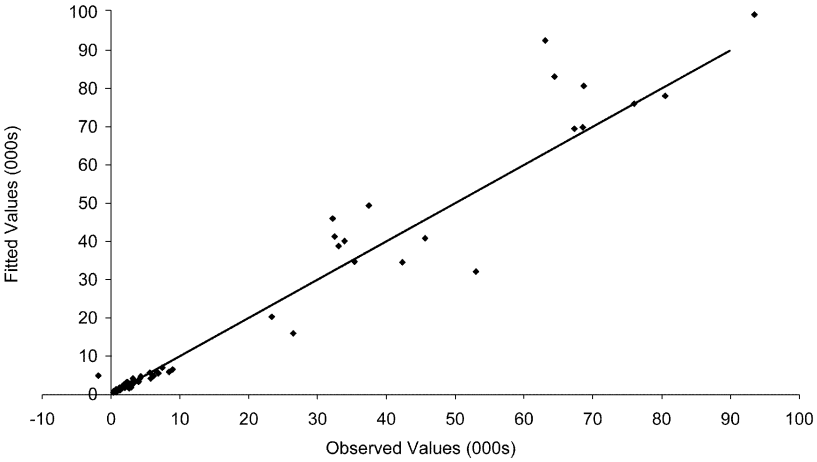


However, the pattern is reversed in the later years, particularly for year 10. The later years contribute by far the largest proportion of the total reserves, which is reflected in the high prediction error of the total.

The fit of the Gamma model is in fact poor in this example, particularly in the early stages of development, where the large incremental observed values are given less weight in the model fitting than in the Poisson model. This is not apparent from an inspection of residual plots (not shown), which look satisfactory for both error structures, but becomes apparent when plotting fitted values against observed values (Figures 7.1 and 7.2), which show clearly the superiority of the Poisson model in this example. This is not always the case, however, and care must be taken in making inferences from these results. For a further example in which prediction errors of claims reserves are compared using different error structures and different methodologies, see England and Verrall [5].

FIGURE 7.2

FITTED VALUES (GAMMA MODEL) VS. OBSERVED VALUES



8. DISCUSSION AND CONCLUSIONS

Given a triangle of data, a simple reserving exercise might proceed by fitting a chain ladder model (usually a 3, 4, or 5 year volume-weighted average chain ladder) and looking at the resultant development factors. It would then be common to smooth the factors manually and consider the necessity of a tail factor for projecting beyond the range of data observed. Judgment is used to smooth the factors with the aim of smoothing out random variations, particularly in the later stages of development, while leaving the systematic trend intact. A tail factor might be chosen by calculating the ratio of cumulative incurred claims to cumulative paid claims for the oldest accident year, or by fitting a curve to the later development factors and extrapolating (see, for example, Craighead [3] and Sherman [18]). Advantages of this procedure are that it is extremely flexible, and it forces the actuary to look at the data. Disadvantages are that it is time consuming, statistically inefficient, and it is not always easy to

be consistent over the level of smoothing (or confident in the results).

The main strength of the method presented in this paper is that both the smoothing and extrapolating can be performed at the same time in the same model. The actuary simply has to choose one parameter for smoothing across the whole range of development time, choose an error distribution, and choose how far to extrapolate (an additional parameter is necessary if smoothing over accident years). Further advantages are that it is also possible to obtain measures of precision of the reserve estimates, and investigate where the data deviate from the fitted model by viewing residual plots. The fact that standard models can be fitted by choosing smoothing parameters at the extremes is a useful additional feature, if only for clarity of understanding, since at one extreme the model can be considered overparameterized, and at the other that the structure is too rigid. However, we do not consider the method to be a panacea. A thorough reserving exercise will involve an in depth investigation of the data, an understanding of the class of business under review, and a comparison of the results of several reserving methods relying on complementary sets of data. We believe the method proposed here is simply an extremely useful additional tool for the reserving specialist.

Incremental data are used for the method put forward in this paper: this is both an advantage and a disadvantage. It is advantageous since the method can be used when the data history is incomplete. If incremental data were recorded by accident year only after a certain date, accident years prior to that date will have incomplete runoff information, and a section of the claims triangle in the northwest corner will be missing (this is a reasonably common occurrence). This presents difficulties using standard deterministic techniques that rely on cumulative data, but is not a problem for stochastic techniques which treat the unobserved data as "missing" and estimate the data as part of the fitting procedure. The disadvantage is that negative incremental values

sometimes occur in data based on paid losses, and frequently occur in data based on incurred losses where case estimates are often set on a conservative basis and overestimated. The method proposed is robust to a small number of negative incremental claims (as in the example), but will always produce positive fitted values (due to the use of the logarithmic link function) and hence will always produce development factors greater than one. For this reason, the techniques are often not suitable for use with incurred data which often include a series of negative incremental losses in the later stages of development requiring development factors less than one.

In the framework proposed in this paper, a constant scale parameter has been used. This is for ease of exposition; the assumption can be relaxed to allow the scale parameter to be modeled as part of an extended procedure. The difference between the prediction errors of the overdispersed Poisson and Gamma models in Section 7 is partly due to the use of a constant scale parameter, and further research is needed to evaluate how much of the difference can be ameliorated by joint modeling.

The main use of stochastic reserving methods is in the provision of estimates of reserve variability, not in the reserve estimates themselves. Until recently, measures of variability have been of little interest to most general insurance actuaries, but interest is likely to increase as the need to parameterize and calibrate dynamic financial analysis (DFA) models becomes routine. Part of a DFA exercise is quantifying reserving risk, and to do this, it is necessary to have a model that simulates the likely payments of outstanding liabilities. Stochastic reserving techniques provide a model structure and a way of calibrating the model to real data, from which payments can be simulated (taking care to allow for process and estimation error).

As outlined in Section 2, there is a wide variety of methods available for stochastic claims reserving. If the use of these methods increases, it is important that the similarities and differences

of the models are understood, and their properties examined. By presenting some of the models within the same framework, and extending to allow flexibility between the extremes of two well-known models, it is hoped that this paper has contributed to the process.

REFERENCES

[1] Chambers, John M. and Trevor J. Hastie, *Statistical Models in S*, Chapman and Hall, London, 1992.

[2] Christofides, Stavros, "Regression Models Based on Log-Incremental Payments," *Claims Reserving Manual* 2, 1990, Institute of Actuaries, London.

[3] Craighead, David H., "Some Aspects of the London Reinsurance Market in World-Wide Short-Term Business," *Journal of the Institute of Actuaries* 106, 1979, pp. 227–287.

[4] De Jong, Piet and Ben Zehnwirth, "Claims Reserving, State-Space Models and the Kalman Filter," *Journal of the Institute of Actuaries* 110, 1983, pp. 157–182.

[5] England, Peter D. and Richard J. Verrall, "Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving," *Insurance: Mathematics and Economics* 25, 1999, pp. 281–293.

[6] Green, Peter J. and Bernard W. Silverman, *Nonparametric Regression and Generalized Linear Models*, Chapman and Hall, London, 1994.

[7] Hastie, Trevor J. and Robert J. Tibshirani, *Generalized Additive Models*, Chapman and Hall, London, 1990.

[8] Kremer, Erhard, "IBNR Claims and the Two Way Model of ANOVA," *Scandinavian Actuarial Journal*, 1982, pp. 47–55.

[9] Mack, Thomas, "A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves," *ASTIN Bulletin* 22, 1, 1991, pp. 93–109.

[10] Mack, Thomas, "Distribution Free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *ASTIN Bulletin* 23, 2, 1993, pp. 213–225.

[11] Mack, Thomas, "Which Stochastic Model is Underlying the Chain Ladder Model?" *Insurance: Mathematics and Economics* 15, 1994, pp. 133–138.

[12] McCullagh, Peter and John Nelder, Generalized Linear Models, 2nd Edition, Chapman and Hall, 1989.

[13] Renshaw, Arthur E., "Chain Ladder and Interactive Modelling (Claims Reserving and GLIM)," *Journal of the Institute of Actuaries* 116, III, 1989, pp. 559–587.

[14] Renshaw, Arthur E., "Claims Reserving by Joint Modelling," *Actuarial Research Paper No. 72*, 1994, Department of Actuarial Science and Statistics, City University, London, EC1V 0HB.

[15] Renshaw, Arthur E., "On the Second Moment Properties and the Implementation of Certain GLIM Based Stochastic Claims Reserving Models," *Actuarial Research Paper No. 65*, 1994, Department of Actuarial Science and Statistics, City University, London, EC1V 0HB.

[16] Renshaw, Arthur E. and Richard J. Verrall, "A Stochastic Model Underlying the Chain Ladder Technique," *Proceedings XXV ASTIN Colloquium*, Cannes, 1994.

[17] Renshaw, Arthur E. and Richard J. Verrall, "A Stochastic Model Underlying the Chain Ladder Technique," *British Actuarial Journal* 4, IV, 1998, pp. 903–923.

[18] Sherman, Richard E., "Extrapolating, Smoothing and Interpolating Development Factors," *PCAS* LXXI, 1984, pp.122–192.

[19] *S-PLUS User's Guide*, Data Analysis Products Division, Mathsoft, Seattle, WA, 1997.

[20] Taylor, Greg C. and F. R. Ashe, "Second Moments of Estimates of Outstanding Claims," *Journal of Econometrics* 23, 1983, pp. 37–61.

[21] Verrall, Richard J., "A State Space Representation of the Chain Ladder Linear Model," *Journal of the Institute of Actuaries* 116, III, 1989, pp. 589–609.

[22] Verrall, Richard J., "On the Unbiased Estimation of Reserves From Log-Linear Models," *Insurance: Mathematics and Economics* 10, 1991, pp. 75–80.

[23] Verrall, Richard J., "Chain Ladder and Maximum Likelihood," *Journal of the Institute of Actuaries* 118, III, 1991, pp. 489–499.

[24] Verrall, Richard J., "Claims Reserving and Generalised Additive Models," *Insurance: Mathematics and Economics* 19, 1996, pp. 31–43.

[25] Whittaker, Edmund T., "On a New Method of Graduation," *Proceedings of the Edinburgh Mathematical Society* 41, 1923, pp. 63–75.

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

[27] Zehnwirth, Ben, "The Chain Ladder Technique—A Stochastic Model," *Claims Reserving Manual* 2, 1989, Institute of Actuaries, London.

[28] Zehnwirth, Ben, *ICRFS Version 4 Manual and Users Guide*, Benhar Nominees Pty. Ltd., Turramurra, NSW, Australia, 1985.

[29] Zehnwirth, Ben, *Interactive Claims Reserving Forecasting System (ICRFS)*, Insureware P/L, E. St. Kilda, Victoria 3183, Australia, 1991.