A Flexible Framework for Stochastic Reserving Models

by Roger Hayne

ABSTRACT

Maximum likelihood estimators provide a powerful statistical tool. In this paper we directly deal with non-linear reserving models, without the need to transform those models to make them tractable for linear or generalized linear methods. We also show how the same general approach can be easily adapted to provide estimates for a very wide range of reserving methods and models, making use of the same framework, and even much of the same computer code. We focus on the triangle of incremental average costs, and show how five common methods can be set in a stochastic framework.

13628-03_Hayne-3rdPgs.indd 123 9/12/14 9:48 AM

1. Introduction

Traditional approaches to estimating unpaid claim liabilities, an exercise often referred to briefly as "reserving," have usually been deterministic in nature and allowed for the exercise of the practitioner's judgment in arriving at estimates. In the end, an honest practitioner recognized that his or her final estimates were just that, and that final payments would likely differ, potentially materially. The loss process under consideration is generally too complex to be adequately described by a single deterministic forecasting method. Typically, practitioners have used more than one forecasting method in reserving exercises and in fact, in the United States, Actuarial Standard of Practice Number 43 states "The actuary should consider the use of multiple methods" in estimating unpaid claim liabilities. The divergence of forecasts among various methods indicated that the underlying loss process violated the assumptions inherent in one or more of the forecast methods. This provided the practitioner indications of further qualitative or quantitative investigation needed to better understand the underlying loss process. Conversely, if different methods provided reasonably similar forecasts, then the practitioner took some comfort that his or her methods captured the critical characteristics underlying the loss process.

There has long been interest in quantifying the uncertainty inherent in these traditional estimates. Traditional practitioners would often get a qualitative sense of the uncertainty in their estimates by reviewing both the volatility of the data used for forecasting and the range of estimates provided by the various forecast methods used. However, practitioners have sought ways to better describe and quantify this uncertainty inherent in reserving projections and have turned to statistical approaches for the answer.

Probably the most commonly used of all the traditional reserving methods is the development factor approach, commonly referred to as the chain ladder method. This method is straightforward to describe, flexible, and easy to implement computationally. It is not surprising, then, that it would be one of the first methods to be considered in a stochastic framework.

Here we would like to make a clear distinction between a stochastic model and a non-stochastic or deterministic model. In this paper, we will use the term *method* to denote specific description of a process underlying the emergence of amounts over time which allows a practitioner to extrapolate from historical experience. For example, the chain ladder method assumes that cumulative amounts at one age will be a specified percentage of cumulative amounts at the immediate prior age. The implementation of the method involves studying the triangle of development factors sometimes referred to as "link ratios," the ratios of the cumulative amounts at one age to the cumulative amounts at a prior age, selecting a representative factor for each age, then computing a forecast from the product of the cumulative amounts to date with the appropriate selected factors.

By a stochastic reserving *model,* we mean a mathematical statement describing how amounts emerge over time along with an explicit statement regarding the uncertainty or variability of the corresponding amounts. For example, the statement of a method may be that amounts at 24 months will be the amounts at 12 months times a fixed factor. A stochastic *model* based on that statement would be as follows: amounts at 12 and 24 months are random variables A_{12} and A_{24}, respectively, with distributions such that $E(A_{24}) = cE(A_{12})$ for some constant c.

In principle, stochastic models describe the full probability distribution of possible future claim payments. However, such models are often too complex to allow closed-form calculation of even key quantities such as the expected value or standard deviation. Instead, the widespread availability of inexpensive computing power has enabled the use of a number of simulation approaches to estimate uncertainty. Probably the most common tool currently employed is the bootstrap applied to the chain ladder model as discussed by England and Verrall (1999), which provides an estimate of the uncertainty of estimates resulting from the chain ladder model.

13628-03_Hayne-3rdPgs.indd 124 9/12/14 9:48 AM

When reviewing uncertainty in a reserving framework, it is important to recognize precisely what sources of uncertainty are being addressed and what sources are not. From the reserving point of view, total uncertainty can be thought of as having contributions from three sources:

- Process uncertainty—random fluctuations inherent in any stochastic process, even if the process and all its related parameters are known with certainty;
- Parameter uncertainty—the possibility that the parameters of the selected model are estimated incorrectly, even if the selected model is correct; and
- Model uncertainty—the possibility that the amounts to be modeled do not arise from the model assumed.

Works such as England and Verrall (1999) consider only a single model, so even though they may appropriately take into account process and parameter uncertainty, they cannot even begin to account for model uncertainty. Just as practitioners applying traditional reserving methods need to apply a variety of methods to assess uncertainty, those using stochastic approaches should not ignore model uncertainty.

In this paper, rather than relying on linear or generalized linear models, as do many of the papers using the bootstrap method, we take advantage of the broad availability of relatively cheap computing power to look directly at non-linear reserving models and make use of the maximum likelihood estimators (MLEs) to derive estimates of both process and parameter uncertainty. We will also use this same framework to consider a range of different models and begin to explore ways we can at least start to recognize model uncertainty in our estimates.

2. Maximum likelihood estimators

Klugman, Panjer and Willmot (1998) present a very clear and concise discussion of the MLE approach. Rather than repeating that exposition here, we will briefly summarize the concept of maximum likelihood estimate and the principle conclusions of the primary theorem appearing on page 62 of Klugman, Panjer and Willmot (1998). We will refer to this result as Theorem 1 in what follows.

Suppose we have a sample drawn from a distribution whose general form is known and described by a set of parameters, but the values of those parameters are themselves unknown. The task then is to estimate those parameters given the sample drawn. One approach to this problem is based on the relative likelihood of drawing the sample $\mathbf{X}$ given a particular choice of the parameters $\boldsymbol{\theta}$, calling this likelihood $\mathrm{L}(\mathbf{X}|\boldsymbol{\theta})$. The maximum likelihood estimator or MLE is the value of $\boldsymbol{\theta}$ that gives the largest value for $\mathrm{L}(\mathbf{X}|\boldsymbol{\theta})$ over all possible choices for $\boldsymbol{\theta}$. Under some mild regularity conditions on the distribution under consideration, the MLE has the following asymptotic properties as the sample size grows to infinity:

1. It asymptotically tends to the true value of the unknown parameters.
2. Its variance asymptotically tends to a value that is no larger than the variance of any other estimator of those parameters.
3. It asymptotically approaches a Gaussian distribution.

The asymptotic Gaussian distribution has mean $\boldsymbol{\theta}$ and covariance matrix equal to the inverse of the Fisher information matrix, which has elements given in (2.1).

$$\mathbf{I}(\boldsymbol{\theta})_{ij} = \mathrm{E}_{\mathbf{X}}\left(\frac{\partial^2}{\partial\theta_i\,\partial\theta_j} - \ln(\mathrm{L}(X|\boldsymbol{\theta}))\right) \tag{2.1}$$

In practice, both the mean and covariance matrix of the limiting Gaussian distribution are calculated assuming that the actual parameter values are equal to the MLE. We will use this property in incorporating parameter uncertainty in the stochastic models we present here.

3. General stochastic reserving model

We will focus on the usual triangular array of amounts, where the amount could be claim counts, paid losses, or case-incurred losses (paid losses plus

13628-03_Hayne-3rdPgs.indd 125 9/12/14 9:48 AM

claim adjuster estimates of outstanding losses). In the remainder of this paper we will use the term "payments" for simplicity, but do not intend any loss of generality. If we are interested in incurred losses then "payments made" during a time span would translate to "change in incurred losses" during that time. Similar adaptations can be made for claim counts or other amounts of interest.

We will denote by C_{ij} payments made for exposure period i during development period j. Here the time spans are unspecified and can represent quarters, half years, years, or other time spans of interest. Similarly, the exposure period could represent a policy period, accident period, underwriting period, or other span of interest. As with payments, we will simply refer to accident years and annual development in what follows.

We will assume that the time span of interest covers m accident years and n years of development. We will denote by n_i the age of the most recent available experience for accident year i. In the case of a square data set with $m = n$ accident years and development years, then $n_i = n - i + 1$.

In addition to the basic paid amount data, we will assume we have some measure of exposure for each accident year, either an exposure count, premium amount, or an estimate of ultimate claim counts. We note that ultimate claim counts often cannot be known with certainty and as such should be treated as random variables. We will not make that generalization here but rather leave it as a future project. In any event, we will denote this measure of relative exposure as W_i for accident year i. We will thus focus on the incremental averages A_{ij} defined by Equation (3.1).

$$A_{ij} = \frac{C_{ij}}{W_i}. \tag{3.1}$$

The incorporation of this additional exposure information widens the variety of methods available for analysis. We can now move beyond the simple chain ladder to include others, such as a version of the Cape Cod variant of the Bornhuetter-Ferguson method, an incremental severity method discussed by Berquist and Sherman (1977), and a variant of the Hoerl curve method discussed by Wright (1992).

At this point, we will assume that the reserving method can be expressed as a matrix-valued function of a parameter vector $\boldsymbol{\theta}$, as expressed in (3.2).

$$A_{ij} = g_{ij}(\boldsymbol{\theta}). \tag{3.2}$$

We will turn this simple method or "recipe" into a stochastic model by introducing random variables with specific probability distributions. It is not unusual to assume that the variances of the incremental amounts are proportional to a power of their expected values (for example, see Venter 2007). We will take this same approach. However, since we will allow the expected values to be negative, we will, without loss of generality, take the variance to be proportional to a power of the square of the mean. Also, we are taking the constant of proportionality among the variances as an exponential, thereby allowing the parameter to take on any value. In addition, we note that the variance of an average of a sample independently drawn from a population with a finite variance will be inversely proportional to the number of items in the sample, so we will take the constant of proportionality to also vary inversely to the number of exposures.

Following the notation in Venter (2007), we will state our stochastic model as shown in (3.3), suppressing subscripts for the moment and letting w denote the natural log of the exposure measure for an accident year.

$$\begin{aligned} \mathrm{E}(A) &= \mu, \\ \mathrm{Var}(A) &= \frac{e^{\kappa}(\mu^2)^p}{W} = e^{\kappa - w}(\mu^2)^p. \end{aligned} \tag{3.3}$$

Note that this model includes an implicit structural heteroscedasticity. The expected values will differ by accident year and development year. Since the variance is a function of the expected value, it will likewise differ by accident and development year. By

13628-03_Hayne-3rdPgs.indd 126 9/12/14 9:48 AM

combining this structure with two variance parameters, κ and p, that can be fit to the data, we provide a mechanism for the variance structure of the model to approximate the variance structure of the data without over-parameterizing our model. We note, though, that the formulae we present here can very easily be modified to allow κ to vary by development period if additional control over the heteroscedasticity is desired.

We note the incremental amounts A under consideration are averages of a number of observations. If we assume the observations are themselves independent, then the central limit theorem would imply that they have asymptotically Gaussian distributions. For this reason, we will assume that the A variables are all independent and have Gaussian distributions given in (3.4), again suppressing subscripts for the moment.

$$A \sim \mathrm{N}\left(\mu, e^{\kappa - w}(\mu^2)^p\right). \tag{3.4}$$

Since we are concerned with MLEs, the negative log likelihood for this distribution will be key to our analysis. The likelihood function for a Gaussian distribution is given by (3.5).

$$\mathrm{f}(x; \mu, \kappa, p) = \frac{1}{\sqrt{2\pi e^{\kappa - w}(\mu^2)^p}} e^{-\frac{(x-\mu)^2}{2e^{\kappa - w}(\mu^2)^p}}. \tag{3.5}$$

This gives a negative log likelihood for a single variable given in (3.6).

$$\begin{aligned} \mathrm{l}(x; \mu, \kappa, p) &= -\ln(\mathrm{f}(x; \mu, \kappa, p)) \\ &= -\ln\left(\frac{1}{\sqrt{2\pi e^{\kappa - w}(\mu^2)^p}} \exp\left(-\frac{(x-\mu)^2}{2e^{\kappa - w}(\mu^2)^p}\right)\right) \\ &= \frac{1}{2}\left(\kappa - w + \ln\left(2\pi(\mu^2)^p\right)\right) + \frac{(x-\mu)^2}{2e^{\kappa - w}(\mu^2)^p}. \end{aligned} \tag{3.6}$$

Using the relationships in (3.2) and (3.4), we have the stochastic statement of our reserving model as shown in (3.7).

$$A_{ij} \sim \mathrm{N}\left(g_{ij}(\boldsymbol{\theta}), e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p\right). \tag{3.7}$$

In formula (3.7) the added constants w_i are included to reflect the fact that the variance of an average of a sample is dependent on the number of elements composing that sample. At this point, if we wanted the constant of proportionality to vary by development lag we would simply substitute a vector for the constant parameter κ.

With observations in a typical loss triangle, we get the negative log likelihood function given in (3.8).

$$\begin{aligned} &\ell(A_{11}, A_{12}, \ldots, A_{m1}; \boldsymbol{\theta}, \kappa, p) \\ &\quad = \frac{1}{2}\sum_{(i,j)\in S} \kappa - w_i + \ln\left(2\pi\left(g_{ij}(\boldsymbol{\theta})^2\right)^p\right) \\ &\qquad + \frac{\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}. \end{aligned} \tag{3.8}$$

The set S in (3.8) denotes the set of all index pairs for which data are available. We denote by T the index pairs over which we want forecasts. If the data were available in a full triangle, with n rows and n columns, then S and T would follow the form given in (3.9).

$$\begin{aligned} S &= \{(i, j) \mid i = 1, 2, \ldots, n, j = 1, 2, \ldots, n - i + 1\} \\ T &= \{(i, j) \mid i = 2, \ldots, n, j = n - i + 2, \ldots, n\} \end{aligned} \tag{3.9}$$

However, we do not need to restrict ourselves to this regular case and will use the more flexible notation.

We select the values of the parameters to be the MLEs, the values that minimize the negative log likelihood function in (3.8). Let us denote these estimates by $\hat{\boldsymbol{\theta}}$, $\hat{\kappa}$, and $\hat{p}$. In practice there are a number of tools available to estimate these parameters. We have used the *nlminb* function in the package MASS written for the statistical programming language and environment R (R Core Team 2012) for this purpose, though other tools might be just as useful. We used analytic rather than numeric representations for the various derivatives to increase the accuracy and speed of the calculation.

If we assume that there is no model or parameter uncertainty, so that the model parameterized

13628-03_Hayne-3rdPgs.indd 127 9/12/14 9:48 AM

with the MLEs gives an exact description of the true loss emergence process, then it is straightforward to obtain estimates of the distribution of outcomes. We have assumed that, for fixed values of the parameters, the variables A_{ij} are independent. Thus, in the absence of parameter or model uncertainty that could introduce correlations between the variables, we can conclude that the distribution of average future payments for each year is given by (3.10) and that the distribution of total future payments is given by (3.11).

$$R_i \sim \mathrm{N}\left(W_i \sum_{\{j|(i,j)\in T\}} g_{ij}\left(\hat{\boldsymbol{\theta}}\right), W_i^2 \sum_{\{j|(i,j)\in T\}} e^{\hat{\kappa}-w_i}\left(g_{ij}\left(\hat{\boldsymbol{\theta}}\right)^2\right)^{\hat{p}} \right). \tag{3.10}$$

$$R_T \sim \mathrm{N}\left(\sum_{i=1}^{m} W_i \sum_{\{j|(i,j)\in T\}} g_{ij}\left(\hat{\boldsymbol{\theta}}\right), \sum_{i=1}^{m} W_i^2 \sum_{\{j|(i,j)\in T\}} e^{\hat{\kappa}-w_i}\left(g_{ij}\left(\hat{\boldsymbol{\theta}}\right)^2\right)^{\hat{p}} \right). \tag{3.11}$$

This then gives the effect of process uncertainty on the total forecast incremental averages by accident year. This does not, however, address the issue of parameter uncertainty.

Just as the standard error provides insight into parameter uncertainty in usual regression applications, the information matrix can be helpful in estimating the variance-covariance matrix of the parameters. Since we have based our model on a Gaussian distribution, the conditions of Theorem 1 will be met as long as the functions g_{ij} are sufficiently regular, which will be the case for the examples we will consider. Thus, to estimate the variance-covariance matrix for the parameters, we first define the Fisher information matrix as the matrix of expected values of the Hessian of the negative log likelihood function.

First, the Hessian is the matrix whose element in the *ith* row and the *jth* column is the second derivative of the negative log likelihood function, once with respect to the *ith* variable and once with respect to the *jth* as shown in (3.12), assuming the vector $\boldsymbol{\theta}$ has k elements.

$$\text{Hessian } \ell = \begin{pmatrix} \frac{\partial^2 \ell}{\partial \theta_1^2} & \cdots & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_k} & \frac{\partial^2 \ell}{\partial \theta_1 \partial \kappa} & \frac{\partial^2 \ell}{\partial \theta_1 \partial p} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_1} & \cdots & \frac{\partial^2 \ell}{\partial \theta_k^2} & \frac{\partial^2 \ell}{\partial \theta_k \partial \kappa} & \frac{\partial^2 \ell}{\partial \theta_k \partial p} \\ \frac{\partial^2 \ell}{\partial \kappa \partial \theta_1} & \cdots & \frac{\partial^2 \ell}{\partial \kappa \partial \theta_k} & \frac{\partial^2 \ell}{\partial \kappa^2} & \frac{\partial^2 \ell}{\partial \kappa \partial p} \\ \frac{\partial^2 \ell}{\partial p \partial \theta_1} & \cdots & \frac{\partial^2 \ell}{\partial p \partial \theta_k} & \frac{\partial^2 \ell}{\partial p \partial \kappa} & \frac{\partial^2 \ell}{\partial p^2} \end{pmatrix} \tag{3.12}$$

Thus the information matrix, which is the expected value of this Hessian, evaluated at the parameter estimates is given in (3.13).

$$\mathbf{I}\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right) = \begin{pmatrix} \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \theta_1^2}\right) & \cdots & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \theta_1 \partial \theta_k}\right) & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \theta_1 \partial \kappa}\right) & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \theta_1 \partial p}\right) \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \theta_k \partial \theta_1}\right) & \cdots & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \theta_k^2}\right) & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \theta_k \partial \kappa}\right) & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \theta_k \partial p}\right) \\ \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \kappa \partial \theta_1}\right) & \cdots & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \kappa \partial \theta_k}\right) & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \kappa^2}\right) & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial \kappa \partial p}\right) \\ \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial p \partial \theta_1}\right) & \cdots & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial p \partial \theta_k}\right) & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial p \partial \kappa}\right) & \mathrm{E}\left(\frac{\partial^2 \ell\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)}{\partial p^2}\right) \end{pmatrix} \tag{3.13}$$

13628-03_Hayne-3rdPgs.indd 128 9/12/14 9:48 AM

We show these expectations, along with both the elements of the gradient and the Hessian of the negative log likelihood function in terms of the general model functions g_{ij} in Appendix A to this paper. The inverse of the information matrix is then an approximation for the variance-covariance matrix for the parameters. In particular, then, an estimate of the standard error of the various parameters is given by the square root of the diagonal of that matrix, that is by $\sqrt{\operatorname{diag}\left(\mathbf{I}\left(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p}\right)^{-1}\right)}$. If we are willing to assume, by virtue of Theorem 1 above, that the parameters asymptotically have a multivariate Gaussian distribution with expected value $(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p})$ and covariance matrix $\mathbf{I}(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p})^{-1}$, then the problem of estimating the distribution of reserve forecasts reduces to estimating the distribution of a random variable whose parameters have a known distribution, a classical problem of estimating a posterior distribution.

There are a number of approaches that can be applied to that problem. Given that we have not placed any real restrictions on the general expected value model, other than having the second derivatives existing, an analytic solution is not easy to obtain. A Markov chain Monte Carlo (MCMC) method such as the Gibbs sampler as discussed by Scollnik (1996) or the Metropolis-Hastings algorithm as discussed by Meyers (2009) could be used for this purpose. However, our assumption that the parameters have a multivariate Gaussian distribution makes a more straightforward approach available. We use direct Monte Carlo simulation to estimate the reserve distribution. We first randomly select a parameter vector $(\boldsymbol{\theta}^*, \kappa^*, \hat{p}^*)$ from a multivariate Gaussian distribution with expected value $(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p})$ and covariance matrix $\mathbf{I}(\hat{\boldsymbol{\theta}}, \hat{\kappa}, \hat{p})^{-1}$ and then randomly select Gaussian variables R_i^* from Gaussian distributions with mean $E_i \sum_{\{j|(i,j)\in T\}} g_{ij}(\boldsymbol{\theta}^*)$ and variance $E_i^2 \sum_{\{j|(i,j)\in T\}} e^{\kappa^*-w_i}\left(g_{ij}(\boldsymbol{\theta}^*)^2\right)^{p^*}$ for each value of i.

At this juncture, if we wished to assume that claim counts were stochastic but independent of the incremental severities, we could simulate the ultimate number of claims by exposure year to add a provision for uncertainty in those estimates in the final forecast.

This then sets a general framework that can be applied to any loss model that is sufficiently differentiable. The next section of this paper will look at five specific examples of commonly applied reserving methods adapted to this framework.

4. Example models

In this section we will show how five commonly used reserving methods can be readily adapted to stochastic models in this framework. The five models we select are a variant of the Bornhuetter-Ferguson method that is akin to the Stanard-Buhlmann or Cape Cod method, a variation of the incremental severity model presented by Berquist and Sherman (1977), two variations of the Hoerl curve models presented by Wright (1992), and a formulation of the traditional chain ladder or development factor method.

We will show the application of these five models to aggregate commercial automobile liability paid loss and defense and cost containment expenses, net of reinsurance, as reported in the 2010 Schedule P by ten U.S. insurers all writing $50 million or more in net premiums for that line in 2010, as shown in Table 1. We used a standard chain ladder applied to reported counts for these carriers to derive the claim count estimates.

4.1. Cape Cod model

Bornhuetter and Ferguson (1972) estimate the future incurred losses for an accident year as a percentage of an a priori estimate of ultimate losses for that year. The percentage is based on historical incurred loss development patterns, and the a priori estimate is based on a selected loss ratio times the premiums for that year. The distinguishing feature of the Cape Cod variant of this method, as presented, for example, by Stanard (1985), is that it derives the a priori loss estimates directly from the data. We will use the term "Cape Cod" in this paper to refer to an approach that has the same general structure of the Bornhuetter-Ferguson approach but derives its a priori ultimate estimates by

13628-03_Hayne-3rdPgs.indd 129 9/12/14 9:48 AM

Table 1. 2010 Schedule P commercial automobile liability data for selected companies

	Cumulative Average Paid Loss & Defense & Cost Containment Expenses per Estimated Ultimate Claim										
Accident	Months of Development										Count
Year	12	24	36	48	60	72	84	96	108	120	Forecast
2001	670	1,480	1,939	2,466	2,838	3,004	3,055	3,133	3,141	3,160	39,161
2002	768	1,593	2,464	3,020	3,375	3,554	3,602	3,627	3,646		38,672
2003	741	1,616	2,346	2,911	3,202	3,418	3,507	3,529			41,801
2004	862	1,755	2,535	3,271	3,740	4,003	4,125				42,263
2005	841	1,859	2,805	3,445	3,950	4,186					41,481
2006	848	2,053	3,076	3,861	4,352						40,214
2007	902	1,928	3,004	3,881							43,599
2008	935	2,104	3,182								42,118
2009	759	1,585									43,479
2010	723										49,492

accident year from the loss and exposure data alone, without reference to outside information. The specific model we use may not necessarily correspond to others used that are also called Cape Cod.

To this purpose, then, we will assume that the incremental average amounts are the product of two different factors, one representing the accident year influence, usually taken as the ultimate losses for the year, and the other corresponding to the lag, usually taken as the percentage of losses emerging that year. With this simple formulation, there are infinitely many sets of parameters that give the same model, since multiplying all the accident year parameters by a constant and dividing all the development year parameters by the same constant gives the same result. For that reason we selected the slightly more complicated parameterization for this model shown in (4.1) with a total of $m + n - 1$ parameters.

$$g_{ij}(\boldsymbol{\theta}) = \begin{cases} \theta_1 \text{ if } i = j = 1 \\ \theta_1\theta_i \text{ if } j = 1 \text{ and } i > 1 \\ \theta_1\theta_{m+j-1} \text{ if } i = 1 \text{ and } j > 1 \\ \theta_1\theta_i\theta_{m+j-1} \text{ if } i > 1 \text{ and } j > 1 \end{cases} \tag{4.1}$$

Just as with the usual Cape Cod method, this model uses the data to determine the accident year expected loss rather than basing those estimates on sources outside the data as in the Bornhuetter-Ferguson method. It therefore allows a separate loss level for each accident year. Table 2 shows the parameter estimates for this model.

In addition to the parameter estimates, Table 2 shows estimates of the standard error of the parameters, taken as the square root of the diagonal of the covariance matrix for the parameters. The standard error is a measure of the standard deviation of particular parameters given the observed data. A large standard error rela-

Table 2. Cape Cod model parameter estimates

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}
Parameter	620.07	1.1603	1.1232	1.3222	1.3757	1.5208	1.5333	1.5800	1.1695	1.1635	1.1805
Std. Error	30.048	0.066	0.064	0.072	0.075	0.082	0.084	0.091	0.082	0.105	0.041
	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_6	θ_{17}	θ_{18}	θ_{19}	κ	p	AIC
Parameter	1.063	0.838	0.534	0.284	0.111	0.067	0.015	0.024	13.105	0.435	619.32
Std. Error	0.040	0.036	0.029	0.023	0.016	0.016	0.009	0.017	1.010	0.083	

13628-03_Hayne-3rdPgs.indd 130 9/12/14 9:48 AM

tive to a parameter estimate may indicate significant uncertainty around that estimate.

We would also like to be able to compare the fits of different models to get an assessment of which models seem to better reflect patterns in the observed data. One measure that is comparable among models is the likelihood of the data given the fitted model. However, one can often increase this likelihood by adding parameters to a model, so a useful comparison should also take into account the number of parameters necessary to obtain a particular fit. The Akaike Information Criterion (AIC; see, for example, Venables and Ripley 2002), is one such statistic that addresses both of these points, particularly when comparing models with the same underlying probability structure as we are doing here. As used in these calculations the AIC is twice the negative log likelihood valued at the selected parameters plus twice the number of parameters estimated. Thus a smaller value of the AIC tends to indicate a better fit.

Table 3 shows the expected average cost forecasts as well as indicated standard deviations by cell implied by the model, *but ignoring parameter uncertainty.*

As mentioned above, we make use of the Fisher Information Matrix to estimate covariance among the parameter estimates. Thus the distribution of forecasts from the Cape Cod method reflects both the process uncertainty shown in Table 3 and the parameter uncertainty captured by the Fisher information matrix. Table 4 summarizes the forecasts for total future payments, both by accident year and in total as well as forecasts for the next calendar year. In this table we

Table 3. Cape Cod model estimates

Accident Year	Months of Development									Total
	24	36	48	60	72	84	96	108	120	
					Forecast Expected Average Losses					
2001										
2002									17.47	17.47
2003								10.70	16.91	27.61
2004							55.18	12.60	19.90	87.68
2005						94.63	57.41	13.10	20.71	185.85
2006					268.01	104.62	63.47	14.49	22.89	473.48
2007				507.50	270.22	105.48	63.99	14.61	23.08	984.87
2008			820.62	522.95	278.44	108.69	65.94	15.05	23.78	1,835.47
2009		770.74	607.41	387.08	206.10	80.45	48.81	11.14	17.60	2,129.33
2010	851.72	766.81	604.31	385.10	205.05	80.04	48.56	11.08	17.51	2,970.19
					Forecast Standard Deviation					
2001										
2002									12.37	12.37
2003								9.61	11.73	15.17
2004							19.52	10.26	12.52	25.36
2005						24.91	20.04	10.54	12.86	36.04
2006					39.79	26.43	21.26	11.18	13.65	55.18
2007				50.45	38.35	25.47	20.49	10.78	13.15	73.31
2008			63.27	52.01	39.53	26.26	21.13	11.11	13.56	98.55
2009		60.59	54.63	44.91	34.14	22.67	18.24	9.59	11.71	104.47
2010	59.32	56.67	51.09	42.00	31.92	21.20	17.06	8.97	10.95	114.30

13628-03_Hayne-3rdPgs.indd 131 9/12/14 9:48 AM

Table 4. Future payment estimates for the Cape Cod model

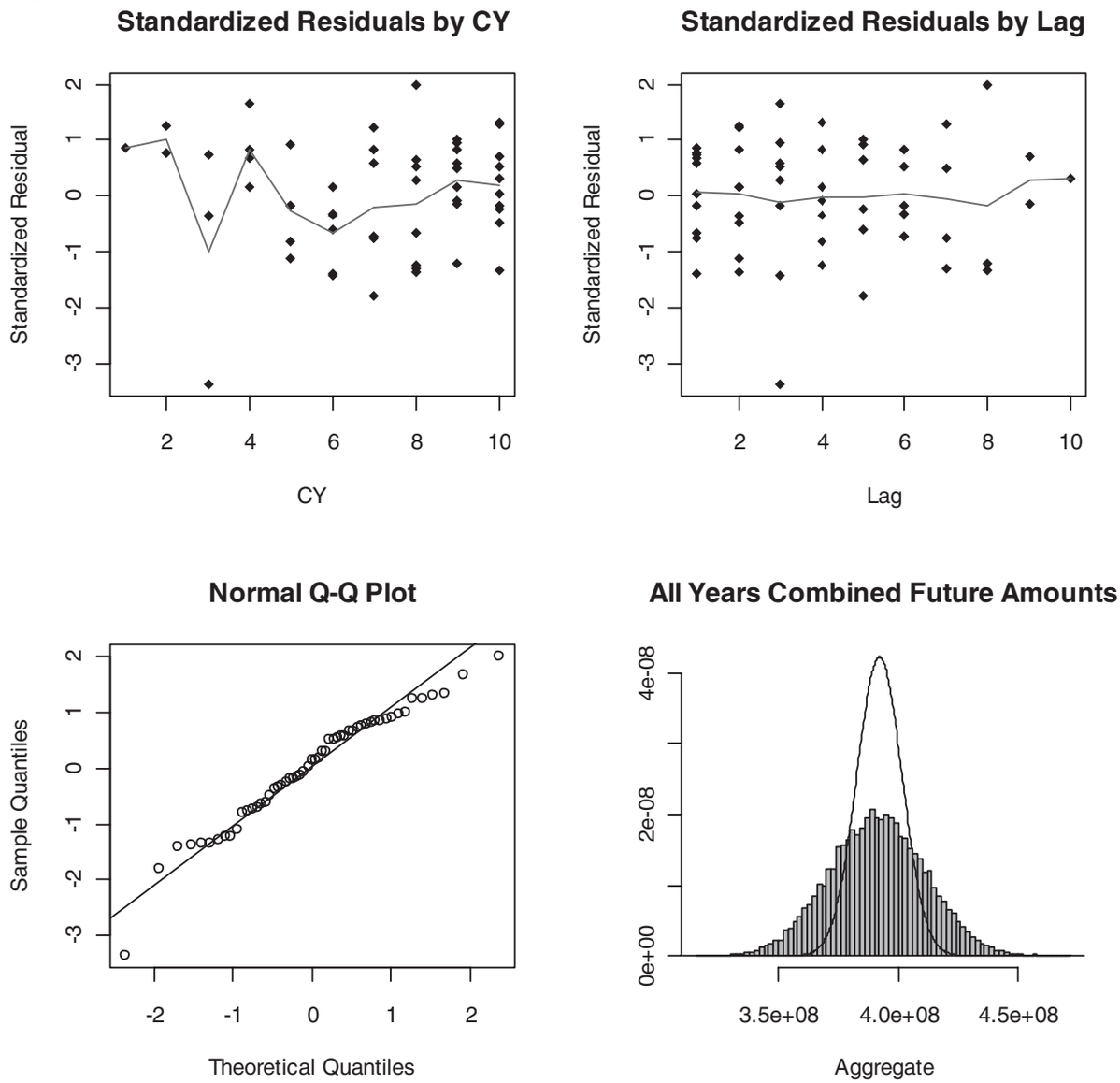
Accident Year	Process Only		Including Parameter Uncertainty			
					Percentile	
	Mean	Standard Deviation	Mean	Standard Deviation	5%	95%
			Total Estimated Accident Year Unpaid Loss Estimates			
2001	0	0	0	0	0	0
2002	675,486	478,362	671,796	696,527	–320,924	1,913,574
2003	1,154,053	633,975	1,152,202	890,392	–222,454	2,664,759
2004	3,705,613	1,071,770	3,699,395	1,426,606	1,397,182	6,059,300
2005	7,709,391	1,494,876	7,677,680	1,917,178	4,559,428	10,852,708
2006	19,040,687	2,219,174	18,975,975	2,776,688	14,419,234	23,599,264
2007	42,938,949	3,196,213	42,875,071	4,099,182	36,166,422	49,674,115
2008	77,307,053	4,150,959	77,161,212	5,779,893	67,818,676	86,842,885
2009	92,581,984	4,542,194	92,324,203	7,389,974	80,291,551	104,664,146
2010	147,002,025	5,656,751	146,768,931	13,408,467	124,818,538	168,746,509
Total	392,115,241	9,434,799	391,306,466	20,297,820	357,781,810	424,885,057
			Estimates of Loss Payments in Next Calendar Year			
2001	0	0	0	0	0	0
2002	675,486	478,362	671,796	696,527	–320,924	1,913,574
2003	447,277	401,848	447,581	511,066	–294,577	1,335,516
2004	2,332,108	824,843	2,323,106	1,004,849	739,639	4,020,687
2005	3,925,311	1,033,279	3,914,543	1,215,317	1,965,746	5,954,827
2006	10,777,934	1,600,217	10,733,932	1,854,372	7,733,279	13,836,243
2007	22,126,161	2,199,643	22,098,025	2,607,290	17,841,494	26,464,229
2008	34,563,211	2,664,703	34,493,116	3,275,767	29,161,789	39,930,102
2009	33,511,431	2,634,555	33,388,362	3,427,547	27,901,041	39,111,620
2010	42,153,713	2,935,691	42,106,937	4,627,754	34,603,882	49,789,830
Total	150,512,633	5,674,264	150,177,398	7,616,666	137,692,029	162,703,904

show the total estimated future loss payments and standard deviation by accident year and in total for all accident years implied by the model, *both with and without consideration of parameter uncertainty.* Also shown are the results of our simulation of the effects of parameter uncertainty on those amounts along with the 90% confidence interval implied by that simulation. The lower half of Table 4 shows the same information for the payments estimated to be made in the next year. This information can provide a test for how well emerging data fits the model. If payments next year fall outside the range indicated, then there may be cause to question whether the Cape Cod model is the appropriate one for this data set.

As with other stochastic models, it is often helpful to visualize the model's fit both in terms of standardized residuals and Q-Q plots. The former help inform if the model is missing significant patterns in the data, while the latter test how well the model captures the statistical characteristics of the data. Generally, a Q-Q plot that follows a straight line indicates that the actual variability observed in the data is consistent with the form of the model. Deviation from a straight line may indicate the actual data have different "tail" characteristics than what is assumed in the model. Figure 1 gives four different charts resulting from our Cape Cod model fit. Calendar year influences may show in the Standardized Residuals by CY plot, while tests of

13628-03_Hayne-3rdPgs.indd 132 9/12/14 9:48 AM

Figure 1. Cape Cod model charts



how well the model fits in the tail are shown in the Standardized Residuals by Lag plot. The normal Q-Q plot shows how close to a Gaussian (straight line) the standardized residuals fall. Finally, the distributions of total reserve forecasts both with (histogram) and without (line) parameter uncertainty are shown graphically.

Appendix B shows the derivatives for this formulation of the Cape Cod model.

All these calculations were carried out using the open-source statistical package R. The code used is shown in Appendix G of this paper.

4.2. Berquist-Sherman incremental severity model

The Cape Cod model has the largest number of parameters of the five models we discuss here. Berquist and Sherman (1977) recognized that average costs may exhibit a reasonably predictable trend over the experience period in question. They developed methods to model incremental severities that allowed for not only different loss levels by lag, as in the Cape Cod model, but also different trends by payment lag as well. We simplify their

13628-03_Hayne-3rdPgs.indd 133 9/12/14 9:48 AM

approach here somewhat and replace the separate loss levels for each accident year by a single parameter, trend from one accident year to the next, and assume that this trend is constant through the entire time frame to be modeled, effectively replacing the m accident year parameters by a single trend parameter for a total of $n + 1$ parameters to be estimated using the data. We thus use the formulation in (4.2) for the Berquist-Sherman incremental severity model.

$$g_{ij}(\boldsymbol{\theta}) = \theta_j e^{i\theta_{n+1}}, i = 1, 2, \ldots, m, j = 1, 2, \ldots, n \quad (4.2)$$

Table 5 through Table 7 and Figure 2 show the results for this model, similar to Table 2 through Table 4 and Figure 1.

Table 5. Berquist-Sherman model parameter estimates

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}
Parameter	620.96	760.66	708.16	553.57	350.00	181.39	70.96	43.88	11.08	15.21	0.0452
Std. Error	40.498	46.552	43.004	35.491	26.169	17.662	10.390	8.735	4.224	7.343	0.0086

	κ	p	AIC
Parameter	11.216	0.6539	643.45
Std. Error	1.0368	0.0846	

Table 6. Berquist-Sherman model estimates

Accident Year	Months of Development									Total
	24	36	48	60	72	84	96	108	120	
					Forecast Expected Average Losses					
2001										
2002									16.65	16.65
2003								12.69	17.42	30.10
2004							52.57	13.27	18.22	84.07
2005						88.96	55.00	13.89	19.07	176.92
2006					237.91	93.08	57.55	14.53	19.95	423.01
2007				480.27	248.91	97.38	60.21	15.20	20.87	922.84
2008			794.75	502.48	260.42	101.88	62.99	15.90	21.84	1,760.27
2009		1,063.70	831.50	525.72	272.46	106.59	65.91	16.64	22.85	2,905.37
2010	1,195.40	1,112.89	869.95	550.03	285.06	111.52	68.95	17.41	23.90	4,235.12
					Forecast Standard Deviation					
2001										
2002									8.72	8.72
2003								7.02	8.64	11.13
2004							17.69	7.19	8.85	21.05
2005						25.19	18.39	7.48	9.20	33.37
2006					48.67	26.35	19.24	7.82	9.62	59.90
2007				74.00	48.15	26.07	19.03	7.74	9.52	94.80
2008			104.66	77.55	50.46	27.32	19.95	8.11	9.98	144.30
2009		124.63	106.09	78.61	51.15	27.69	20.22	8.22	10.11	192.18
2010	126.08	120.32	102.42	75.89	49.38	26.73	19.52	7.94	9.76	224.31

13628-03_Hayne-3rdPgs.indd 134 9/12/14 9:48 AM

Table 7. Future payment estimates for the Berquist-Sherman model

	Total Estimated Accident Year Unpaid Loss Estimates					
	Process Only		Including Parameter Uncertainty			
					Percentile	
Accident Year	Mean	Standard Deviation	Mean	Standard Deviation	5%	95%
			Total Estimated Accident Year Unpaid Loss Estimates			
2001	0	0	0	0	0	0
2002	643,872	334,239	645,956	483,711	–41,004	1,509,475
2003	1,258,405	465,360	1,257,727	654,156	235,852	2,363,675
2004	3,553,041	889,597	3,554,492	1,140,890	1,725,969	5,467,024
2005	7,338,748	1,384,118	7,337,109	1,679,907	4,625,974	10,107,040
2006	17,011,030	2,408,663	17,031,317	2,805,350	12,489,148	21,635,432
2007	40,234,557	4,133,010	40,212,331	4,762,548	32,478,281	48,188,095
2008	74,139,470	6,077,538	74,158,913	7,280,489	62,403,876	86,346,221
2009	126,323,651	8,355,660	126,307,981	10,575,695	109,238,192	143,818,772
2010	209,606,332	11,101,836	209,681,729	15,864,532	184,115,098	236,297,303
Total	480,109,106	15,997,662	480,187,555	29,089,899	433,504,594	528,833,729
			Estimates of Loss Payments in Next Calendar Year			
2001	0	0	0	0	0	0
2002	643,872	337,239	645,956	483,711	–41,004	1,509,475
2003	530,262	293,501	531,608	374,295	–30,941	1,181,397
2004	2,221,885	747,725	2,221,462	899,807	817,280	3,752,606
2005	3,690,224	1,044,868	3,697,592	1,197,655	1,784,899	5,723,241
2006	9,567,350	1,957,360	9,573,094	2,199,851	6,051,225	13,261,677
2007	20,939,191	3,226,274	20,909,924	3,597,284	15,114,772	26,954,489
2008	33,473,518	4,407,909	33,482,822	4,903,004	25,585,777	41,665,309
2009	46,249,116	5,418,932	46,219,106	6,120,291	36,363,999	56,379,075
2010	59,163,420	6,240,073	59,128,030	7,212,388	47,462,722	71,104,164
Total	176,478,837	10,189,397	176,409,595	12,632,905	156,084,211	197,512,110

Appendix C shows the derivatives for this formulation of the Berquist-Sherman incremental severity model.

4.3. Wright's model

The Berquist-Sherman model presented here replaced the Cape Cod's accident year level factors with a uniform annual trend. That model still maintained separate parameters for each development lag. Just as fitting a line replaces a number of points with a line defined by two parameters, Wright (1992) made use of a curve to replace the multiple lag parameters with a fixed number of parameters. Wright considered two similar curves representing loss volume as a year aged, using the variable τ to represent what he calls "operational time." Two of the curves he considered are shown in (4.3).

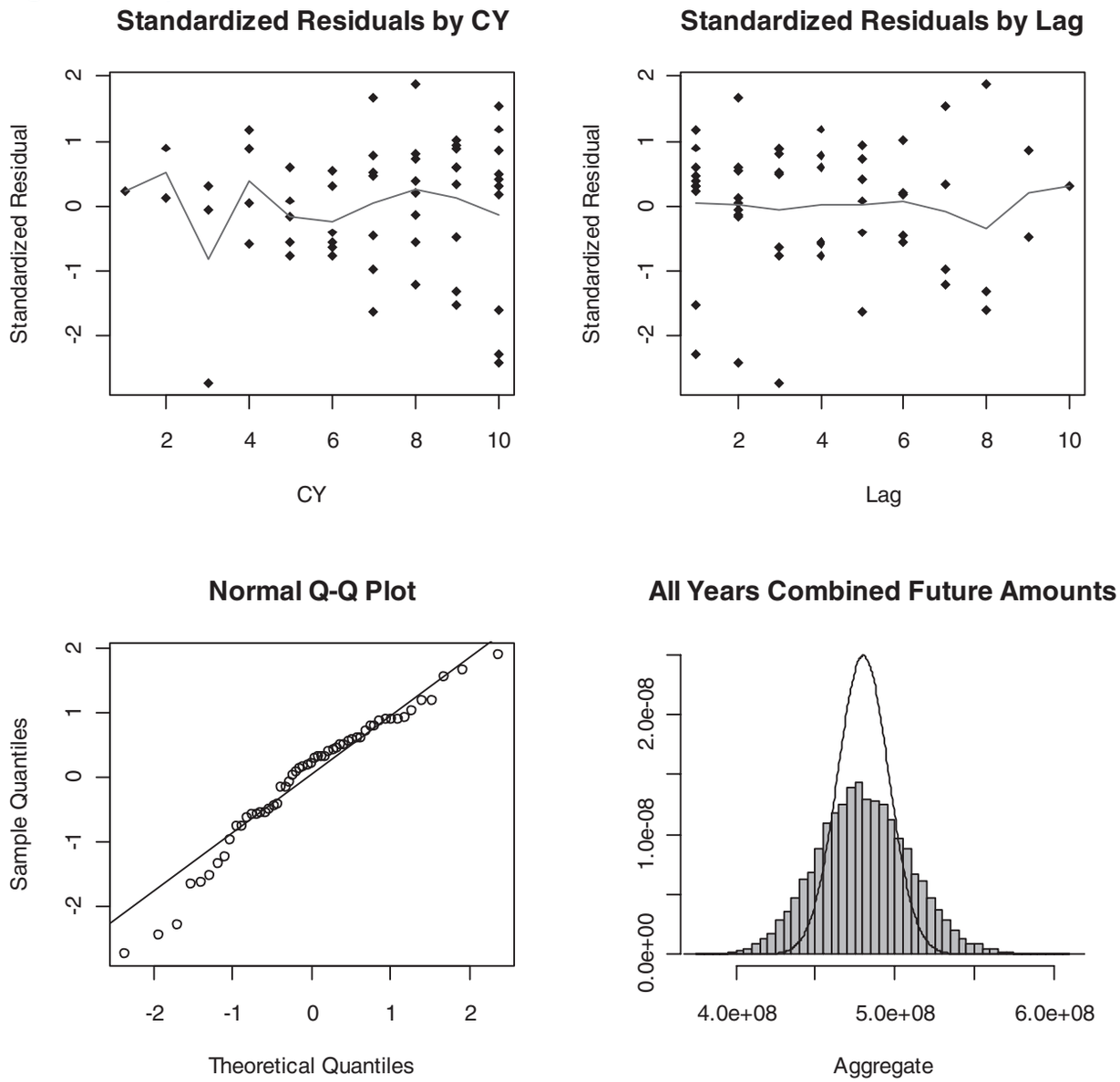
$$\exp(\beta_0 + \beta_1\tau + \beta_2 \ln(\tau)) \qquad \exp(\beta_0 + \beta_1\tau + \beta_2\tau^2) \tag{4.3}$$

Both of these are special cases of the form shown in (4.4).

$$\exp(\beta_0 + \beta_1\tau + \beta_2\tau^2 + \beta_3 \ln(\tau)) \tag{4.4}$$

13628-03_Hayne-3rdPgs.indd 135 9/12/14 9:48 AM

Figure 2. Berquist-Sherman model charts



Incorporating this curve with separate level parameters by accident year, we call the function in (4.5) Wright's model.

$$g_{ij}(\theta) = \exp\left(\theta_i + \theta_{m+1} j + \theta_{m+2} j^2 + \theta_{m+3} \ln(j)\right), \quad i = 1, \ldots, m,\ j = 1, \ldots, n \qquad (4.5)$$

We note that, in contrast to the prior two models, this requires that the expected losses be positive in all cells. We also note that we use the formulation from England and Verrall (2001), which uses the age directly rather than the operational time variant used by Wright (1992). If we wished to replace the parameter j in (4.5) by a more general operational time independent variable τ_{ij}, the formulae in Appendix D still hold with τ_{ij} replacing j.

Table 8 through Table 10 and Figure 3 show the results for this model similar to Table 2 through Table 4 and Figure 1.

Appendix D shows the derivatives for this formulation of Wright's model.

13628-03_Hayne-3rdPgs.indd 136 9/12/14 9:48 AM

Table 8. Wright model parameter estimates

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}
Parameter	6.3169	6.4758	6.4403	6.5919	6.6407	6.7428	6.7468	6.7756	6.4808	6.4732	0.1864
Std. Error	0.1674	0.1665	0.1666	0.1662	0.1668	0.1670	0.1660	0.1634	0.1655	0.1836	0.1825

	θ_{12}	θ_{13}	K	p	AIC
Parameter	–0.078	0.2975	14.583	0.3199	612.33
Std. Error	0.0152	0.2322	0.9101	0.0746	

Table 9. Wright model estimates

Accident Year	Months of Development									Total
	24	36	48	60	72	84	96	108	120	
				Forecast Expected Average Losses						
2001										
2002									3.55	3.55
2003								12.03	3.43	15.46
2004							41.95	14.00	3.99	59.94
2005						112.49	44.05	14.70	4.19	175.43
2006					270.78	124.59	48.79	16.28	4.64	465.08
2007				501.71	271.86	125.09	48.98	16.35	4.65	968.64
2008			806.18	516.40	279.82	128.75	50.42	16.83	4.79	1,803.18
2009		787.26	600.33	384.54	208.37	95.87	37.54	12.53	3.57	2,130.00
2010	847.16	781.34	595.81	381.64	206.80	95.15	37.26	12.44	3.54	2,961.15
				Forecast Standard Deviation						
2001										
2002									11.20	11.20
2003								15.91	10.65	19.15
2004							23.60	16.61	11.11	30.93
2005						32.66	24.19	17.03	11.40	45.52
2006					43.93	34.27	25.39	17.87	11.96	64.89
2007				51.39	42.24	32.95	24.41	17.19	11.50	80.83
2008			60.85	52.77	43.37	33.84	25.07	17.65	11.81	102.92
2009		59.43	54.50	47.26	38.85	30.31	22.45	15.81	10.58	109.68
2010	57.03	55.57	50.96	44.19	36.32	28.34	20.99	14.78	9.89	117.34

13628-03_Hayne-3rdPgs.indd 137 9/12/14 9:48 AM

Table 10. Future payment estimates for Wright model

	Total Estimated Accident Year Unpaid Loss Estimates					
	Process Only		Including Parameter Uncertainty			
					Percentile	
Accident Year	Mean	Standard Deviation	Mean	Standard Deviation	5%	95%
	Total Estimated Accident Year Unpaid Loss Estimates					
2001	0	0	0	0	0	0
2002	137,270	432,966	146,803	500,614	–651,264	952,947
2003	646,137	800,325	678,423	905,856	–784,462	2,165,711
2004	2,533,412	1,306,997	2,600,209	1,478,254	219,628	5,027,027
2005	7,277,123	1,888,006	7,393,861	2,222,554	3,797,995	11,100,278
2006	18,702,982	2,609,470	18,835,439	3,162,250	13,713,787	24,078,401
2007	42,231,067	3,524,225	42,372,657	4,389,321	35,280,203	49,702,153
2008	75,946,730	4,334,693	76,107,096	5,819,740	66,651,205	85,842,577
2009	92,611,271	4,768,645	92,884,369	7,316,102	81,036,320	105,112,191
2010	146,554,330	5,807,424	147,221,999	13,359,651	125,978,317	169,973,663
Total	386,640,322	10,029,257	388,240,855	20,375,406	355,694,226	422,510,275
	Estimates of Loss Payments in Next Calendar Year					
2001	0	0	0	0	0	0
2002	137,270	432,966	146,803	500,614	–651,264	952,947
2003	502,940	665,167	519,746	733,651	–674,295	1,725,485
2004	1,773,153	997,326	1,792,630	1,079,363	29,163	3,557,836
2005	4,666,383	1,354,560	4,706,444	1,457,674	2,307,722	7,115,227
2006	10,889,367	1,766,422	10,920,602	1,890,455	7,829,137	14,061,135
2007	21,873,673	2,240,328	21,906,284	2,454,013	17,890,404	25,974,136
2008	33,955,023	2,562,715	33,989,988	3,006,619	29,069,635	39,023,999
2009	34,229,549	2,584,086	34,316,900	3,252,448	29,052,654	39,729,019
2010	41,928,126	2,822,425	42,069,559	4,461,885	34,945,491	49,564,857
Total	149,955,483	5,727,985	150,368,956	7,586,869	138,022,721	162,924,093

4.4. Generalized Hoerl curve model

Just as the Berquist-Sherman model presented here replaced the Cape Cod's accident year level factors with a uniform annual trend, we can further refine Wright's model and replace separate accident year levels by an expected trended amount. We will use the four-parameter curve in (4.4) and incorporate trend as well in the model that we call a generalized Hoerl curve model shown in (4.6).

$$g_{ij}(\boldsymbol{\theta}) = \exp(\theta_1 + \theta_2 j + \theta_3 j^2 + \theta_4 \ln(j) + i\theta_5), \quad i = 1, \ldots, m,\ j = 1, \ldots, n \qquad (4.6)$$

We note that, as with the last model, this requires that the expected losses be positive in all cells. Again, if we wished to replace the parameter j in (4.6) by a more general operational time independent variable τ_{ij}, the formulae in Appendix E still hold with τ_{ij} replacing j.

Table 11 through Table 13 and Figure 4 show the results for this model similar to Table 2 through Table 4 and Figure 1.

Appendix E shows the derivatives for this formulation of the generalized Hoerl curve model.

4.5. Chain ladder model

The fifth model we will consider here is the chain ladder. In contrast to England and Verrall (1999), our purpose is to state the classical *method* in a general

13628-03_Hayne-3rdPgs.indd 138 9/12/14 9:48 AM

Figure 3. Wright model charts

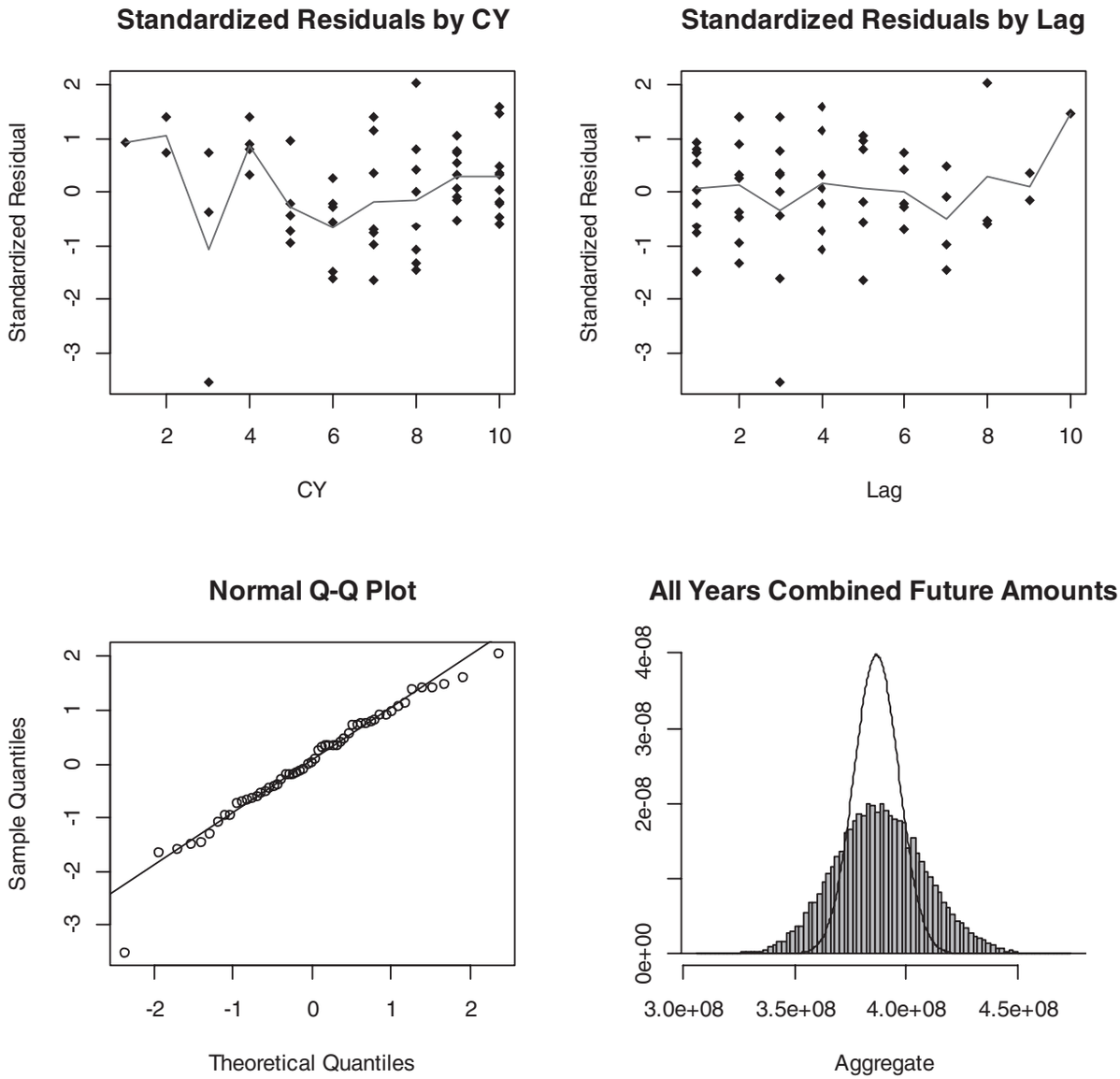


Table 11. Generalized Hoerl curve model parameter estimates

	θ_1	θ_2	θ_3	θ_4	θ_5	κ	p	AIC
Parameter	6.4977	0.0034	–0.065	0.5984	0.0430	13.142	0.5059	639.71
Std. Error	0.2195	0.2395	0.0185	0.3229	0.0084	1.0148	0.0826	

13628-03_Hayne-3rdPgs.indd 139 9/12/14 9:48 AM

Table 12. Generalized Hoerl curve model estimates

Accident Year	Months of Development 24	36	48	60	72	84	96	108	120	Total
					Forecast Expected Average Losses					
2001										
2002									4.39	4.39
2003								14.80	4.59	19.38
2004							43.43	15.45	4.79	63.66
2005						110.84	45.33	16.12	5.00	177.29
2006					245.28	115.70	47.32	16.83	5.22	430.35
2007				468.52	256.05	120.78	49.40	17.57	5.45	917.77
2008			766.66	489.10	267.29	126.09	51.57	18.34	5.68	1,724.73
2009		1,059.53	800.32	510.57	279.03	131.62	53.83	19.15	5.93	2,860.00
2010	1,197.91	1,106.05	835.47	532.99	291.28	137.40	56.20	19.99	6.19	4,183.49
					Forecast Standard Deviation					
2001										
2002									7.68	7.68
2003								13.65	7.55	15.60
2004							23.41	13.88	7.67	28.28
2005						37.97	24.15	14.32	7.91	47.88
2006					57.63	39.41	25.07	14.86	8.21	76.10
2007				76.80	56.57	38.68	24.61	14.58	8.06	107.13
2008			100.24	79.85	58.82	40.22	25.58	15.16	8.38	149.86
2009		116.21	100.83	80.32	59.16	40.45	25.73	15.25	8.43	190.33
2010	115.90	111.31	96.58	76.94	56.67	38.75	24.65	14.61	8.08	216.03

stochastic framework and we are not interested in reproducing the selection of classic weighted average development factors. We note that the chain ladder method is equivalent to assuming the incremental amounts at a particular age are equal to some factor times the aggregate amount for the accident year. So far, this is precisely the formulation of the Cape Cod model above. That model has $m + n - 1$ parameters. What sets the chain ladder apart, though, is the additional constraint that the expected amounts to date equal the actual amounts to date for each accident year. In effect, this constraint fixes m of the parameters that reflect accident year levels, reducing the remaining model to $n - 1$ parameters. The implication of this formulation is that the risk that the actual and expected amounts to date differ is *not* captured in process or parameter uncertainty but rather *remains in the realm of model uncertainty.*

Thus, we formulate the chain ladder model here using (4.7). Here we denote the actual average amount paid to date per unit of exposure for accident year i by P_i and denote by n_i the age of the most mature available entry for that accident year. For example, in a complete "square" triangle with annual development for m accident years through m years of development we have $n_i = m - i + 1$.

$$g_{ij}(\boldsymbol{\theta}) = \begin{cases} P_1\theta_j \text{ if } j < n \text{ and } i = 1 \\ P_1\left(1 - \sum_{k=1}^{n-1}\theta_k\right) \text{ if } j = n \text{ and } i = 1 \\ \dfrac{P_i\theta_j}{\sum_{k=1}^{n_i}\theta_k} \text{ if } j < n \text{ and } i \neq 1 \\ \dfrac{P_i}{\sum_{k=1}^{n_i}\theta_k}\left(1 - \sum_{k=1}^{n-1}\theta_k\right) \text{ if } j = n \text{ and } i \neq 1 \end{cases} \tag{4.7}$$

13628-03_Hayne-3rdPgs.indd 140 9/12/14 9:48 AM

Table 13. Future payment estimates for the generalized Hoerl curve model

Total Estimated Accident Year Unpaid Loss Estimates						
	Process Only		Including Parameter Uncertainty			
					Percentile	
Accident Year	Mean	Standard Deviation	Mean	Standard Deviation	5%	95%
Total Estimated Accident Year Unpaid Loss Estimates						
2001	0	0	0	0	0	0
2002	169,866	296,971	188,570	355,767	–379,492	767,244
2003	810,146	652,176	857,782	775,477	–377,190	2,141,572
2004	2,690,392	1,195,121	2,763,349	1,414,141	478,798	5,114,808
2005	7,354,087	1,986,075	7,464,912	2,340,903	3,715,734	11,364,279
2006	17,306,357	3,060,366	17,411,757	3,539,256	11,635,512	23,236,441
2007	40,013,505	4,670,729	40,235,193	5,423,020	31,473,809	49,279,771
2008	72,642,848	6,311,652	72,860,486	7,488,778	60,699,576	85,272,138
2009	124,351,005	8,275,308	124,733,603	10,496,421	107,725,640	142,061,952
2010	207,051,137	10,691,966	207,505,148	15,415,859	182,654,820	233,213,854
Total	472,389,343	16,115,325	473,722,319	29,454,831	426,676,462	523,060,721
Estimates of Loss Payments in Next Calendar Year						
2001	0	0	0	0	0	0
2002	169,866	296,971	188,570	355,767	–379,492	767,244
2003	618,476	570,763	645,372	650,329	–398,957	1,719,855
2004	1,835,320	989,529	1,859,707	1,080,411	96,052	3,616,012
2005	4,597,553	1,574,902	4,631,492	1,686,375	1,869,355	7,415,311
2006	9,863,747	2,317,720	9,862,558	2,426,617	5,911,497	13,887,916
2007	20,426,916	3,348,200	20,501,981	3,556,848	14,739,516	26,404,805
2008	32,290,474	4,221,997	32,326,281	4,541,447	24,882,439	39,820,696
2009	46,067,803	5,052,593	46,113,592	5,530,294	37,072,543	55,181,750
2010	59,287,652	5,736,070	59,368,323	6,572,320	48,655,391	70,225,964
Total	175,157,807	9,834,234	175,497,877	12,385,515	155,435,156	196,021,497

13628-03_Hayne-3rdPgs.indd 141 9/12/14 9:48 AM

Figure 4. Generalized Hoerl curve model charts

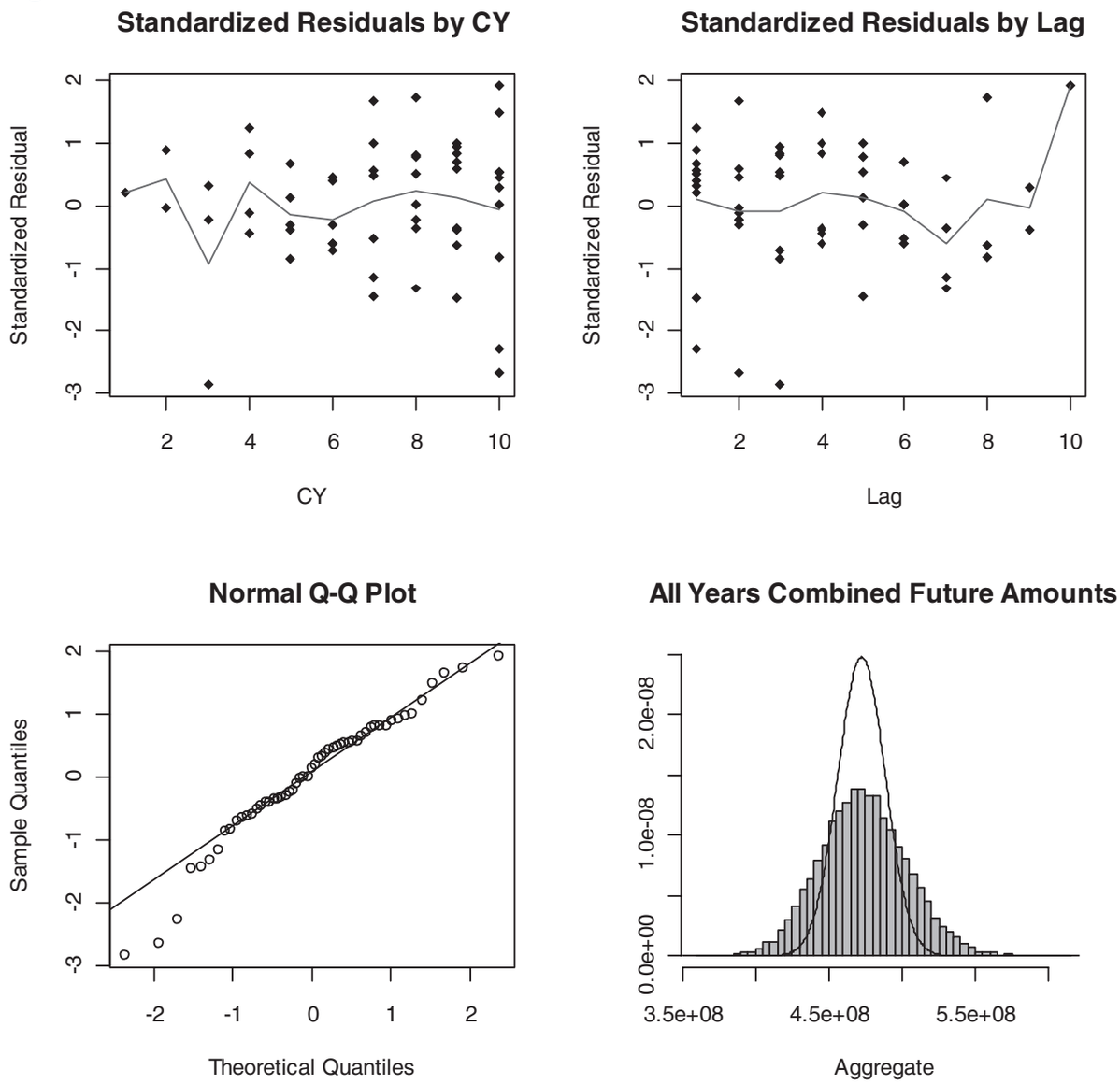


Table 14 through Table 16 and Figure 5 show the results for this model, similar to Table 2 through Table 4 and Figure 1.

Appendix F shows the derivatives for this formulation of the chain ladder model.

5. Some other MLE applications

This is not the only paper to discuss the use of MLEs in the context of actuarial projections. Weisner (1978) shows the use of MLEs in fitting probability distributions to the emergence of reported claims over time, similar to the Hoerl curve model presented here, but limited to a single accident year. Clark (2003) also used parametric curves to model emergence of losses but in the context of loss triangles. In the models presented here, the Berquist-Sherman model can be seen as a special case of the Cape Cod where accident year variations are replaced by trend. Wright's model can be seen as another special case of the Cape Cod where the development year variations are replaced by a smooth curve. The Hoerl curve model can be

13628-03_Hayne-3rdPgs.indd 142 9/12/14 9:48 AM

Table 14. Chain ladder model parameter estimates

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	κ	p	AIC
Parameter	0.1955	0.2307	0.2077	0.1637	0.1043	0.0555	0.0217	0.0132	0.0030	13.074	0.4378	599.37
Std. Error	0.0049	0.0052	0.0052	0.0051	0.0047	0.0040	0.0031	0.0030	0.0018	1.0074	0.0824	

seen as a special case of both of these, with both development variations and accident year variations replaced by parametric curves. Wright's model most closely tracks the approach taken in Clark (2003).

Instead of a Hoerl curve, Clark considers the log-logistic curve and the Weibull curve as cumulative emergence patterns. Clark also presents another model, labeled as "Cape Cod," which assumes constant expected accident year loss ratios. In terms of the models we have here, this is quite similar to the Hoerl curve model, with the assumption of no trend and using premiums as an exposure measure in the denominator.

Instead of allowing the variance of incremental amounts to be proportional to a power of the square of their mean, Clark requires them to be proportional to the mean, resulting in the selection of the over-dispersed Poisson as an underlying statistical model. Given the general framework presented in this paper, it would not be difficult to include additional example models here. All that would be necessary would be to appropriately define the various functions corresponding to (3.2). In the implementation we would need to derive the Hessian of those expected value functions and to incorporate them in the R code in Appendix G.

Table 15. Chain ladder model estimates

					Forecast Expected Average Losses					
2001										
2002									17.39	17.39
2003								10.71	16.89	27.59
2004							55.46	12.69	20.00	88.15
2005						94.70	57.55	13.16	20.76	186.17
2006					267.91	104.51	63.51	14.53	22.91	473.36
2007				507.58	270.22	105.41	64.05	14.65	23.10	985.01
2008			821.57	523.52	278.70	108.72	66.07	15.11	23.83	1,837.52
2009		772.46	608.70	387.88	206.49	80.55	48.95	11.20	17.66	2,133.88
2010	853.88	768.69	605.74	385.99	205.48	80.16	48.71	11.14	17.57	2,977.36
					Forecast Standard Deviation					
2001										
2002									12.25	12.25
2003								9.53	11.63	15.04
2004							19.47	10.21	12.46	25.27
2005						24.84	19.98	10.47	12.78	35.91
2006					39.78	26.35	21.18	11.11	13.56	55.07
2007				50.54	38.35	25.40	20.42	10.71	13.07	73.29
2008			63.48	52.12	39.55	26.19	21.06	11.04	13.48	98.71
2009		60.82	54.79	44.98	34.14	22.61	18.18	9.53	11.63	104.68
2010	59.56	56.88	51.25	42.07	31.93	21.14	17.00	8.91	10.88	114.60

13628-03_Hayne-3rdPgs.indd 143 9/12/14 9:48 AM

Table 16. Future payment estimates for the chain ladder model

	Total Estimated Accident Year Unpaid Loss Estimates					
	Process Only		Including Parameter Uncertainty			
					Percentile	
Accident Year	Mean	Standard Deviation	Mean	Standard Deviation	5%	95%
			Total Estimated Accident Year Unpaid Loss Estimates			
2001	0	0	0	0	0	0
2002	672,556	473,869	671,147	689,049	–319,900	1,885,721
2003	1,153,495	628,724	1,150,636	885,614	–217,651	2,666,293
2004	3,725,552	1,068,159	3,721,634	1,426,223	1,416,270	6,097,255
2005	7,722,556	1,489,549	7,732,341	1,901,538	4,646,017	10,908,311
2006	19,036,072	2,214,503	19,027,321	2,701,443	14,571,405	23,467,318
2007	42,945,172	3,195,515	42,949,561	3,840,408	36,740,390	49,297,398
2008	77,393,393	4,157,471	77,391,906	4,945,416	69,245,975	85,536,306
2009	92,779,952	4,551,418	92,773,213	5,324,172	84,080,881	101,546,274
2010	147,356,871	5,671,774	147,474,496	7,340,340	135,630,736	159,650,144
Total	392,785,618	9,447,957	392,892,256	15,703,578	367,309,051	418,819,212
			Estimates of Loss Payments in Next Calendar Year			
2001	0	0	0	0	0	0
2002	672,556	473,869	671,147	689,049	–319,900	1,885,721
2003	447,637	398,443	446,263	502,414	–301,293	1,310,455
2004	2,343,910	823,025	2,339,002	1,004,344	759,164	4,030,521
2005	3,928,277	1,030,573	3,930,214	1,201,908	2,003,801	5,954,578
2006	10,773,902	1,599,744	10,760,299	1,822,639	7,789,085	13,773,068
2007	22,129,708	2,203,317	22,125,964	2,479,555	18,086,585	26,225,425
2008	34,603,222	2,673,798	34,598,460	2,968,528	29,750,109	39,518,563
2009	33,585,957	2,644,331	33,596,235	2,894,785	28,889,389	38,392,903
2010	42,260,699	2,947,786	42,311,316	3,318,030	36,884,904	47,842,602
Total	150,745,869	5,689,259	150,778,901	6,405,816	140,279,071	161,360,024

6. Observations and conclusions

As with traditional analyses, the various stochastic models considered here give different estimates of expected total future payments, ranging from $388 million for Wright's model to $480 million for the Berquist-Sherman model, and AIC values ranging from 643 for the Berquist-Sherman model to 599 for the chain ladder model. Comparing the total indicated distributions of projected future payments, as shown in Figure 6, provides valuable information.

Figure 6 shows that not only are the expected amounts different among the various models but so are the underlying distributions of forecast outcomes. Under the Hoerl and Berquist-Sherman models, there is little likelihood for the bulk of the potential forecasts from the Cape Cod, Wright, and chain ladder models and the converse is also true. In short, differences in the forecasts are likely not due to random fluctuations or even parameter uncertainty but to differences among the models themselves. This is model uncertainty. Although the chain ladder model shows the tightest distribution and lowest AIC value, care should be taken in jumping to conclusions. As mentioned before, the formulation of the chain ladder has the requirement that the expected amount

13628-03_Hayne-3rdPgs.indd 144 9/12/14 9:48 AM

Figure 5. Chain ladder model charts

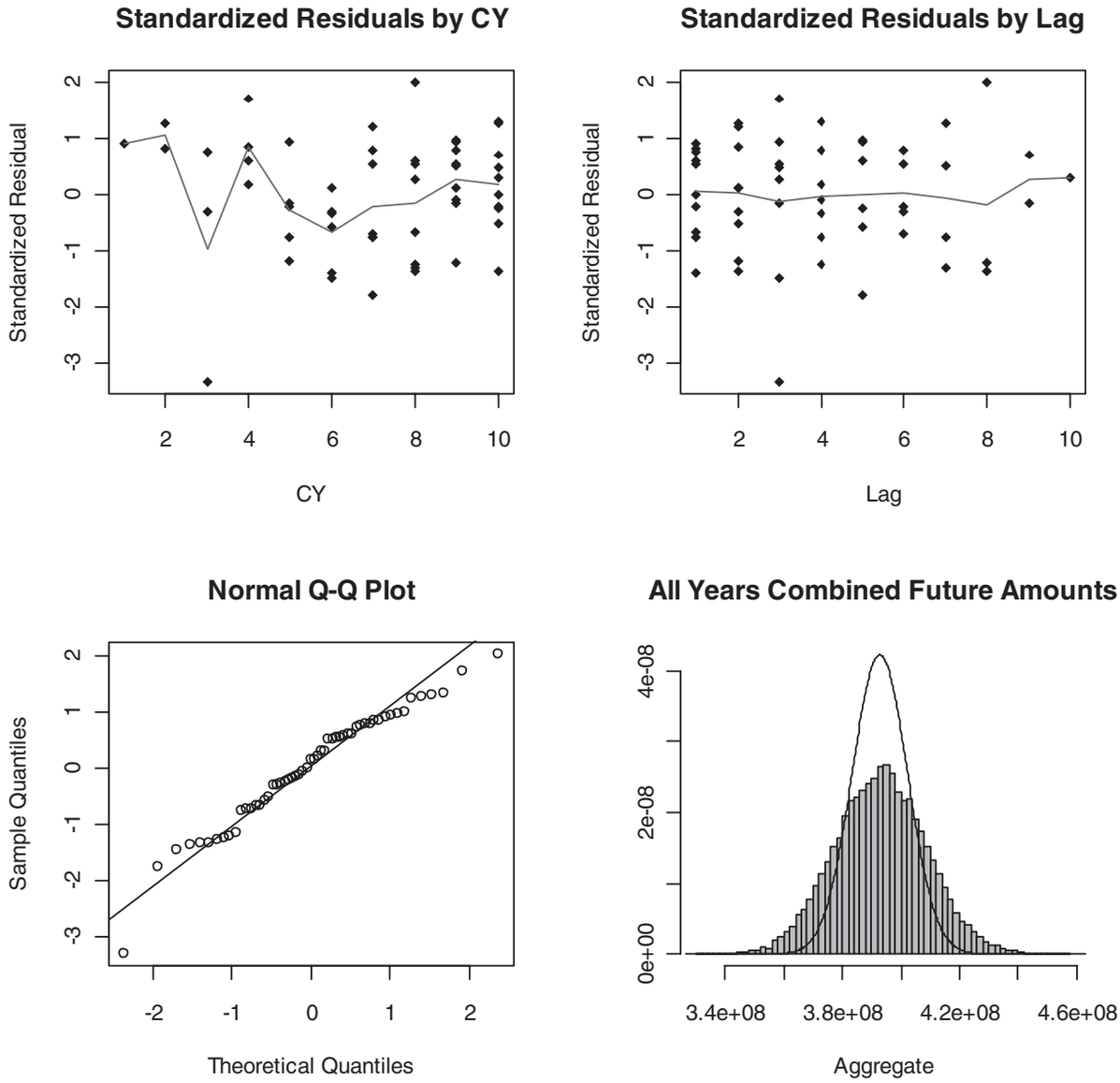
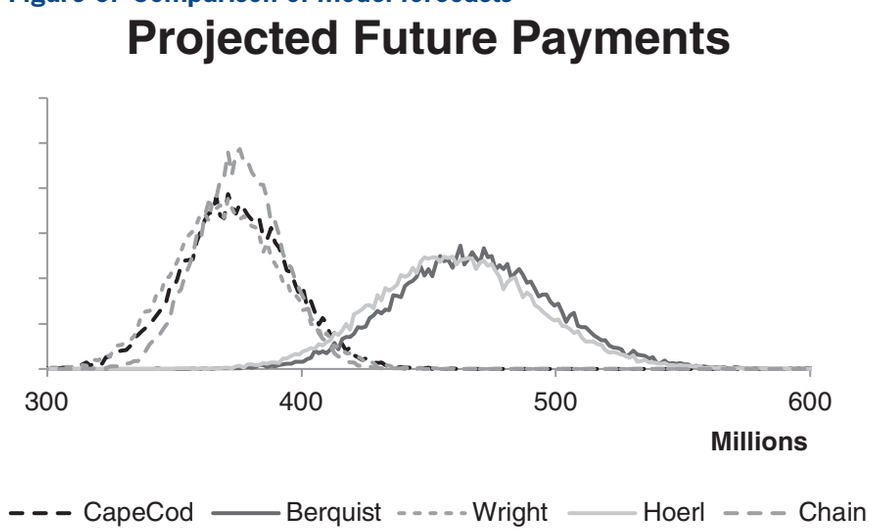


Figure 6. Comparison of model forecasts



13628-03_Hayne-3rdPgs.indd 145 9/12/14 9:48 AM

to date equals the actual amount for each accident year, a restriction that the other three models presented do not have.

Reviewing the data in Table 1 gives some insight as to the reason for the differences among the models. We see average costs steadily increasing through accident year 2008 and then dropping off in the most recent two accident years. Both the Berquist-Sherman incremental severity and the generalized Hoerl curve models shown here assume a uniform trend throughout. The drop-off in the last two accident years increases the errors for these two models and hence the standard deviation of the forecasts. The Cape Cod, Wright, and chain ladder models all key on the amounts to date for the more recent accident years and thus their forecasts react to the change in the pattern. The fundamental question then becomes whether this drop-off is due to some characteristic of the underlying data or random noise. If the former is the case, then the uniform trend assumption of the Berquist and Hoerl models would be violated. If the latter is the case, then the Cape Cod and chain ladder models will be fitting to noise and not signal. The data alone as presented does not answer this question.

In spite of the additional information provided by stochastic models, the practitioner is faced with the same question that the practitioner of traditional methods faced: the applicability of a particular model or method to a particular data set. Unfortunately, a single data set probably is not sufficient to answer the question in either situation. This is a significant drawback to this or any frequentist statistical approach.

This entire discussion goes to probably the greatest source of uncertainty in reserving projections—the question of model uncertainty. We hope the greater ease that the approach outlined in this paper gives to considering a range of models in reserving exercises will make it easier to explore and quantify the issue of model uncertainty in the future.

These models applied to a single data set may not give insight into model uncertainty. However, successive applications of these models over time might give additional insight. One significant advantage these stochastic models have over their nonstochastic counterparts is the ability to assess new data. Table 4, Table 7, Table 13, and Table 16 show not only estimates and statistics on total future payments by accident year, but also estimates and statistics on forecast payments during the next year. If emerging losses deviate significantly, say outside the 90% confidence interval shown, then one might come to suspect that a particular model may not be capturing what is actually going on in the underlying data. This observation also opens the way toward a potential weighting of the various models as time progresses. If we assess that none of these models are better than any of the others in deriving an ultimate loss forecast, then a Bayesian outlook would say that the model that is a straight average of the five might in some sense be better than any of the models alone. As losses emerge next year, this same Bayesian approach would indicate how to modify those a priori weights to reflect how well or how poorly a particular model performed in forecasting the future. If one applied such a Bayesian evolution to a broad stable of models, possibly incorporating some not shown here, it is likely that the resulting blend could produce forecasts most consistent with the underlying data. Models that have a good track record of predicting the next diagonal receive increasing weight while those with poor records receive decreasing weight.

There is still much work to be done. At least until the loss-generating process in insurance is understood and can be adequately and accurately modeled, there continues to be the need to look at a range of potential loss emergence models, not just the results from a single one. There also continues to be the need to find a way to sort through all those various models to best track and predict future payments or claims. Frequentist approaches such as those presented here by their nature are limited to observed data. They can, however, form the basis for a richer analysis, incorporating not only data specific to a particular situation but also experience from similar situations using Bayesian methods.

13628-03_Hayne-3rdPgs.indd 146 9/12/14 9:48 AM

Appendix A. Derivatives of the likelihood function

In order to derive estimates of parameter uncertainty, we need the matrix of second derivatives of the negative log likelihood function. In this appendix we list those derivatives.

Recall from (3.8) the negative log likelihood function is given by

$$\ell(A_{11}, A_{12}, \ldots, A_{m1}; \boldsymbol{\theta}, \kappa, p)$$
$$= \frac{1}{2}\sum_{(i,j)\in S} \kappa - w_i + \ln\left(2\pi\left(g_{ij}(\boldsymbol{\theta})^2\right)^p\right)$$
$$+ \frac{\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}.$$

Suppressing arguments and parameters, we thus have the following first partial derivatives:

$$\frac{\partial \ell}{\partial \theta_r} = \sum_{(i,j)\in S} \frac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})\left(\frac{p}{g_{ij}(\boldsymbol{\theta})} - \frac{\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p} - \frac{p\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p g_{ij}(\boldsymbol{\theta})}\right)$$

$$\frac{\partial \ell}{\partial \kappa} = \sum_{(i,j)\in S}\left(\frac{1}{2} - \frac{\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2}{2e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}\right)$$

$$\frac{\partial \ell}{\partial p} = \sum_{(i,j)\in S} \frac{\ln\left(g_{ij}(\boldsymbol{\theta})^2\right)}{2}\left(1 - \frac{\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}\right)$$

These then give the following second derivatives:

$$\frac{\partial^2 \ell}{\partial \theta_r \partial \theta_s} = \sum_{(i,j)\in S}\left(\begin{array}{l} \left(\dfrac{p\dfrac{\partial^2 g_{ij}}{\partial \theta_r \partial \theta_s}(\boldsymbol{\theta})}{g_{ij}(\boldsymbol{\theta})} + \dfrac{\dfrac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})\dfrac{\partial g_{ij}}{\partial \theta_s}(\boldsymbol{\theta})}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p} - \dfrac{p\dfrac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})\dfrac{\partial g_{ij}}{\partial \theta_s}(\boldsymbol{\theta})}{g_{ij}(\boldsymbol{\theta})^2}\right) \\ + \left(\dfrac{4p\dfrac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})\dfrac{\partial g_{ij}}{\partial \theta_s}(\boldsymbol{\theta})}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p g_{ij}(\boldsymbol{\theta})} - \dfrac{\dfrac{\partial^2 g_{ij}}{\partial \theta_r \partial \theta_s}(\boldsymbol{\theta})}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}\right)\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right) \\ + \left(\dfrac{p(2p+1)\dfrac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})\dfrac{\partial g_{ij}}{\partial \theta_s}(\boldsymbol{\theta})}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^{p+1}} - \dfrac{p\dfrac{\partial^2 g_{ij}}{\partial \theta_r \partial \theta_s}(\boldsymbol{\theta})}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p g_{ij}(\boldsymbol{\theta})}\right)\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2 \end{array}\right)$$

$$\frac{\partial^2 \ell}{\partial \kappa \partial \theta_r} = \sum_{(i,j)\in S}\left(\frac{\dfrac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right) + \frac{p\dfrac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p g_{ij}(\boldsymbol{\theta})}\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2\right)$$

$$\frac{\partial^2 \ell}{\partial p \partial \theta_r} = \sum_{(i,j)\in S}\left(\frac{\dfrac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})}{g_{ij}(\boldsymbol{\theta})} + \frac{\ln\left(g_{ij}(\boldsymbol{\theta})^2\right)\dfrac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right) + \frac{\left(p\ln\left(g_{ij}(\boldsymbol{\theta})^2\right) - 1\right)\dfrac{\partial g}{\partial \theta_r}(\boldsymbol{\theta})}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p g_{ij}(\boldsymbol{\theta})}\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2\right)$$

13628-03_Hayne-3rdPgs.indd 147 9/12/14 9:48 AM

$$\frac{\partial^2 \ell}{\partial \kappa^2} = \sum_{(i,j)\in S} \frac{\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2}{2e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}$$

$$\frac{\partial^2 \ell}{\partial p \partial \kappa} = \sum_{(i,j)\in S} \frac{\ln\left(g_{ij}(\boldsymbol{\theta})^2\right)\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2}{2e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}$$

$$\frac{\partial^2 \ell}{\partial p^2} = \sum_{(i,j)\in S} \frac{\ln\left(g_{ij}(\boldsymbol{\theta})^2\right)^2\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2}{2e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p}$$

The information matrix then requires the expected values of these derivatives. To this end, recall that because of (3.7) we have the following relationships:

$$\mathrm{E}\left(A_{ij}\right) = g_{ij}(\boldsymbol{\theta}), \text{ and}$$

$$\mathrm{Var}\left(A_{ij}\right) = \mathrm{E}\left(\left(A_{ij} - g_{ij}(\boldsymbol{\theta})\right)^2\right) = e^{\kappa - w_i} g_{ij}(\boldsymbol{\theta})^{2p}.$$

We can then derive the entries of the information matrix as follows:

$$\mathrm{E}\left(\frac{\partial^2 \ell}{\partial \theta_r \partial \theta_s}\right) = \mathrm{E}\left(\frac{\partial^2 \ell}{\partial \theta_s \partial \theta_r}\right)$$

$$= \sum_{(i,j)\in S} \frac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta}) \frac{\partial g_{ij}}{\partial \theta_s}(\boldsymbol{\theta}) \left(\frac{1}{e^{\kappa - w_i}\left(g_{ij}(\boldsymbol{\theta})^2\right)^p} + \frac{2p^2}{g_{ij}(\boldsymbol{\theta})^2}\right)$$

$$\mathrm{E}\left(\frac{\partial^2 \ell}{\partial \kappa \partial \theta_r}\right) = \mathrm{E}\left(\frac{\partial^2 \ell}{\partial \theta_r \partial \kappa}\right) = \sum_{(i,j)\in S} \frac{p \frac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})}{g_{ij}(\boldsymbol{\theta})}$$

$$\mathrm{E}\left(\frac{\partial^2 \ell}{\partial \theta_r \partial p}\right) = \mathrm{E}\left(\frac{\partial^2 \ell}{\partial p \partial \theta_r}\right) = \sum_{(i,j)\in S} \frac{\frac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})\, p \ln\left(g_{ij}(\boldsymbol{\theta})^2\right)}{g_{ij}(\boldsymbol{\theta})}$$

$$\mathrm{E}\left(\frac{\partial^2 \ell}{\partial \kappa^2}\right) = \sum_{(i,j)\in S} \frac{1}{2}$$

$$\mathrm{E}\left(\frac{\partial^2 \ell}{\partial \kappa \partial p}\right) = \mathrm{E}\left(\frac{\partial^2 \ell}{\partial p \partial \kappa}\right) = \sum_{(i,j)\in S} \frac{\ln\left(g_{ij}(\boldsymbol{\theta})^2\right)}{2}$$

$$\mathrm{E}\left(\frac{\partial^2 \ell}{\partial p^2}\right) = \sum_{(i,j)\in S} \frac{\ln\left(g_{ij}(\boldsymbol{\theta})^2\right)^2}{2}$$

Appendix B. Cape Cod model

A variation of the Stanard-Buhlmann Cape Cod model can be stated as the following set of functions.

$$g_{ij}(\boldsymbol{\theta}) = \begin{cases} \theta_1 \text{ if } i = j = 1 \\ \theta_1\theta_i \text{ if } j = 1 \text{ and } i > 1 \\ \theta_1\theta_{m+j-1} \text{ if } i = 1 \text{ and } j > 1 \\ \theta_1\theta_i\theta_{m+j-1} \text{ if } i > 1 \text{ and } j > 1 \end{cases}$$

The various first and second partial derivatives of this function are as follows.

$\nabla g_{11}(\boldsymbol{\theta}) = (1, 0, \ldots, 0)$

$\nabla g_{i1}(\boldsymbol{\theta}) = (\theta_i, 0, \ldots, 0, \theta_1, 0, \ldots, 0)$, non zero in first and *ith* position, $i > 1$

$\nabla g_{1j}(\boldsymbol{\theta}) = (\theta_{m+j-1}, 0, \ldots, 0, \theta_1, 0, \ldots, 0)$, non zero in first and $m + j - 1st$ position, $j > 1$

$\nabla g_{ij}(\boldsymbol{\theta}) = (\theta_i\theta_{m+j-i}, 0, \ldots, 0, \theta_1\theta_{m+j-1}, 0, \ldots, 0, \theta_i\theta_1, 0, \ldots, 0)$, non zero in first, *ith* and $m + j - 1st$ position, $i > 1, j > 1$

$$\text{Hessian}\left(g_{11}(\boldsymbol{\theta})\right) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$\text{Hessian}\left(g_{i1}(\boldsymbol{\theta})\right) = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

only non zero in *ith* position, $i > 1$

13628-03_Hayne-3rdPgs.indd 148 9/12/14 9:48 AM

$$\text{Hessian}\left(g_{1j}(\boldsymbol{\theta})\right)=\begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

only non zero in $m+j-1st$ position, $\mathrm{j}>1$

$$\text{Hessian}\left(g_{ij}(\boldsymbol{\theta})\right)=$$

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & \theta_{m+j-1} & 0 & \cdots & 0 & \theta_i & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \theta_{m+j-1} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \theta_1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \theta_i & 0 & \cdots & 0 & \theta_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

only non zero in ith and $m+j-1st$ position, $i>1, j>1$

Appendix C. Berquist-Sherman incremental average cost model

A simplified model from Berquist and Sherman (1977) can be represented using the following set of functions.

$$g_{ij}(\boldsymbol{\theta})=\theta_j e^{i\theta_{n+1}},\ i=1,2,\ldots,m,\ j=1,2,\ldots,n$$

The various first and second partial derivatives of this function are as follows.

$$\frac{\partial g_{ij}}{\partial \theta_r}(\boldsymbol{\theta})=\begin{cases} 0\ if\ \ j\neq r, r\leq n \\ e^{i\theta_{n+1}}\ \ if\ \ j=r, r\leq n \\ i\theta_j e^{i\theta_{n+1}}\ \ if\ \ r=n+1 \end{cases}$$

$$\text{Hessian}\left(g_{ij}(\boldsymbol{\theta})\right)=$$

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & i\exp(i\theta_{n+1}) \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & i\exp(i\theta_{n+1}) & 0 & \cdots & 0 & i^2\theta_j\exp(i\theta_{n+1}) \end{pmatrix}$$

Appendix D. Wright's model

A generalization of two of the curves used by Wright (1992). can be described using the following set of functions.

$$g_{ij}(\boldsymbol{\theta})=\exp\left(\theta_i+\theta_{m+1}j+\theta_{m+2}j^2+\theta_{m+3}\ln(j)\right),$$
$$i=1,\ldots,m, j=1,\ldots,n$$

The various first and second partial derivatives of this function are as follows.

$$\nabla g_{ij}(\theta)=\left(0,\ldots,0,\underset{i}{\underset{\smile}{1}},0,\ldots,0,\ j,\ j^2,\ln(j)\right)g_{ij}(\theta)$$

$$\text{Hessian}\left(g_{ij}(\theta)\right)=$$

$$\begin{pmatrix} 0\ldots0 & 0 & 0\ldots0 & 0 & 0 & 0 \\ \vdots\ddots\vdots & \vdots & \vdots\ddots\vdots & \vdots & \vdots & \vdots \\ 0\ldots0 & 0 & 0\ldots0 & 0 & 0 & 0 \\ 0\ldots0 & 1 & 0\ldots0 & j & j^2 & \ln(j) \\ 0\ldots0 & 0 & 0\ldots0 & 0 & 0 & 0 \\ \vdots\ddots\vdots & \vdots & \vdots\ddots\vdots & \vdots & \vdots & \vdots \\ 0\ldots0 & 0 & 0\ldots0 & 0 & 0 & 0 \\ 0\ldots0 & j & 0\ldots0 & j^2 & j^3 & j\ln(j) \\ 0\ldots0 & j^2 & 0\ldots0 & j^3 & j^4 & j^2\ln(j) \\ 0\ldots0 & \ln(j) & 0\ldots0 & j\ln(j) & j^2\ln(j) & \ln(j)^2 \end{pmatrix} g_{ij}(\theta)$$

13628-03_Hayne-3rdPgs.indd 149 9/12/14 9:48 AM

Appendix E. Generalized Hoerl curve model

A generalization of two of the curves used by Wright (1992) can be described using the following set of functions, replacing separate levels by accident year by trend.

$$g_{ij}(\boldsymbol{\theta}) = \exp(\theta_1 + \theta_2 j + \theta_3 j^2 + \theta_4 \ln(j) + i\theta_5),$$

$$i = 1,\ldots,m,\ j = 1,\ldots,n$$

The various first and second partial derivatives of this function are as follows.

$$\nabla g_{ij}(\boldsymbol{\theta}) = (1, j, j^2, \ln(j), i)\, g_{ij}(\boldsymbol{\theta})$$

$$\text{Hessian}\left(g_{ij}(\boldsymbol{\theta})\right) = \begin{pmatrix} 1 & j & j^2 & \ln(j) & i \\ j & j^2 & j^3 & j\ln(j) & ij \\ j^2 & j^3 & j^4 & j^2\ln(j) & ij^2 \\ \ln(j) & j\ln(j) & j^2\ln(j) & \ln(j)^2 & i\ln(j) \\ i & ij & ij^2 & i\ln(j) & i^2 \end{pmatrix} g_{ij}(\boldsymbol{\theta})$$

Appendix F. Chain ladder model

The incremental averages from the chain ladder model can be described by the following functions.

$$g_{ij}(\boldsymbol{\theta}) = \begin{cases} P_1\theta_j \text{ if } j < n \text{ and } i = 1 \\ P_1\left(1 - \sum_{k=1}^{n-1}\theta_k\right) \text{ if } j = n \text{ and } i = 1 \\ \dfrac{P_i\theta_j}{\sum_{k=1}^{n_i}\theta_k} \text{ if } j < n \text{ and } i \neq 1 \\ \dfrac{P_i}{\sum_{k=1}^{n_i}\theta_k}\left(1 - \sum_{k=1}^{n-1}\theta_k\right) \text{ if } j = n \text{ and } i \neq 1 \end{cases}$$

The various first and second partial derivatives of this function are as follows.

$$\nabla g_{1j}(\boldsymbol{\theta}) = P_1 \begin{cases} \mathbf{V}_{1j} \text{ if } j < n \\ -\mathbf{V}_{21} \text{ if } j = n \end{cases}$$

$$\nabla g_{ij}(\boldsymbol{\theta}) = \frac{P_i}{\sum_{k=1}^{n_i}\theta_k} \begin{cases} \left(\mathbf{V}_{1j} - \dfrac{\theta_j}{\sum_{k=1}^{n_i}\theta_k}\mathbf{V}_{2i}\right) \text{ if } j \le n_i \text{ and } i > 1 \\ \mathbf{V}_{1j} \text{ if } n > j > n_i \text{ and } i > 1 \end{cases}$$

$$\nabla g_{in}(\boldsymbol{\theta}) = -\frac{P_i}{\sum_{k=1}^{n_i}\theta_k}\left(\frac{1}{\sum_{k=1}^{n_i}\theta_k}\left(1 - \sum_{k=n_i+1}^{n-1}\theta_k\right)\mathbf{V}_{2i} + \mathbf{V}_{3i}\right)$$

Where

$$\mathbf{V}_{1j} = \left(0,\ldots,0,\underset{j}{1},0,\ldots,0\right),\ \mathbf{V}_{2i} = \left(\underbrace{1,\ldots,1}_{n_i},0,\ldots,0\right)$$

$$\mathbf{V}_{3i} = \left(\underbrace{0,\ldots,0}_{n_i},1,\ldots 1\right) = (1,\ldots,1) - \mathbf{V}_{2i}$$

$$\text{Hessian}\left(g_{1j}(\boldsymbol{\theta})\right) = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

$$\text{Hessian}\left(g_{ij}(\boldsymbol{\theta})\right) = \frac{P_i}{\left(\sum_{k=1}^{n_i}\theta_k\right)^2} \begin{cases} \dfrac{2\theta_j}{\sum_{k=1}^{n_i}\theta_k}\mathbf{M}_{1i} - \left(\mathbf{M}_{2ij} + \mathbf{M}_{3j}\right) \text{ if } j \le n_i \\ -\mathbf{M}_{4ij} \text{ if } n > j > n_i \end{cases}$$

$$\text{Hessian}\left(g_{in}(\boldsymbol{\theta})\right) = \frac{P_i}{\left(\sum_{k=1}^{n_i}\theta_k\right)^2}\left(\frac{2}{\sum_{k=1}^{n_i}\theta_k}\left(1 - \sum_{k=n_i+1}^{n-1}\theta_k\right)\mathbf{M}_{1i} + \mathbf{M}_{5i}\right)$$

Where

$$\mathbf{M}_{1i} = \begin{pmatrix} 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1_{n_i n_i} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

13628-03_Hayne-3rdPgs.indd 150 9/12/14 9:48 AM

$$\mathbf{M}_{2ij} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 1 & \cdots & 1 & 1_{jj} & 1 & \cdots & 1_{jn_i} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1_{n_i j} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$\mathbf{M}_{3j} = \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1_{jj} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$\mathbf{M}_{4ij} = \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1_{n_i j} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 1_{jn_i} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$\mathbf{M}_{5i} = \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1_{n_i n_i+1} & \cdots & 1_{n_i n-1} \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1_{n_i+1 n_i} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1_{n-1 n_i} & 0 & \cdots & 0 \end{pmatrix}$$

Appendix G. R Code

The R code of Appendix G can be found on the *Variance* web site at http://www.variancejournal.org/issues/07-02/R-Example.txt

References

Actuarial Standards Board, *Actuarial Standard of Practice No. 43,* "Property/Casualty Unpaid Claim Estimates," June 2007, 3.6.1, p. 5.

Berquist, J. R., and R. E. Sherman, "Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach," *Proceedings of the Casualty Actuarial Society* 64, 1977, pp. 123–184.

Bornhuetter, R. L., and R. E. Ferguson, "The Actuary and IBNR," *Proceedings of the Casualty Actuarial Society* 59, 1972, pp. 181–195.

Clark, D. R., "LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach," Casualty Actuarial Society *Forum,* Fall 2003, pp. 41–92.

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

England, P. D., and R. J. Verrall, "A Flexible Framework For Stochastic Claims Reserving," *Proceedings of the Casualty Actuarial Society* 88, 2001, pp. 1–38.

Klugman, S. A., H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions* (1st ed.), New York: Wiley, 1998, pp. 55–67.

Meyers, G. G., "Stochastic Loss Reserving with the Collective Risk Model," *Variance* 3, 2009, pp. 239–269.

R Core Team, "R: A Language and Environment for Statistical Computing," Vienna, Austria: R Foundation for Statistical Computing, 2012, ISBN 3-900051-07-0, URL http://www.r-project.org. The MASS package that is used with the R Code: Venables, W. N., and B. D. Ripley, *Modern Applied Statistics with S* (4th ed.), New York: Springer, 2002, ISBN 0-387-95457-0.

Scollnik, D. P. M., "Introduction to Markov Chain Monte Carlo Methods and Their Actuarial Applications," *Proceedings of the Casualty Actuarial Society* 83, 1996, pp. 114–165.

Stanard, J. N., "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *Proceedings of the Casualty Actuarial Society* 72, 1985, pp. 124–148.

Venables, W. N., and B. D. Ripley, *Modern Applied Statistics with S* (4th ed.), New York: Springer, 2002, p. 174.

Venter, G., "Generalized Linear Models beyond the Exponential Family with Loss Reserve Applications," *CAS E-Forum,* Summer 2007.

Weisner, E. W., "Estimation of the Distribution of Report Lags by the Method of Maximum Likelihood," *Proceedings of the Casualty Actuarial Society* 65, 1978, pp. 1–9.

Wright, T. S., "Stochastic Claims Reserving When Past Claim Numbers are Known," *Proceedings of the Casualty Actuarial Society* 79, 1992, pp. 255–361.

13628-03_Hayne-3rdPgs.indd 151 9/12/14 9:48 AM