Hierarchical Growth Curve Models for Loss Reserving

James Guszcza, FCAS, MAAA

Abstract

Hierarchical or multilevel modeling extends traditional GLM or non-linear models by giving certain of the model parameters their own probability sub-models. Hierarchical modeling can be viewed as an extension of Bayesian credibility theory that allows one to build models for data that are grouped along a dimension containing multiple levels. In particular, hierarchical modeling can be used to analyze longitudinal datasets containing multiple observations for each of several subjects. A contention of this paper is that traditional loss reserving triangles are most naturally regarded as longitudinal datasets. Non-linear hierarchical models – known also as non-linear mixed effects models – therefore provide a natural and flexible framework in which to model loss development across multiple accident years. The use of non-linear growth curves together with multilevel modeling techniques allows one to build models that are at once parsimonious and easy to interpret. Finally, because they incorporate growth curves, such models obviate the need to specify tail factors.

Keywords: Stochastic loss reserving, hierarchical models, multilevel models, nonlinear mixed effects models, growth models, repeated measurements, longitudinal data, Bayesian credibility, shrinkage, R.

1. INTRODUCTION

Loss reserving theory and practice is undergoing a renaissance due to a recent proliferation of stochastic reserving techniques. To cite but a few examples, recent authors have applied regression analysis (Barnett and Zehnwirth [1]), generalized linear models (England and Verrall [2]), loss development growth curves together with maximum likelihood estimation (Clark [3]), and Bayesian methods (Meyers [4]) to model loss development data. Statistical modeling techniques are increasingly supplementing or supplanting spreadsheet-based projection methods for estimating ultimate losses.

This paper will propose yet another statistical framework for modeling loss triangles: *nonlinear hierarchical models*. These models are also commonly known as *nonlinear mixed effects* [NLME] *models*. The contention of this paper is that this class of models provides a highly flexible and natural framework within which the loss development process can be analyzed. The goals of this paper are twofold: to introduce the concept of hierarchical models and to illustrate how hierarchical models can be used in loss reserving.

Section 2 will sketch some of the basic theory of hierarchical models and also provide a hypothetical example illustrating how hierarchical modeling can be used to analyze longitudinal (or "repeated measurements") data. The relationship between hierarchical modeling and Bayesian credibility theory will also be discussed. These topics are not specific to loss reserving in particular, but are discussed in order to set the stage for Section 3.

Section 3, the main section of the paper, will broaden the discussion of hierarchical models to include non-linear model forms. Motivated by the very interesting presentation of Clark [3], our hierarchical models will incorporate the Weibull and Loglogistic "growth curve" functional forms. These models will be applied to the same loss triangle data analyzed by Clark and others. Many such growth curves are possible, but the Weibull and Loglogistic are two natural options. One way of understanding the loss reserving models proposed here is that they add "random effects" to the types of growth curve models introduced by Clark.

No attempt will be made in this paper to estimate reserve variability, which is beyond the scope of this introductory paper. This will be the subject of a future paper.

2. HIERARCHICAL MODELS

Generally speaking, hierarchical models are used when the data at hand are *grouped* in some important way. Examples include:

- The relationship between standardized test scores and prior grades of students from different high schools.
- Performance of a state's high schools, where schools are grouped into school districts.
- Expected workers compensation claims for exposures with various NCCI class codes.
- Expected loss ratio relativities for a personal auto carrier's various state territories.
- The growth of a collection of soybean plants, measured at various times since planting.

The first two examples are typical of the examples discussed in the social science literature (e.g., Gelman and Hill [5]). The third and fourth examples are classic problems of actuarial science, but are similar in form to the first two examples.

The final example is typical of hierarchical modeling applications in such fields as biology (e.g., modeling the growth of plants and animals) and pharmacology (e.g., modeling the effect of a drug over time). Many such examples are given in the book by Pinheiro and Bates [6]. In cases such as these, we have multiple measurements of each subject, performed at different points in time. Such multilevel datasets are commonly referred to as "longitudinal," "panel," or "repeated measurements" datasets. The primary goal of this paper is to convince the reader that loss reserving triangles can reasonably be regarded as longitudinal datasets, to which hierarchical modeling techniques naturally

apply.

The central concept of hierarchical models is that certain model parameters are *themselves* modeled. In other words, not all of the parameters in a hierarchical model are directly estimated from the data. Rather, (some of) the model parameters are calculated from estimates of the model's *hyperparameters*, which are in turn estimated from the data. Model parameters that are themselves given models are sometimes referred to as "random effects." They are to be distinguished from "fixed effects," which are not modeled, but are instead estimated directly from the data. "Mixed effects models," therefore, refer to models that contain both modeled and non-modeled parameters.

A note on terminology: this paper generally follows Gelman and Hill in favoring the language of hierarchical models over the "random/fixed/mixed effects" terminology. However, the phrase "random effects" will occasionally be used as shorthand for model parameters that are given sub-models. Many authors, including Pinheiro and Bates [6], speak mainly in terms of "mixed effects models." Note that Pinheiro and Bates wrote the "nlme" R function that was used to fit the hierarchical models described in this paper.

At this point an example might aid the discussion. Consider a hypothetical company that sells personal auto insurance in each of eight roughly equal-sized regions. We have data for the number of policies in force by region as of January 1, 2005, 2006, 2007, and 2008. We thus have 8*4=32 data points in all. We would like to build a model that could be used to forecast the number of policies in force, by region, in the coming years.

Using notation suggested by Gelman and Hill, let i denote the data point number and range from 1 to 32; similarly let j denote the region number. The term $j[i]$ will denote the group to which data point i belongs. For example, $j[5]=2$ because the fifth data point is an observation from Region 2. Two modeling strategies immediately suggest themselves.

Model 1 (complete pooling of data): First, we could simply pool the data from all eight regions and regress PIF (policies in force) on time.

$$PIF = \alpha + \beta t + \varepsilon$$

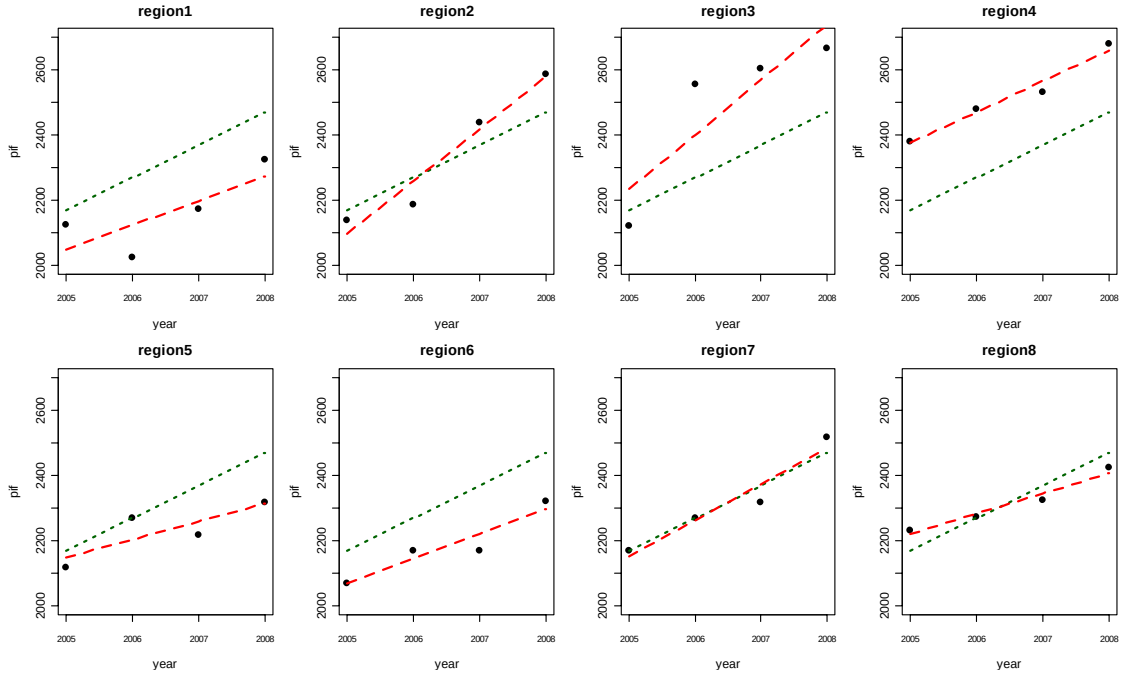
where $\varepsilon \sim N(0,\sigma^2)$. In this case the 32 data points would be used to estimate the three parameters $\{\alpha, \beta, \sigma\}$. Here we are effectively ignoring region.

Model 2 (separate models by region): Second, we could run eight separate regression models, one for each region.

$$PIF = \left\{\alpha^j + \beta^j t + \varepsilon^j\right\}_{j=1,2,\ldots,8}.$$

Note that each of these eight regression models is fit using only four data points.

These models are plotted in the figure below. The dotted lines represent Model 1 and are the same across all regions. The dashed lines represent Model 2 and vary from region to region. This plot illustrates why neither option is entirely satisfactory. At one extreme, the "pooled" model clearly provides poor fits in, for example, regions 1 and 4. At the other extreme, one might doubt that the data is sufficiently credible to support the fitting of eight region-specific models. For example, the first data point in region 3 appears to exert too much leverage on that model's parameters. A slope closer to that of the "pooled" model might be more believable.



Model 3 (include region indicator variables): Of course other conventional strategies are possible. For example, one could fit a no-intercept model that includes a separate indicator variable for each of the eight regions:

$$PIF = \beta_1(region == 1) + \ldots + \beta_8(region == 8) + \beta_9 t + \varepsilon.$$

This is a compromise between models 1 and 2. Like Model 1, it is a single "pooled" model that is fit to all of the data. Like Model 2, it allows us to capture region-specific aspects of the data. This is an improvement, but perhaps is still not ideal. We are still estimating 10 parameters – $\{\beta_1, \ldots, \beta_8, \beta_9, \sigma\}$

– using 32 data points. We face the danger of building an over-parameterized model. (Of course not all of the eight region indicators will necessarily be significant in the model. One or more of the indicators might be dropped.) The need to potentially add region/time interaction terms presents the possibility of further over-parameterization. In the extreme case where we need a separate intercept term and interaction with time for each region, we would need to estimate a model eight different intercepts and eight different slopes. This would essentially return us to Model 2.

Model 4 (random intercepts): Hierarchical modeling offers a different type of compromise. In this simple example, rather than estimate a separate "β" parameter for each region directly from the data, we specify a *Gaussian sub-model* of which eight region-specific intercept parameters $\{\alpha_1, \ldots, \alpha_8\}$ are random draws. Therefore, unlike $\{\beta_1, \ldots, \beta_8\}$ in Model 3, these so-called "random intercepts" $\{\alpha_1, \ldots, \alpha_8\}$ are not "estimated directly from the data." Rather, they are derived from the *hyperparameters* of the Gaussian sub-model.

Explicitly, this "random intercepts" hierarchical model can be written:

$$PIF = \alpha_1 + \ldots + \alpha_8 + \beta t + \varepsilon \quad where \quad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \quad and \quad \varepsilon \sim N(0, \sigma^2).$$

Or more compactly:

$$PIF_i \sim N(\alpha_{j[i]} + \beta t_i, \sigma^2) \quad where \quad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2).$$

In some circles it is conventional to call such a model a "mixed effects" model. The "slope" parameter β is called a "fixed effect," while the $\{\alpha_1, \ldots, \alpha_8\}$ parameters are called "random effects."

This hierarchical model contains four hyperparameters which can be estimated using maximum likelihood or a related optimization technique:

$$\hat{\mu}_\alpha = 2068.0 \quad \hat{\beta} = 100.06 \quad \hat{\sigma} = 81.13 \quad \hat{\sigma}_\alpha = 123.94.$$

Compare this with the 10 parameters estimated from the non-hierarchical regression model with a separate indicator variable for each region.

As noted above, the intercept "random effect" parameters $\{\alpha_1, \ldots, \alpha_8\}$ are derived using the model's estimated hyperparameters. Readers familiar with credibility theory might have anticipated that the formula used to do this is:

$$\hat{\alpha}_j = Z_j \cdot (\bar{y}_j - \beta \bar{t}_j) + (1 - Z_j) \cdot \mu_\alpha \quad where \quad Z_j = \frac{n_j}{n_j + \sigma^2 / \sigma_\alpha^2}.$$

In actuarial parlance, each random intercept α_j is a credibility-weighted average of the region-

specific intercept and the average (μ_α) of all of the region-specific intercepts. The credibility factor Z_j is determined in a familiar way using the number of observations for each region (n_j), the variance of the region specific intercepts (σ^2_α), and the residual variation σ^2.

Models 1 and 2, illustrated above, are special cases of this hierarchical model in a precise sense. As $\sigma^2_\alpha \rightarrow 0$, $Z_j \rightarrow 0$ and the hierarchical model approaches Model 1. As $\sigma^2_\alpha \rightarrow \infty$, $Z_j \rightarrow 1$ and the hierarchical model approaches Model 2 (Gelman and Hill [5] p. 258).

As an aside, it should be apparent that Bühlmann's credibility model is a specific instance of hierarchical models. If we remove the time covariate *t*, Model 4 becomes

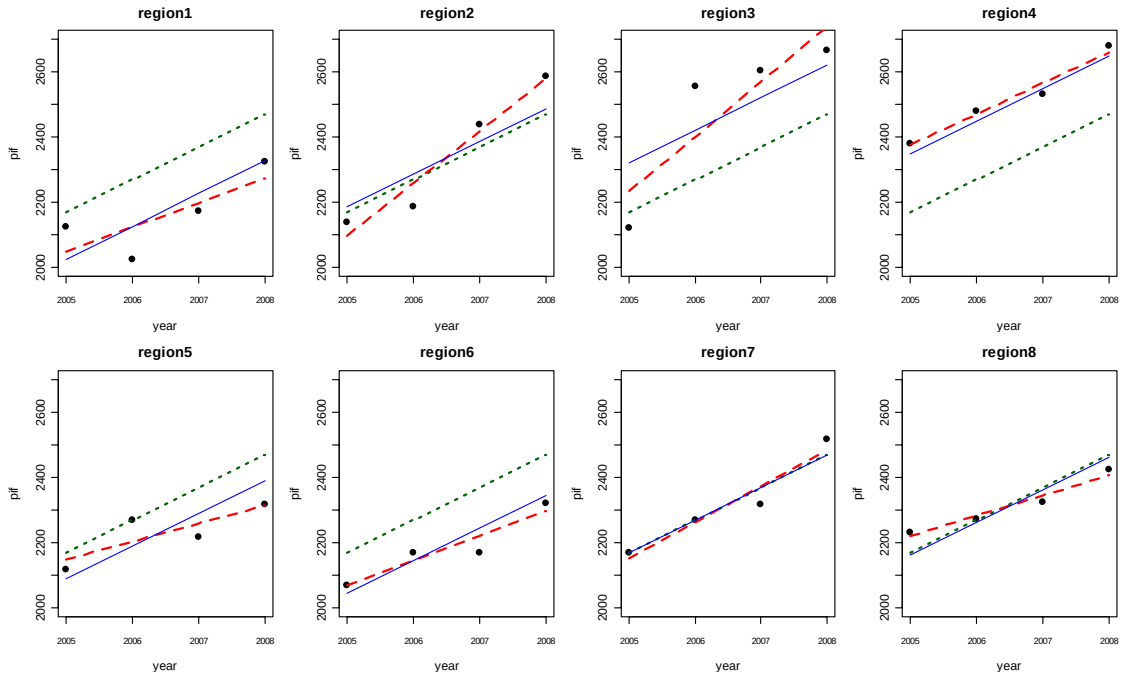
$$PIF_i \sim N(\alpha_{j[i]}, \sigma^2) \quad where \quad \alpha_j \sim N(\mu_\alpha, \sigma^2_\alpha).$$

And the credibility weighting expression becomes:

$$\hat{\alpha}_j \approx Z_j \cdot \bar{y}_j + (1 - Z_j) \cdot \mu \quad where \quad Z_j = \frac{n_j}{n_j + \sigma^2 / \sigma^2_\alpha}.$$

Frees [7 section 4.7] provides a helpful discussion of the ways in which several well-known credibility models are specific types of hierarchical models.

In the figure below, the predicted values of Model 4 (solid line) are added to the predicted values of Models 1 and 2. In certain cases (such as Regions 1 and 3) Model 4 appears to be an improvement over Model 2. This is because the more parsimonious Model 4 is not seriously leveraged by these regions' "2005" data points. For regions 2 and 8, on the other hand, Model 2 seems to fit the data better.



Model 5 (random slopes and intercepts): Because of the seemingly suboptimal fit on the model in Regions 2 and 8, one might consider adding a "slope random effect" to model 4. Explicitly:

$$PIF_i \sim N\left(\alpha_{j[i]} + \beta_{j[i]} \cdot t_i, \sigma^2\right) \quad where \quad \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_\alpha, \mu_\beta \end{bmatrix}, \Sigma\right) \quad , \quad \Sigma = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{bmatrix}.$$

Model 5 contains six hyperparameters: $\{\mu_\alpha,\mu_\beta,\sigma_\alpha,\sigma_\beta,\sigma_{\alpha\beta},\sigma\}$, two more than Model 4. Because Models 4 and 5 are nested models, we can compare their expected predictive accuracy by comparing their log-likelihoods and Akaike Information Criterion [AIC] statistics.

	LL	*d.f.*	*AIC*
Model 4	-186.20	4	380.40
Model 5	-184.32	6	380.64

Recall that AIC = -2*LL + 2*d.f., as can be confirmed from the above table. In a phrase, AIC is log-likelihood penalized for the number of hyperparameters in the model. The model with the lower AIC statistic is thought to make a better trade-off between complexity and goodness-of-fit, and is therefore expected to make more accurate predictions of future data.

Adding the further "random effect" to vary the slopes (in addition to varying the intercepts)

results in an improved log-likelihood; but a slightly worse AIC. This comparison suggests that it would be wise to favor the more parsimonious Model 4 above the slightly better fitting Model 5. The AIC comparison suggests that Model 5 might over-fit the data.

General Observations: Before turning to loss reserving, it is worth making a few general observations about the implications of hierarchical modeling for actuarial work. First, the hierarchical/multilevel modeling framework is a generalization of current actuarial modeling practice in two important ways.

- Actuaries often face a dilemma when faced with multilevel modeling situations. For example, should one pool one's data and build a single countrywide predictive model to be used in all states? Or should one build separate models by state? These options are analogous to Models 1 and 2 above. In the light of the above discussion, it should be clear that these two options are extreme cases (as the variance of a hierarchical model's random effects approach 0 and ∞, respectively) of a suitably specified hierarchical model.
- Bayesian credibility models are specific types of hierarchical models. Just as generalized linear models (GLMs) have provided a unifying framework for traditional minimum bias calculations, hierarchical modeling theory provides a unifying framework for Bayesian credibility modeling. This is helpful both pedagogically and practically. Pedagogically, it is helpful to understand the connection between Bayesian credibility and linear modeling. Practically, multilevel modeling packages can be used to perform Bayesian credibility calculations. In the same way that GLM modeling is less cumbersome than performing minimum bias calculations, hierarchical modeling packages allow one to perform Bayesian credibility calculations with a minimum of ad hoc programming. Furthermore, multilevel modeling packages make it easy to employ rigorous statistical methodology – such as graphical diagnostics and comparison of goodness-of-fit statistics – in one's work.

A second observation is that multilevel modeling potentially allows one to achieve a much better fit at the expense of adding only a few additional hyperparameters to a conventional GLM model. In the above example, Model 4 contains only one more hyperparameter than Model 1, but it provides a much better fit to the data. This is because the "scoring equation" for Model 4 contains nine parameters $\{\alpha_1, \ldots, \alpha_8, \beta\}$ as opposed to Model 1's insufficient $\{\alpha, \beta\}$ parameters. In short, actuaries can consider specifying hierarchical GLM models (HGLMs) as an alternative to purely "fixed effects" GLM models.

A related point is that the hierarchical modeling framework works well even if one's data contains a very large number of levels. The above example could easily be modified to involve four years of PIF data in each of 1000 counties. Model 4, with its four hyperparameters, or Model 5, with its six hyperparameters, would be no less applicable to this data. By comparison a traditional, non-hierarchical model would potentially need hundreds of indicator variables. In short, the hierarchical modeling framework provides a natural way to handle "massively categorical" variables in one's modeling work. This is because hierarchical modeling implicitly allows one to perform Bayesian credibility weighting within a GLM model building context.

These observations are not specific to loss reserving, but they set the stage for the hierarchical growth curve approach to loss reserving to be outlined in the next section.

3. HIERARCHICAL MODELS FOR LOSS RESERVING

The preceding section might have seemed like a long detour away from the topic of loss reserving. But it reviewed some of the concepts needed to build a hierarchical model of the loss development process. Consider a garden variety 10-by-10 loss triangle. Each of the 55 non-missing cells contains cumulative losses (CL), indexed by accident year AY and development period dev. We will treat this loss triangle as a multilevel dataset, in which each of the 10 accident years is a separate level. This will allow us to build a hierarchical model in which we "regress" cumulative losses CL on development period dev. The major disanalogy with the illustrative example in the previous section is that we must replace the linear regression with a non-linear model.

Pinheiro and Bates discuss three advantages of nonlinear hierarchical models, each of which apply in the context of loss reserving:

- **Interpretability.** The modeling approach to be outlined here requires that one explicitly model the loss development process in a specific functional form. Judgment as well as background empirical or theoretical knowledge can be used to guide the choice of nonlinear functional form.
- **Parsimony.** A well-chosen nonlinear function can model a non-linear process with fewer parameters than a linear model with multiple polynomial terms. In addition, as illustrated in the previous section, the hierarchical modeling approach potentially allows one to replace a potentially large number of subject-specific indicator variables and interaction terms with a small number of hyperparameters.

- **Validity beyond the observed range of the data.** Of course it is always dangerous to use a model to extrapolate beyond the data. However, the approach to be outlined here at least offers a framework within which one can harness one's background knowledge when specifying a model. Such an approach is less likely to lead one astray than a less parsimonious or more atheoretical "curve-fitting" approach.

Sample Dataset: To illustrate, we will work with the sample loss reserving dataset analyzed by Clark [3]. For ease of viewing, the cumulative loss numbers in the table below numbers have been divided by 1,000. These numbers are rounded only for the purpose of display; no rounding was done in performing the calculations.

Cumulative Losses in 1000's

AY	*12*	*24*	*36*	*48*	*60*	*72*	*84*	*96*	*108*	*120*	*reported*	*est ult*	*reserve*
1991	358	1,125	1,735	2,183	2,746	3,320	3,466	3,606	3,834	3,901	3,901	3,901	0
1992	352	1,236	2,170	3,353	3,799	4,120	4,648	4,914	5,339		5,339	5,434	95
1993	291	1,292	2,219	3,235	3,986	4,133	4,629	4,909			4,909	5,379	470
1994	311	1,419	2,195	3,757	4,030	4,382	4,588				4,588	5,298	710
1995	443	1,136	2,128	2,898	3,403	3,873					3,873	4,858	985
1996	396	1,333	2,181	2,986	3,692						3,692	5,111	1,419
1997	441	1,288	2,420	3,483							3,483	5,672	2,189
1998	359	1,421	2,864								2,864	6,787	3,922
1999	377	1,363									1,363	5,644	4,281
2000	344										344	4,971	4,627
chain link	3.491	1.747	1.455	1.176	1.104	1.086	1.054	1.077	1.018	1.000	34,358	53,055	18,697
chain ldf	14.451	4.140	2.369	1.628	1.384	1.254	1.155	1.096	1.018	1.000			
growth curve	6.9%	24.2%	42.2%	61.4%	72.2%	79.7%	86.6%	91.3%	98.3%	100.0%			

To provide a baseline for comparison, the results of a simple chain ladder calculation are displayed along with the raw data. All data was used to calculate each of the link ratios; and the 120→ultimate "tail factor" is assumed to be 1.0. According to this calculation, the expected total outstanding losses are approximately \$18.7M. The implied "growth curve" is simply the reciprocal of the sequence of loss development factors.

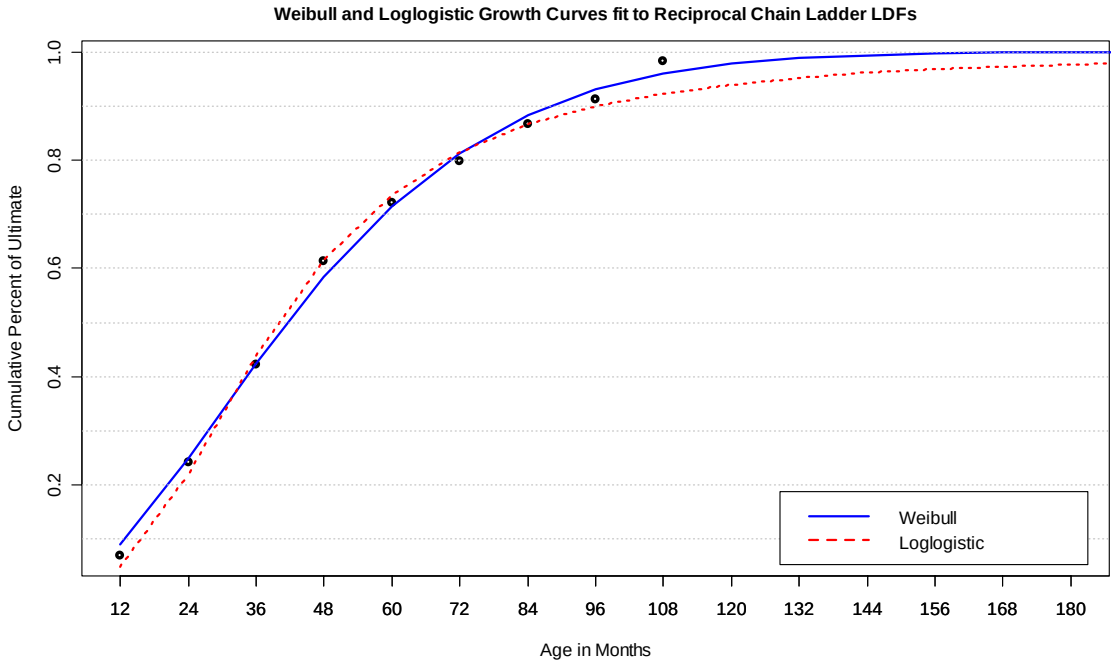
Clark's Models: The nine "growth curve" numbers resulting from the simple chain ladder exercise can be viewed as a piecewise linear approximation to a continuous growth curve. Clark considers two such growth curves, the Weilbull and Loglogistic, and integrates each of them into two models of the loss triangle data. The Weilbull growth curve has the form:

$$G(x \mid \omega, \theta) = 1 - \exp\left(-(x/\theta)^{\omega}\right).$$

The Loglogistic curve has the form:

$$G(x \mid \omega, \theta) = \frac{x^{\omega}}{x^{\omega} + \theta^{\omega}}.$$

Purely for illustration, we can fit each of these curves to the reciprocal of the chain ladder loss development factors (LDFs) displayed above. The resulting curves are displayed below, together with the reciprocal of the nine chain ladder LDFs.



This plot confirms that both the Weibull and Loglogistic growth curves are plausible candidates for modeling the loss development process. Each of the curves fits the reciprocal LDF pattern reasonably well. Note that the Loglogistic growth curve has a "heavier tail" than the Weibull, implying a longer loss development process and higher estimated ultimate losses. Note also that neither of the curves fits empirical development pattern perfectly. The Loglogistic curve fits the earlier data points better; whereas the Weibull curve is a bit closer to the final data point. In practice, one's background knowledge of the likely length of the loss development process would be used to decide between these, or other, growth curves. Following Clark, we fit sample models incorporating each of these growth curves and compare the results.

Clark proposes two models of the loss data. The first is called the "Loss Development Factor" (LDF) model, and can be expressed:

$$CL_{AY,dev} = ULT_{AY}\left[1 - G(dev \mid \omega, \theta)\right].$$

The function G can be the Weibull, the Loglogistic, or any other suitable growth function. The LDF model contains 12 parameters: $\{ULT_{1991}, \ldots, ULT_{2000}, \omega, \theta\}$.

(Note that Clark in fact models incremental rather than cumulative losses, and therefore specifies a formula that differs accordingly. Specifically, Clark's formula is

$$IL_{AY;x,y} = ULT_{AY}\left[(y \mid \omega, \theta) - G(x \mid \omega, \theta)\right]$$

where $IL_{AY;x,y}$ denotes the incremental losses in accident year AY between ages x and y. This is advantageous in that random noise at age x will not be propagated through ages x+1, x+2, and so on. For readability and ease of exposition, the models discussed in this paper are cast in terms of cumulative, rather than incremental, losses. However, it is a simple exercise to recast these models, as done above, in terms of incremental losses.)

Clark's calls his second model a "Cape Cod" model. Here the unknown parameters ULT_{AY} are replaced with $PREM_{AY} \cdot ELR$:

$$CL_{AY,dev} = PREM_{AY} \cdot ELR\left[1 - G(dev \mid \omega, \theta)\right].$$

$PREM_{AY}$ denotes on-leveled premium for accident AY (a known quantity). This model incorporates the Cape Cod assumption of a constant expected loss ratio (ELR) across all accident years. As a result, this model contains only three unknown parameters, {ELR,ω,θ}, as opposed to the LDF model's 12. The Cape Cod model is therefore less prone to overfitting the available data (in this illustration, 55 data points) than the LDF model. Clark points out that the less parsimonious LDF model results in more parameter variance, in turn resulting in more variance around the estimated reserves.

Baseline Hierarchical Model: It is possible to build hierarchical counterparts to each of Clark's models. Let us begin with Clark's LDF model. Rather than estimate the 10 parameters $\{ULT_{1991}, \ldots, ULT_{2000}\}$ directly from the data, we can model them in hierarchical fashion. Explicitly:

$$\begin{gathered} CL_{AY,dev} = ULT_{AY}\left[1 - \exp\left(-(x/\theta)^{\omega}\right)\right] + \varepsilon_{AY,dev} \\ ULT_{AY} \sim N(\mu_{ULT}, \sigma^2_{ULT}) \\ Var(\varepsilon_{AY,dev}) = \sigma^2 \hat{CL}_{AY,dev} \end{gathered}$$

This will be our baseline model. All of the alternate models to be discussed subsequently will be modifications of this baseline. The baseline model contains five unknown hyperparameters that must be estimated from the data: $\{\mu_{ULT}, \omega, \theta, \sigma_{ULT}, \sigma\}$. Specifying a sub-model of ULT_{AY} in the

above fashion is analogous to replacing the region-specific indicator variables in the previous section's PIF example with the "random intercepts" α_j.

Note that rather than assuming constant variance for each loss amount, we are assuming that the within-variance is proportional to the fitted value, where σ^2 is the proportionality constant. This corresponds to the over-dispersed Poisson assumption found in both England and Verrall [2] and Clark. We will relax this assumption shortly.

This model can easily be fit using the "nlme" ("non-linear mixed effects") function in R. (Please refer to the note at the end of this paper for information on how to obtain R and the nlme function.) The R code needed to do this is quite straightforward:

```
start.vals <- c(ult=5000, omega=1.4, theta=45)
w1 <- nlme(cum ~ ult*(1 - exp(-(dev/theta)^omega))
           , fixed = list(ult~1, omega~1, theta ~ 1)
           , random = ult ~ 1 | AY
           , weights = varPower(fixed=.5)
           , data=dat, start = start.vals)
```

Note most stochastic reserving techniques, this one included, require that one organize one's data in matrix rather than triangular form. The appendix to this paper displays the data in the form that it is read in prior to submitting the above R code.

We must supply starting values in order to estimate the parameters of a non-linear hierarchical model (starting values are not needed for linear hierarchical models). Choosing the appropriate starting values is something of an art. Still, in this particular case the model converges to the correct solution for a wide range of starting values. For example, replacing the above starting values with {10000, 2.0, 100} does not change the resulting model. However further changing the starting value of "ult" to 15000 causes the model not to converge. Changing the starting value for "omega" to 3.0, on the other hand, causes the model to converge to an incorrect solution. (A quick glance at a residual plot makes it clear that the solution is incorrect.) In most cases it should be possible to select a workable set of starting values using the estimated ultimate losses and implied growth curve from a simple chain ladder analysis.

Submitting the above R code yields the following estimates of the model's five hyperparameters. The model runs in seconds.

$$\mu_{ULT} = 5306.6 \quad \omega = 1.306 \quad \theta = 46.64 \quad \sigma_{ULT} = 543.03 \quad \sigma = 2.955$$

This model's AIC statistic is 725.76. Also, the *p*-values associated with μ, ω, and θ are all less than

.0001. The parameter error associated with this model is therefore fairly low.

We note in passing that the ω and θ parameter estimates for Clark's Weilbull LDF model are 1.297 and 48.885, respectively. These are reasonably consistent with our results.

The parameters and estimated ultimate losses and loss reserves resulting from the baseline model are displayed in the table below. The model's parameters (not hyperparameters) are listed in the omega, theta, and ULT columns. Because they were not given "random effects," omega and theta are the same for each accident year. We will shortly investigate the effect of adding random effects to the ω and θ parameters.

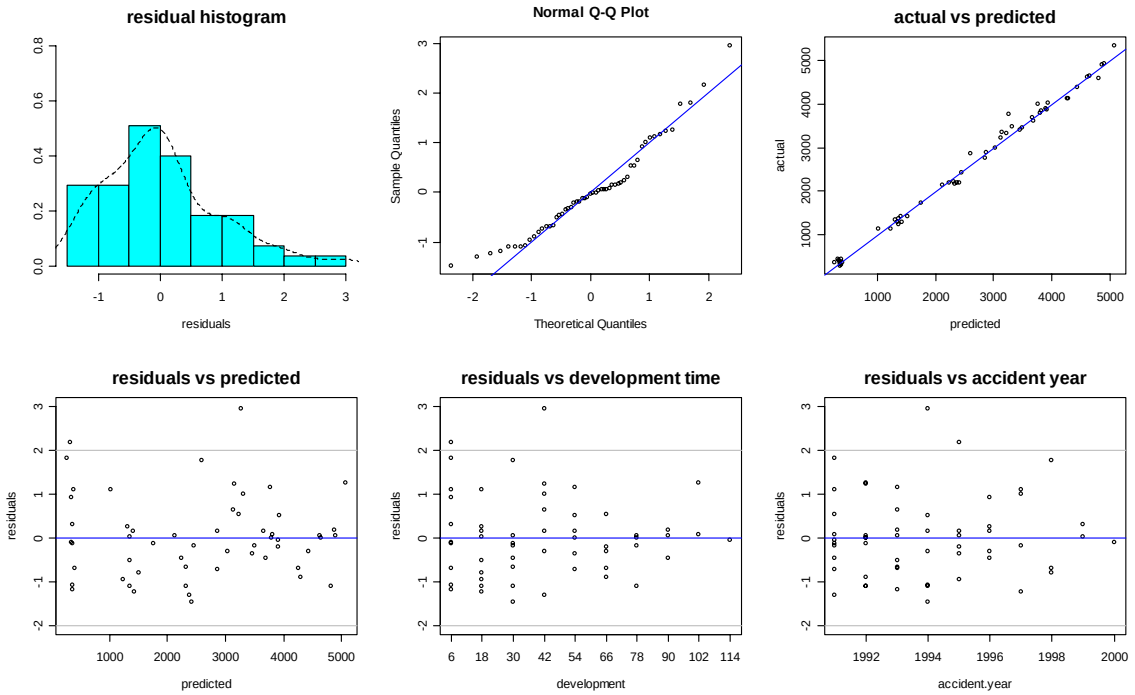
The key difference between this model and Clark's LDF model, is that here the estimated ultimate losses in the *ULT* column are *not* estimated directly from the data. Rather, they are derived from the estimates of the model's hyperparameters. Note that the average value of the *ULT* column is 5306.6, which is the same as the estimate of μ_{ULT}.

Parameters and Estimated Reserves - Baseline Model

AY	*dev*	*omega*	*theta*	*growth*	*reported*	*eval120*	*eval240*	*ULT*	*reserves*
1991	114	1.306	46.638	96.0%	3,901	3,943	4,073	4,074	172
1992	102	1.306	46.638	93.8%	5,339	5,239	5,412	5,413	74
1993	90	1.306	46.638	90.6%	4,909	5,207	5,379	5,380	470
1994	78	1.306	46.638	85.9%	4,588	5,423	5,602	5,603	1,015
1995	66	1.306	46.638	79.3%	3,873	4,777	4,935	4,936	1,062
1996	54	1.306	46.638	70.2%	3,692	5,052	5,219	5,220	1,528
1997	42	1.306	46.638	58.2%	3,483	5,512	5,694	5,695	2,212
1998	30	1.306	46.638	43.0%	2,864	5,850	6,043	6,044	3,180
1999	18	1.306	46.638	25.0%	1,363	5,255	5,429	5,430	4,067
2000	6	1.306	46.638	6.6%	344	5,101	5,270	5,271	4,927
total								53,066	**18,708**

The baseline model's estimate of the total outstanding losses is roughly $18.7M. This is virtually identical to the chain ladder's outstanding loss estimate displayed above. However, this similarity is a coincidence. The two models' reserve estimates differ considerably by accident year. For example, the chain ladder model's estimate of accident year 1998's outstanding losses is $3.92M, in contrast with the baseline hierarchical model's estimate of $3.18M.

Next we can inspect the standardized residuals and fitted values:



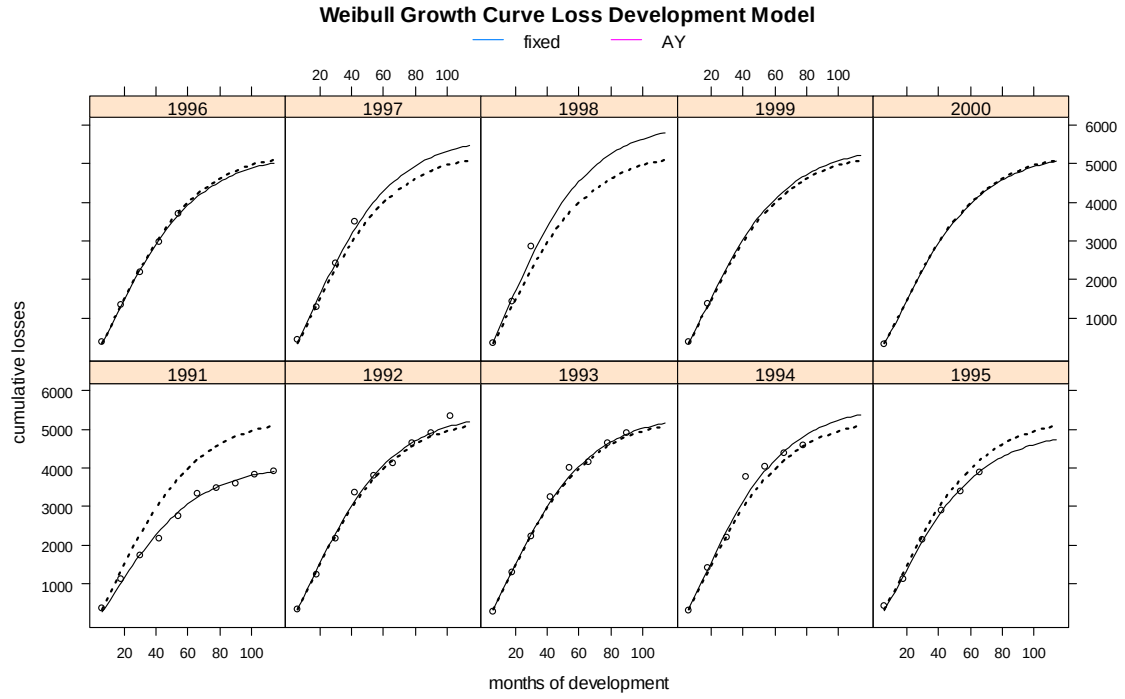
These diagnostic plots together indicate that the model fits the data reasonably well. However, the model is not perfect. The upper left two plots indicate that the standardized residuals are not quite normally distributed. Still, the deviation from normality is perhaps within the realm of acceptability. The "actual vs predicted" plot indicates a good fit. Consistent with this, the "residuals vs predicted" plot indicates that most of the standardized residuals are less than 2.0 in absolute value. A close inspection of this plot reveals an undulating pattern in the residuals: the model has a slight but systematic tendency for the model to under-estimate cumulative losses in the range of $1M-$3M and over-estimate cumulative losses in the $3M-$4M range. This suggests that the Weibull curve does not perfectly characterize the development of cumulative losses.

The general conclusion while the model could perhaps be improved upon, the overall fit is good. Four points are worth emphasizing:

- The model fits the data well despite the fact that it contains only five hyperparameters. In contrast, Clark's non-hierarchical LDF model contains 12 parameters; and the chain ladder analysis requires us to estimate nine link ratios (not including the arbitrary tail factor that must be added).
- Unlike Cape Cod-type models (to be described below), it is not necessary to bring in premium data or assume a constant expected loss ratio across accident years.

- This five-parameter model can be used to project losses to their ultimate values (or any intermediate value) without the need for a tail factor.
- The model's parsimony is made possible both by the hierarchical modeling methodology as well as the use of a non-linear growth function *G*.

Relating to this last point, another way to evaluate the model's fit is to superimpose each accident year's estimated growth curve on top of the cumulative loss observations. In the plot below, the (identical) dotted curves represent the "fixed" Weibull curve implied by the hyperparameters {ω, θ, μ_{ULT}}. The solid curves are the accident year-specific Weibull curves implied by ω and θ as well as the derived parameters $\{ULT_{1991},\ldots,ULT_{2000}\}$.



These plots further support the conclusion that our baseline hierarchical growth model fits the data well. In addition, they illustrate the basic intuition motivating the approach. Following Clark, we are modeling loss development as a *growth process*, in much the same way that a biostatistician would model the growth of a group of trees or soybean plants. In the latter cases, each "subject" is an individual tree or soybean plant and each observation is a measurement of size at various ages. In loss reserving, each "subject" is the aggregate claims from an accident year and each observation is the aggregate cumulative losses at various development ages.

Before continuing, it is worth commenting on the growth curve plots for accident years 1991 and

1998. Note that the 1991 growth curve is different from the other year's growth curves. This is reflected the ULT_{1991} parameter of 4.074M, which is more than 20% lower than the average μ_{ULT}=5.3066M. However, we have 10 AY 1991 observations, all of which fall squarely on the 1991-specific growth curve. This suggests that the low value of ULT_{1991} is justified.

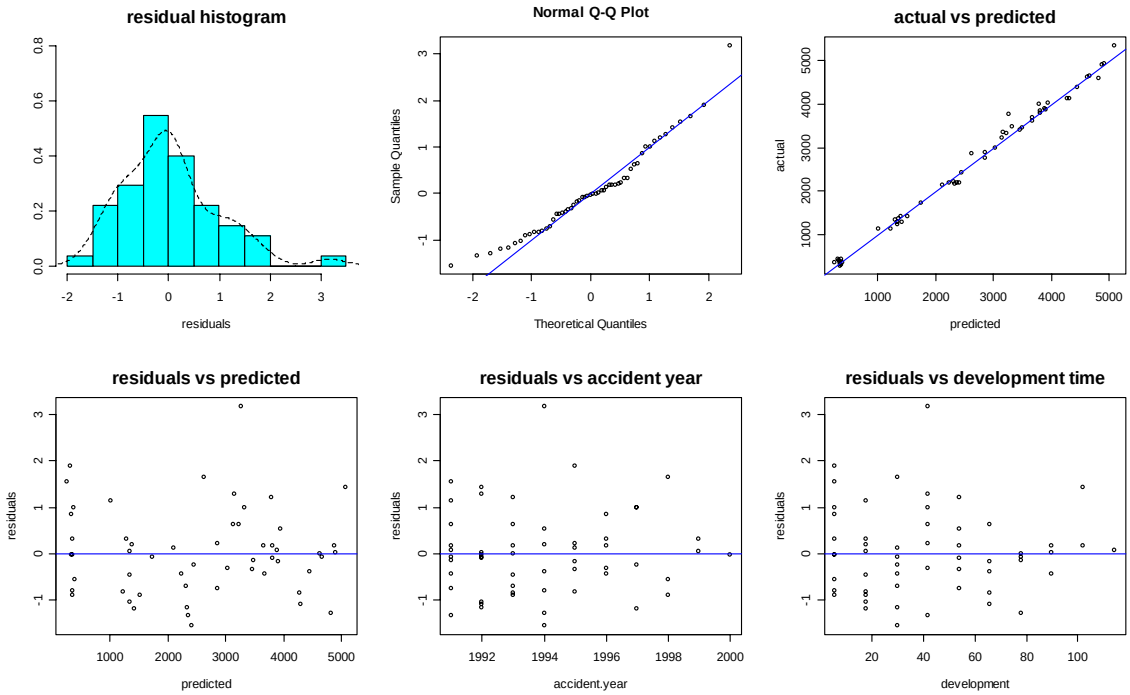
In contrast, the estimated ultimate losses for 1998 are approximately 6.044M, 14% higher than average. This is of course driven by only three data points, which have greater 12→24 and 24→36 developments than their counterparts in other accident years. The chain ladder method produces an even higher estimate of 1998 ultimate losses: 6.787M. The hierarchical model's estimate therefore falls between the global average μ_{ULT} and the chain ladder estimate. This is illustrative of the way in which the hierarchical model implicitly uses a type of "credibility weighting" to "shrink" the accident-year specific estimates towards the global mean. The amount of "shrinkage" is more pronounced for more recent accident years. The most extreme amount of shrinkage occurs for accident year 2000: the estimated ultimate losses for this year are \$5.271M, only a fraction of a percent lower than the global mean of \$5.3066M. Little credibility is given to the single data point for accident year 2000.

Relaxing the Process Variance Assumption: Recall that the baseline model contains the assumption that the within-variance is proportional to the fitted value. We can replace this with the weaker assumption that:

$$Var(\varepsilon_{AY,dev}) = \sigma^2 (\hat{CL}_{AY,dev})^{2\varsigma}.$$

In other words, rather than pre-specify that ζ=0.5, we can introduce ζ as a further model hyperparameter to be estimated. This means that our model will contain six, rather than five, hyperparameters. (In R, this is achieved by simply removing the "fixed=0.5" from inside the "varPower" expression.)

The resulting estimate is $\zeta \approx 0.37$. Although not displayed here, the estimated loss reserves of this model are, in aggregate, only \$100,000 (or 0.5%) less than that of the baseline model. The residual plot indicated an improved residual histogram, but otherwise little difference in the goodness of fit. For simplicity we will therefore continue with the baseline model.



Random Shape Effect: The baseline model incorporates the assumption that the different accident years' ultimate losses vary randomly about a mean value: $ULT_{AY} \sim N(\mu_{ULT}, \sigma^2_{ULT})$. It also incorporates the assumption that the shape (ω) and scale (θ) characterizing the loss development process do *not* vary by accident year. Just as we were able to vary slope – in addition to intercept – by region in the previous section's PIF example, here we have the option of allowing ω and/or θ to vary by accident year.

To illustrate, we expand our model to include varying shape parameters $\{\omega_{1991},\ldots, \omega_{2000}\}$ by accident year. Specifically:

$$CL_{AY,dev} = ULT_{AY}\left[1-\exp\left(-(x/\theta)^{\omega}\right)\right]+\varepsilon_{AY,dev}$$

$$\begin{pmatrix} ULT_{AY} \\ \omega_{AY} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{ULT} \\ \mu_{\omega} \end{pmatrix},\Sigma\right), \quad \Sigma = \begin{pmatrix} \sigma^2_{ULT} & \sigma_{ULT,\omega} \\ \sigma_{ULT,\omega} & \sigma^2_{\omega} \end{pmatrix}.$$

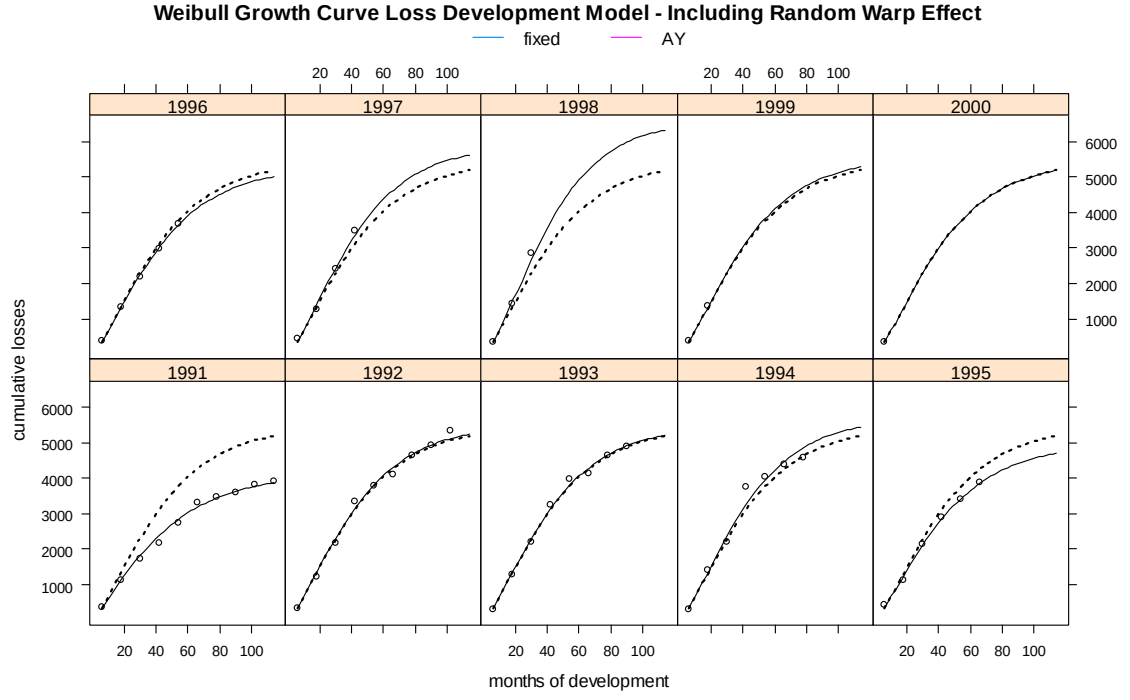
$$Var(\varepsilon_{AY,dev}) = \sigma^2 \hat{CL}_{AY,dev}$$

This model contains the two new hyperparameters σ_{ω} and $\sigma_{\omega,ULT}$ in addition to the baseline model's five hyperparameters. The resulting model parameters and associated loss reserve estimates are displayed below:

Model Parameters and Estimated Reserves

AY	dev	omega	theta	growth	reported	eval120	eval240	ULT	reserves
1991	114	1.189	47.202	95.8%	3,901	3,907	4,101	4,105	203
1992	102	1.313	47.202	93.5%	5,339	5,281	5,462	5,463	124
1993	90	1.311	47.202	90.2%	4,909	5,259	5,440	5,441	532
1994	78	1.332	47.202	85.5%	4,588	5,491	5,667	5,668	1,080
1995	66	1.265	47.202	78.8%	3,873	4,744	4,933	4,935	1,061
1996	54	1.292	47.202	69.7%	3,692	5,052	5,236	5,238	1,546
1997	42	1.347	47.202	57.6%	3,483	5,662	5,835	5,835	2,352
1998	30	1.410	47.202	42.5%	2,864	6,368	6,525	6,525	3,661
1999	18	1.317	47.202	24.7%	1,363	5,325	5,504	5,505	4,142
2000	6	1.308	47.202	6.5%	344	5,229	5,410	5,411	5,067
total								54,126	**19,768**

Note that the parameters in the "omega" column now vary by accident year. The expected ultimate reserves are approximately $1M (5%) higher than those of the more parsimonious baseline model. It is interesting to note that nearly half of this increase comes from the increase in AY 1998 reserves from $3.18M in the baseline model to $3.661M here. At the same time the ω shape parameter for AY 1998 is 1.410 – the highest of all accident years. Allowing the shape parameter to vary by accident year therefore results in an accident year 1998 reserve estimate that is nearly as high as that of the chain ladder model's estimate.



The AY 1998 component of the above plot suggests that adding the random shape effect gives the most recent observation from AY 1998 more leverage over that accident year's growth curve. This

might be appropriate – perhaps accident year 1998's claims are expected to be of higher ultimate severity. We can also note that the random shape model's AIC is 720.79, down from the baseline model's 725.76. This suggests that the random shape model offers a better tradeoff between complexity and goodness of fit.

Of course, it is equally possible that the most recent 1998 observation is an outlier, in which case we would want to mitigate its leverage on the ultimate loss estimate. Assuming the latter is correct, we would favor the baseline model over the random shape alternative model. For simplicity, we will continue to work with the baseline model.

Random Scale Effect: We can similarly allow the scale parameter θ to vary by accident year. Doing so causes the AIC measure to deteriorate from 725.76 to 729.76. Therefore on this dataset, allowing θ to vary by accident year does not offer a sufficient improvement in fit to justify the additional complexity. It is interesting to note that the (θ, σ_θ) hyperparameters of this model are 46.6375 and 0.0000094, respectively. In other words the estimate of θ is nearly identical in the baseline and random scale models; and the estimated size of the random scale effect is negligible.

To summarize, where Clark's LDF model requires a separate ultimate loss parameter for each accident year, we allow ultimate loss (*ULT*) to randomly vary by accident year using a Gaussian sub-model. In addition, there is perhaps some justification for allowing the shape parameter (ω) to similarly vary by accident year. But doing so heightens the danger of overfitting the data. In the absence of compelling prior knowledge in support of including a random shape effect, one might be inclined to exclude it. Finally, the data indicates that the scale parameter (θ) does not vary by accident year. There is therefore no justification for including a random scale effect.

Loglogistic Growth Curves: Next, we can test the effect of replacing the Weibull growth curve with a Loglogistic growth curve: $G(x|\omega,\theta)=x^\omega/(x^\omega+\theta^\omega)$. This is achieved by changing a single line of our R code:

```
start.vals <- c(ult=5000, omega=1.4, theta=45)
l1 <- nlme(cum ~ ult*(dev^omega)/((dev^omega) + (theta^omega))
           , fixed = list(ult~1, omega~1, theta ~ 1)
           , random = ult ~ 1 | AY
           , weights = varPower(fixed=.5)
           , data=dat, start = start.vals)
```

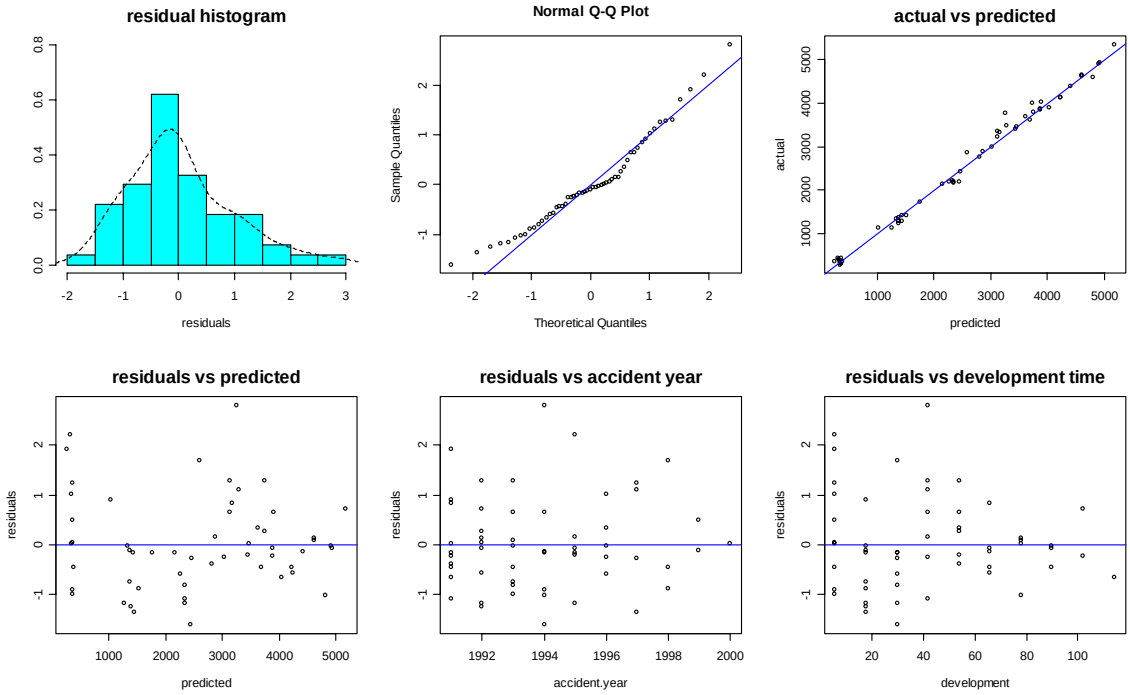
As with the baseline Weibull model, we allow only the *ULT* parameter to vary by accident year – no random shape or scale effects are included. The resulting hyperparameter estimates are:

$$\mu_{ULT} = 6898.3 \quad \omega = 1.403 \quad \theta = 49.14 \quad \sigma_{ULT} = 702.8 \quad \sigma = 3.109 .$$

Note in passing that Clark reports ω and θ parameter estimates of 1.434 and 48.63, respectively for his Loglogistic LDF model.

Immediately we can see that the Loglogistic model will result in considerably higher loss reserve estimates than the Weibull model: the μ_{ULT} hyperparameter was 5306.6 for the Weibull model, compared with 6898.3 for the Loglogistic model.

The residual plots suggest that the Loglogistic model also fits the data fairly well. It is not clear from these plots that the Loglogistic model fits the data substantially better or worse than the baseline Weibull model.



The model parameters and expected loss reserves are displayed below:

Model Parameters and Estimated Reserves

AY	*dev*	*omega*	*theta*	*growth*	*reported*	*eval120*	*eval240*	*ULT*	*reserves*
1991	114	1.404	49.135	76.5%	3,901	4,099	4,756	5,269	1,368
1992	102	1.404	49.135	73.6%	5,339	5,471	6,348	7,034	1,694
1993	90	1.404	49.135	70.0%	4,909	5,458	6,333	7,017	2,107
1994	78	1.404	49.135	65.7%	4,588	5,696	6,609	7,322	2,734
1995	66	1.404	49.135	60.2%	3,873	5,020	5,825	6,454	2,580
1996	54	1.404	49.135	53.3%	3,692	5,294	6,142	6,805	3,113
1997	42	1.404	49.135	44.5%	3,483	5,742	6,662	7,381	3,898
1998	30	1.404	49.135	33.3%	2,864	6,055	7,026	7,784	4,920
1999	18	1.404	49.135	19.6%	1,363	5,454	6,329	7,012	5,648
2000	6	1.404	49.135	5.0%	344	5,372	6,234	6,906	6,562
total								68,984	**34,626**

Again, these results are broadly consistent with those reported by Clark. As anticipated, the estimated reserve amount – $34.6M – is quite a bit higher than the $18.7 estimated by the baseline Weibull model. But as Clark points out, one should be careful using a heavy-tailed model such as the Loglogistic to extrapolate too many years beyond the data. If, following Clark, we compute the reserves using losses projected to 240 months (the "eval240" column in the table above), the resulting reserve estimate is $27.9M. Again, this is broadly consistent with Clark's result ($28.9M). This is more realistic than using the Loglogistic model to extrapolate the results "to infinity." However, the result is still somewhat disconcerting: the reserve estimate after arbitrarily truncating the Loglogistic growth curve at 240 months is still nearly 50% higher than the corresponding Weibull models' reserve estimate.

The moral is that much hinges on the form of the growth curve one chooses for one's model. The advantage discussed by Pinheiro and Bates – validity of the model beyond the observed range of the data – is meaningful only to the extent that the model has been chosen wisely. In practice the considerations one would use to choose a growth curve are similar to considerations that are used in choosing a tail factor. The above display shows that, according to the Loglogistic model, the losses are only 76.5% developed as of 120 months. In contrast, the baseline Weibull model implies that the losses are 96% developed as of 120 months. One's general knowledge of how rapidly the types of claims being modeled develop should be considered when deciding which is the more appropriate growth curve, or whether additional growth curves should be investigated.

"Cape Cod" Models: If we have access to exposure information in addition to loss development data, it is easy to recast our hierarchical growth model into what might be called "Cape Cod" form. In the Cape Cod method, one assumes that expected ultimate loss ratio is constant across accident years and either estimates it from the data or simply introduces it as a model

assumption. In the hierarchical modeling framework, we can dispense with the assumption that loss ratio is common across accident years. Rather, we can provide a sub-model for the various accident years' loss ratios, just as we provided a sub-model for the various accident years' ultimate losses in the baseline model. Still, we are acting in the original spirit of the Cape Cod method in the sense that we include the average loss ratio across all accident years as a model hyperparameter.

We will modify our original baseline Weibull model:

$$\begin{aligned} CL_{AY,dev} &= ULT_{AY}\left[1-\exp\left(-(x/\theta)^{\omega}\right)\right]+\varepsilon_{AY,dev} \\ ULT_{AY} &\sim N(\mu_{ULT},\sigma^2_{ULT}) \\ Var(\varepsilon_{AY,dev}) &= \sigma^2 \hat{CL}_{AY,dev} \end{aligned} \quad .$$

The "Cape Cod" counterpart is:

$$\begin{aligned} CL_{AY,dev} &= prem_{AY} LR_{AY}\left[1-\exp\left(-(x/\theta)^{\omega}\right)\right]+\varepsilon_{AY,dev} \\ LR_{AY} &\sim N(\mu_{LR},\sigma^2_{LR}) \\ Var(\varepsilon_{AY,dev}) &= \sigma^2 \hat{CL}_{AY,dev} \end{aligned} \quad .$$

In other words, we replace the hyperparameters $\{\mu_{ULT}, \sigma_{ULT}\}$ with $\{\mu_{LR}, \sigma_{LR}\}$.

The corresponding modification of our R code is equally minor:

```
prem <- seq(from=0, length=10, by=400) + 10000
prem <- rep(prem, 10:1)
start.vals <- c(lr=.5, omega=1.4, theta=45)
cc.w1 <- nlme(cum ~ prem*lr*(1 - exp(-(dev/theta)^omega))
            , fixed = list(lr~1, omega~1, theta ~ 1)
            , random = lr ~ 1 | AY
            , weights = varPower(fixed=.5)
            , data=dat, start = start.vals)
```

(Note that the loss triangle analyzed by Clark and others was originally not accompanied by premium information. Clark therefore assumed that the premium was \$10M in 1991 and increased by \$400,000 in each subsequent year. This is done in the first two lines of code above.)

Recall that the parameter estimates for the baseline Weibull model are:

$$\mu_{ULT} = 5306.6 \quad \omega = 1.306 \quad \theta = 46.64 \quad \sigma_{ULT} = 543.03 \quad \sigma = 2.955 .$$

In contrast the parameter estimates for the "Cape Cod" Weibull model are:

$$\mu_{LR} = 0.4634 \quad \omega = 1.317 \quad \theta = 49.91 \quad \sigma_{LR} = 0.0383 \quad \sigma = 2.977 .$$

It is comforting to note that the estimates of both process error (σ) and of the parameters determining the average shape of the loss development curve ($\{\omega,\theta\}$) are fairly consistent across both of these models.

Although it will not be reproduced here, the residual plot for the "Cape Cod" model is virtually identical to that of the baseline model. The various parameters and resulting loss reserve estimates for the Cape Cod Weibull model are displayed below:

Model Parameters and Estimated Reserves -- Cape Cod Weibull Model

AY	*dev*	*prem*	*omega*	*theta*	*lr*	*growth*	*reported*	*eval120*	*lr*prem*	*reserves*
1991	114	10,000	1.317	46.910	0.408	96.0%	3,901	3,952	4,082	181
1992	102	10,400	1.317	46.910	0.519	93.8%	5,339	5,229	5,401	62
1993	90	10,800	1.317	46.910	0.498	90.5%	4,909	5,208	5,380	470
1994	78	11,200	1.317	46.910	0.501	85.8%	4,588	5,433	5,611	1,023
1995	66	11,600	1.317	46.910	0.429	79.1%	3,873	4,818	4,977	1,103
1996	54	12,000	1.317	46.910	0.440	70.0%	3,692	5,114	5,283	1,591
1997	42	12,400	1.317	46.910	0.467	57.9%	3,483	5,608	5,792	2,309
1998	30	12,800	1.317	46.910	0.486	42.6%	2,864	6,016	6,215	3,350
1999	18	13,200	1.317	46.910	0.439	24.7%	1,363	5,613	5,798	4,435
2000	6	13,600	1.317	46.910	0.446	6.4%	344	5,871	6,064	5,720
total									54,604	**20,245**

The total reserves estimate by this model is $20.2M: about 8% higher than the baseline Weibull result. Most of the additional $1.5M of estimated reserves come the increased reserve estimates for accident years 1998-2000. This is an expected and sensible result. The ultimate loss estimates for the earlier accident years, where more loss development information is available, are less affected by the premium information. Conversely, the more recent the accident year, the less loss development data is available. Therefore, the ultimate loss estimates depend more heavily on the model's *LR* hyperparameter (the "Cape Cod" loss ratio estimate) together with the premium information.

Recall that Clark's Cape Cod model contains only three parameters (ω, θ, *ELR*) in contrast with his LDF model's 11 parameters. Because we are building hierarchical models there is not such a dramatic difference between our baseline model and its "Cape Cod" counterpart. Each of these models contains five hyperparameters.

Each of these models – the baseline and the Cape Cod variant – offers an advantage over its non-hierarchical counterpart:

- The hierarchical baseline model is less prone to overparameterization because it does require not a separate ultimate loss parameter for each accident year. The parameters $\{ULT_{1991}, \ldots, ULT_{2000}\}$ are replaced with the $\{\mu_{ULT},\sigma_{ULT}\}$ hyperparameters.

- The "Cape Cod" hierarchical model does not require one to assume a constant loss ratio across all accident years. This hierarchical model approaches the Clark Cape Cod model as the hyperparameter $\sigma_{LR} \rightarrow 0$.

We are not arguing that the hierarchical Cape Cod model is not an improvement on its baseline counterpart. On the contrary, the Cape Cod hierarchical model is preferable because adding exposure information will typically yield improved ultimate loss estimates, especially for more the recent, data-sparse, accident years. This, not greater parsimony, is the benefit it offers over the baseline hierarchical model.

Reserve Variability: Estimating the variability around a hierarchical growth model's loss reserve estimates (reserve variability) will be the topic of a future paper. For now a few brief comments must suffice. The problem of estimating reserve variability is twofold: we must estimate the variability resulting from the stochastic nature of the loss development process (process variance); and we must also estimate the variability resulting from the uncertainty around our models' hyperparameters (parameter variance). The future paper will outline a simulation-based approach to estimate the variability arising from both of these sources. In particular, Markov Chain Monte Carlo (MCMC) simulation, a technique widely used in contemporary Bayesian statistics, will be used to estimate parameter variance.

Of course, model risk – illustrated above by the dramatic effect that the choice of growth functions has on one's ultimate loss estimate – will remain a serious issue even after process and parameter variance have been accounted for.

4. CONCLUSION

Hierarchical modeling in actuarial science is an idea whose time has come. Hierarchical models encompass Bayesian credibility theory and therefore allow actuaries to perform credibility calculations within a statistical modeling framework. Moreover, hierarchical models allow one to easily integrate credibility concepts into one's GLM or non-linear modeling activities. By incorporating sub-models of various model parameters, hierarchical models allow one to strengthen an estimate for a sparsely populated segment of one's data by appropriately weighting it with the overall average estimate for the population as a whole. This integrates the fundamental insight of Bayesian credibility into a statistical modeling framework. For classification ratemaking and predictive modeling applications, actuaries can consider adding hierarchical structure to their

generalized linear models in order to account for the variation along such "massively categorical" dimensions as territory or class code.

Turning to loss reserving, hierarchical modeling is useful in that it provides a natural way to analyze longitudinal (or "repeated measures") datasets. The point of view of this paper is that traditional loss reserving triangles can be viewed as longitudinal datasets in which each accident year is a "subject" and the cumulative or incremental losses at various development times constitute a series of repeated observations.

Unlike ratemaking and other general insurance predictive modeling applications, loss reserving is best approached using non-linear models. Following Clark, we have explored the use of the Weibull and Loglogistic growth curves for modeling the loss development process. We have done this in a non-linear hierarchical modeling (or "non-linear mixed effects models" – NLME) context. Hierarchical modeling allows us to specify sub-models for one or more of the parameters that determine the loss development process. The result is a natural and flexible framework in which to build parsimonious loss reserving models. Furthermore, the use of growth curves eliminates the need to specify arbitrary tail factors.

A Note Regarding Software

All models discussed in this paper were fit using the freely available R statistical computing package. R is available at http://www.r-project.org. Once the base R package has been installed, the multilevel modeling packages "lmer" and "lme" can easily be added.

Acknowledgments

I would like to thank Bill Stergiou for several enjoyable conversations about hierarchical modeling and also for introducing me to the Gelman-Hill textbook. I also thank Dave Clark, Keith Curley, Ravi Kumar, Glenn Meyers, Bill Stergiou, and Steve Talley for their many helpful comments on an earlier draft of this paper. When I have not taken their advice, it has been at my peril. Any errors are entirely my own.

Appendix: Raw Loss Triangle Data, as Imported into R

AY	dev	cum
1991	6	357.848
1991	18	1124.788
1991	30	1735.33
1991	42	2182.708
1991	54	2745.596
1991	66	3319.994
1991	78	3466.336
1991	90	3606.286
1991	102	3833.515
1991	114	3901.463
1992	6	352.118
1992	18	1236.139
1992	30	2170.033
1992	42	3353.322
1992	54	3799.067
1992	66	4120.063
1992	78	4647.867
1992	90	4914.039
1992	102	5339.085
1993	6	290.507
1993	18	1292.306
1993	30	2218.525
1993	42	3235.179
1993	54	3985.995
1993	66	4132.918
1993	78	4628.91
1993	90	4909.315
1994	6	310.608
1994	18	1418.858
1994	30	2195.047
1994	42	3757.447
1994	54	4029.929
1994	66	4381.982
1994	78	4588.268
1995	6	443.16
1995	18	1136.35
1995	30	2128.333
1995	42	2897.821
1995	54	3402.672
1995	66	3873.311
1996	6	396.132
1996	18	1333.217
1996	30	2180.715
1996	42	2985.752
1996	54	3691.712
1997	6	440.832
1997	18	1288.463
1997	30	2419.861
1997	42	3483.13
1998	6	359.48
1998	18	1421.128
1998	30	2864.498
1999	6	376.686
1999	18	1363.294
2000	6	344.014

5. REFERENCES

[1] Barnett, G., and B. Zehnwirth, "Best Estimates for Reserves," *PCAS* 2000, Vol. 87, 245-321.
[2] England, P. D., and R. J. Verrall, "Stochastic Claims Reserving in General Insurance," *British Actuarial Journal* 2002, Vol. 8, 443-544.
[3] Clark, D. R., "LDF Curve Fitting and Stochastic Loss Reserving: A Maximum Likelihood Approach," Casualty Actuarial Society *Forum*, Fall 2003, 41-91.
[4] Meyers, G., "Estimating Predictive Distributions for Loss Reserve Models" *Variance*, Vol 1, No. 2, Fall 2007, 248-272.
[5] Gelman, A. and J. Hill (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, New York.
[6] Pinheiro, J. C. and D. M. Bates (2000). Mixed-Effects Models in S and S-Plus, Springer-Verlag, New York.
[7] Frees, E. (2006). Longitudinal and Panel Data Analysis and Applications in the Social Sciences, Cambridge University Press, New York.