A Multivariate Bayesian Claim Count Development Model With Closed Form Posterior and Predictive Distributions

Stephen J. Mildenhall
Filename: cdm.tex (975)

January 2, 2005

Abstract

We present a rich, yet tractable, multivariate Bayesian model of claim count development. The model combines two conjugate families: the gamma-Poisson distribution for ultimate claim counts and the Dirichlet-multinomial distribution for emergence. We compute closed form expressions for all distributions of actuarial interest, including the posterior distribution of parameters and the predictive multivariate distribution of future counts given observed counts to date and for each of these distributions give a closed form expression for the moments. A new feature of the model is its explicit sensitivity to ultimate claim count variability and the uncertainty surrounding claim count emergence. Depending on the value of these parameters, the posterior mean can equal the Borhuetter-Ferguson or chain-ladder reserve. Thus the model provides a continuum of models interpolating between these common methods. We give an example to illustrate use of the model.

JEL Classification G - Financial Economics; G220 - Insurance; Insurance Companies

Keywords Loss Development, Chain-Ladder Method, Borhuetter-Ferguson Method, Dirichlet-multinomial, Poisson-gamma

1 Introduction

We present a Bayesian model of claim count development. The model is rich enough to provide a realistic model for the practitioner but at the same time it is mathematically tractable and we give explicit equations for the posterior and predictive distributions. The predictive distribution is an example of a generalized power series distribution and a generalized hypergeometric distribution. The method in the paper will be of interest to practicing actuaries because it is easy to implement and it provides explicit posterior distributions for unreported claims, and hence Bayesian means and confidence intervals, and a rationale for choosing between existing reserving methods. The model is theoretically interesting because the posterior mean generalizes three common reserving methods (the peg, the Borhuetter-Ferguson and the chain-ladder) in an intuitive and insightful manner.

Actuaries today are asked to provide a distribution of potential outcomes or a confidence interval around the point estimates they have traditionally supplied. The push towards greater quantification of uncertainty is particularly marked in the property and casualty loss reserving practice. Understanding reserve uncertainty and linking the pricing actuary's prior estimate of ultimate losses to the reserving actuary's posterior estimates is therefore becoming more and more important.

These recent demands on the profession have played up some shortcomings of the traditional chain-ladder method of determining loss reserves. The chain-ladder method is simple to apply and easy to explain, and is the *de facto* standard reserve method. Mack's 1993 paper [15] showing how to compute the standard error of chain ladder reserves was an important enhancement to the method. However, the chain-ladder is still not well suited to providing explicit posterior distributions, nor does it provide diagnostic information to assess model fit. The latter point is a severe weakness in practice. There is no *one* chain-ladder method; the technique can be applied to a variety of different loss development triangles in slightly different ways. (Academic discussions usually assume link ratios are stable over time—something rarely seen in practice—and use the weighted average of all years link ratios.) When the various chain-ladder related estimates do not agree there is no statistical guidance on which method to prefer. The shortcomings of the chain-ladder have been discussed in the literature. Mack [14] identifies the stochastic assumptions which underlie the chain-ladder method. Venter [26] discusses the assumptions required for the chain-ladder estimates to produce least-squares optimal reserve estimates, and discusses some alternative methods when the conditions are not met. Renshaw and Verrall [23] describe a statistical procedure which is exactly equivalent to the chain-ladder in almost all circumstances. We will discuss their model more in Section 6.

In order to address these shortcomings, and respond to the demand for more precise quantification of uncertainty, both practicing actuaries and academics have explored alternative models. Zehnwirth [31] and Zehnwirth and Barnett [32] construct general linear models of reserve development based on log-incremental data. Kunkler [12] uses a mixed model to include zero claims in a log-incremental model. England and Verrall [7] and Wright [30] discuss generalized linear models, the latter taking an operation time point of view. Norberg [20] models the claims process as a non-homogeneous marked Poisson process. There has been considerable interest in Bayesian models of development. Reserving involves the periodic update of estimates based on gradually emerging information—a naturally Bayesian situation. Bayesian methods have been explored by Robbins [24], de Alba [5], Dellaportas and Ntzoufras [21], Renshaw and Verrall [23], and Verrall [27, 28], amongst others. Stephens et al. [25] use a survival time approach to modeling claim closure in a Bayesian framework. As Robbins points out, the mathematics of Bayesian models often becomes intractable. One advantage of this paper's model is the closed mathematical form of all the distributions of interest. For the less tractable models the WinBUGs MCMC system has been applied. See Verrall [28] for a very detailed explanation of how to do this.

Despite all of these advances, no model has come close to challenging the chain ladder method. In part this reflects the difficulties a new method faces before it becomes accepted practice. It also reflects the technical complexity of some of the alternative models. Practicing actuaries can be uncomfortable with the assumptions[1] and the number of parameters. The chain-ladder method has one parameter for each development period: the link-ratios and the tail factor. Regression models may produce a model with fewer parameters, but the model itself is often selected from a very large number of potential models. This can lead to generalization error where a particular model can over-fit artifacts in a small data set. The advantages of "simple" models are discussed in Balasubramanian [2] and Domingos [6].

There is, therefore, a need for a simple statistical model of loss development to augment and enhance the chain-ladder method. A new model should have a similar number of parameters to the chain-ladder, should be fit to the data using a statistical technique such as maximum likelihood, should be able to incorporate prior information from the pricing department, and should be easily updated with observed loss information as it becomes available. We will present such a model for claim count development in this paper. The model is introduced in Section 2. The explicit form of the marginal distributions of claims reported in each period is proved in Section 3. Sections 4 and 5 prove results about the conditional and predictive distributions. Section 6 discusses where our model fits within the continuum of reserving models, from the "book plan" peg method through the chain-ladder method. Section 7 discusses parameter estimation. Section 8 applies the model to a specific triangle. Finally, Section 9 will discuss extending the model to loss development, rather than just claim count development.

[1] A difficulty with *explicit* assumptions is the disquiet they can cause!

This paper focuses on the theoretical development of the claim count model. However, I want to stress how easy and useful this model will be in practice. I spell out exactly how to apply the model and provide several more examples in Mildenhall [18]

Notation

The following notational convention will be use extensively in the paper. For any n-tuple $x_1, \dots, x_n$ define

$$x(t) = \sum_{i=1}^{t} x_i, \quad x'(t) = \sum_{i=t+1}^{n} x_i$$

and let $x := x(n)$. Thus $x = x(n) = x(t) + x'(t)$ for all $t = 1, \dots, n$. This notation will apply to B, b, π, and v. It will be re-iterated before it is used.

The letters p and $q := 1-p$ will be used as parameters of a gamma distribution and will never have subscripts.

2 The GPDM Bayesian Claim Count Model

The gamma-Poisson Dirichlet-multinomial, or GPDM, claim count model is a combination of a gamma-Poisson random variable for total claims and a Dirichlet-multinomial distribution for the distribution of claims by report period. For a particular accident year, let B_i be the incremental number of claims reported in period $i = 1, \dots, n$. We assume that nth report is ultimate and will not model further claim emergence. Let

$$B(n) = B_1 + \cdots + B_n \tag{1}$$

denote the ultimate number of claims.

The GPDM is defined as a combination of two conjugate models. The ultimate number of claims $B(n) \mid \Lambda = \lambda$ has a Poisson distribution with mean λ.

Λ has a gamma prior distribution. Conditional on $B(n)$ and parameters $\Pi_1 = \pi_1, \ldots, \Pi_{n-1} = \pi_{n-1}$, $\pi_n = 1 - \sum_{i-1}^{n-1} \pi_i$, the claim emergence $B_1, \ldots, B_n$ has a multinomial distribution with parameters $B(n), \pi_1, \ldots, \pi_n$. $\Pi_1, \ldots, \Pi_{n-1}$ have a Dirichlet prior distribution. The full vector of parameters is $\Theta = (\Lambda, \Pi_1, \ldots, \Pi_{n-1})$. Conditional on

$$\Theta = \theta := (\lambda, \pi_1, \ldots, \pi_{n-1}) \qquad \pi_n = 1 - \sum_{i=1}^{n-1} \pi_i. \tag{2}$$

the GPDM probability density is

$$\Pr(B_1, \ldots, B_n \mid \Theta = \theta) = \frac{e^{-\lambda} \lambda^{b(n)}}{b_1! \ldots b_n!} \pi_1^{b_1} \ldots \pi_n^{b_n} \tag{3}$$

where $b(n) = \sum_{i=1}^{n} b_i$ and the two $b(n)!$ terms have cancelled.

The prior densities for the parameter vector $\Theta = (\Lambda, \Pi_1, \ldots, \Pi_{n-1})$ are

$$\Lambda \sim \text{Gamma}(r, p/q), \quad q = 1 - p, \tag{4}$$

$$\Pr(\Lambda = \lambda) = \frac{p^r}{q^r \Gamma(r)} \lambda^{r-1} e^{-\lambda p/q} \tag{5}$$

and

$$(\Pi_1, \ldots, \Pi_{n-1}) \sim \text{Dirichlet}(v_1, \ldots, v_n), \tag{6}$$

$$\Pr(\Pi_1 = \pi_1, \ldots, \Pi_{n-1} = \pi_{n-1}) = \frac{\Gamma(v_1 + \cdots + v_n)}{\Gamma(v_1) \cdots \Gamma(v_n)} \pi_1^{v_1 - 1} \cdots \pi_n^{v_n - 1}. \tag{7}$$

Λ and the Π_i are *a priori* independent.

The form of the gamma distribution in Eqn. (4) is chosen so that the predictive distribution for $B(n)$ is

$$\Pr(B(n) = b(n)) = \binom{r + b(n) - 1}{b(n)} p^r q^{b(n)}. \tag{8}$$

Thus $\mathrm{E}(B(n)) = rq/p$ and $\mathrm{Var}(B(n)) = rq/p^2$. If $\mathrm{E}(B(n)) = m$ then $p = r/(r+m)$, $q = m/(r+m) = 1/(1+m/r)$ and $\mathrm{Var}(B(n)) = m(1+m/r)$. The coefficient of variation of the gamma distribution is $1/\sqrt{r}$. The expression $1/r$ is sometimes called the contagion, see Mildenhall [17, Section 2.2].

Compared to traditional methods of reserving the GPDM includes two new parameters: r which controls the variability of ultimate claim counts and the extra Dirichlet paramter which controls the variability of claim emergence. The Borhuetter-Ferguson method of reserving, by contrast, assumes a prior estimate of the ultimate number of claims but no measure of its variability. The chain-ladder does not assume a prior estimate of ultimate claims, but gives full credibility to observed claim emergence which corresponds to a high degree of confidence in estimates of Π_i. These two extra parameters determine the behaviour of the GPDM model.

Pricing actuaries often have prior estimates of expected frequency because the frequency-severity approach is a common pricing method. Thus reserving actuaries can usually obtain a prior mean for the number of ultimate claims expected from a block of business. We want to be able to incorporate this information into our claim count model. Eqn. (3) assumes that the ultimate claim count, $B(n)$, has a negative binomial distribution. The r parameter for $B(n)$ is a measure of the inhomogeneity of insureds or of non-diversifiable parameter risk; it could be estimated based on line of business studies. The negative binomial has been suggested as a more flexible alternative to the Poisson distribution for modeling claim counts by many authors, including Klugman, Panjer and Willmot [10]. Also see the references in Johnson et al. [8].

The second part of Eqn. (3) is the multinomial with Dirichlet conjugate prior. Basic properties of Dirichlet-multinomial (DM) are given in Bernardo and Smith [3] and Johnson et al. [9, Section 35.13.1]. For more details on the Dirichlet see Kotz et al. [11]. The Dirichlet distribution has n free parameters (compared to only $n-1$ free π_i because of the condition $\sum_i \pi_i = 1$), and the extra parameter controls uncertainty in the proportions. When $n = 2$ the Dirichlet becomes a beta distribution.

The DM distribution with parameters $(b(n); v_1, \dots, v_n)$ has predictive probability density function

$$\Pr(B_1 = b_1, \dots, B_n = b_n) = \frac{b(n)!}{b_1! \cdots b_n!} \frac{\Gamma(\sum v_i)}{\Gamma(b(n) + \sum v_i)} \prod_{i=1}^{n} \frac{\Gamma(b_i + v_i)}{\Gamma(v_i)} \tag{9}$$

where $b(n) = \sum_i b_i$. We can write Eqn. (9) more succinctly using the Pochhammer symbol $(r)_k$. For a real r and non-negative integer k define

$$(r)_k := r(t+1) \cdots (r+k-1) = \frac{\Gamma(r+k)}{\Gamma(r)}. \tag{10}$$

Then Eqn. (9) becomes

$$\Pr(B_1 = b_1, \dots, B_n = b_n) = \frac{b(n)!}{b_1! \cdots b_n!} \frac{1}{(\sum v_i)_{b(n)}} \prod_{i=1}^{n} (v_i)_{b_i}. \tag{11}$$

We will use the Pochhammer symbol extensively.

The marginal distributions of the DM are beta-binomial mixtures. Let $v = \sum_i v$, then

$$\mathrm{E}(B_i) = n v_i / v, \tag{12}$$

$$\mathrm{Var}(B_i) = \frac{b(n) + v}{1 + v} \left(\frac{b(n) v_i (v - v_i)}{v^2} \right) \tag{13}$$

$$\mathrm{Cov}(B_i, B_j) = -\frac{b(n) + v}{1 + v} \frac{b(n) v_i v_j}{v^2} \tag{14}$$

$$\mathrm{Corr}(B_i, B_j) = -\sqrt{\frac{v_i v_j}{(v - v_i)(v - v_j)}}. \tag{15}$$

The marginal and conditional distributions of a DM are also DMs. See Johnson et al. [9, Section 35.13.1] for these facts.

The next lemma, and its obvious generalizations, follows from the properties of the Dirichlet and multinomial distributions. We will use it several times in various guises.

Lemma 1 *Let* $B_1, \dots, B_n \mid \Theta$ *be a GPDM. Then* $B_1+B_2, \dots, B_n \mid \Theta' = (\lambda, \pi_1+\pi_2, \pi_3, \dots, \pi_n)$ *is also a GPDM.*

Proof: This follows from [9, Chapter 35 Section 13.1]. ■

We end this section by computing the predictive distribution of $B_1, \dots, B_n$ given no observations. Let $v = \sum v_i$. Then

Proposition 1 *Let* $B_1, \dots, B_n \mid \Theta$ *be a GPDM. Then*

$$\Pr(B_1 = b_1, \dots, B_n = b_n) = \frac{b(n)!\Gamma(v)}{\Gamma(b(n)+v)} \prod_{i=1}^{n} \frac{\Gamma(b_i+v_i)}{b_i!\Gamma(v_i)} \binom{r+b(n)-1}{b(n)} p^r q^{b(n)}. \tag{16}$$

Proof: We have

$$\begin{aligned}
&\Pr(B_1 = n_1, \dots, B_n = b_n) \\
&\quad = \int \dots \int \Pr(B_1, \dots, B_n \mid \Theta) f(\lambda) f(\pi_1, \dots, \pi_n) d\lambda d\pi_1 \dots d\pi_{n-1} \\
&\quad = \int \dots \int \left(\frac{e^{-\lambda}\lambda^{b(n)}}{b(n)!}\right) \left(\frac{b(n)!}{b_1! \dots b_n!} \pi_1^{b_1} \dots \pi_n^{b_n}\right) \frac{p^r}{\Gamma(r) q^r} \lambda^{r-1} e^{-p\lambda/q} \\
&\qquad \times \frac{\Gamma(v)}{\prod \Gamma(v_i)} \prod_{i=1}^{n} \pi_i^{v_i-1} d\lambda d\pi_1 \dots d\pi_{n-1} \\
&\quad = \frac{b(n)!}{b_1! \dots b_n!} \frac{\Gamma(v)}{\prod \Gamma(v_i)} \frac{p^r}{\Gamma(r) q^r b(n)!} \\
&\qquad \times \int \dots \int \left(\int e^{-\lambda(1+p/q)} \lambda^{b(n)+r-1} d\lambda \right) \prod_{i=1}^{n} \pi_i^{b_i+v_i-1} d\pi_1 \dots d\pi_{n-1} \\
&\quad = \frac{b(n)!\Gamma(v)}{\Gamma(b(n)+v)} \prod_{i=1}^{n} \frac{\Gamma(b_i+v_i)}{b_i!\Gamma(v_i)} \binom{r+b(n)-1}{b(n)} p^r q^{b(n)}
\end{aligned}$$

since the inner integral with respect to λ equals $\Gamma(b(n)+r) q^{b(n)+r}$ and $1 + p/q = 1/q$. ■

Eqn. (16) can also be written more compactly as

$$\Pr(B_1 = b_1, \dots, B_n = b_n) = p^r q^b \frac{(r)_{b(n)}}{(v)_{b(n)}} \prod_{i=1}^{n} \frac{(v_i)_{b_i}}{b_i!}. \tag{17}$$

3 Marginal Distributions

The GPDM is a tractable distribution because it is possible to write down closed-form and easy-to-compute expressions for its conditional marginal distributions and its predictive distribution of future claims given observed claims to date. The marginals are necessary to compute likelihoods from whole or partial claim count development triangles. The predictive distributions provide a conditional distribution for ultimate claims given counts to date. We now prove these two important results, starting with marginal distributions in this section.

The marginal and conditional distributions of an GPDM are hypergeometric distribution and use the Gaussian hypergeometric functions ${}_2F_1(a, b; c; z)$. Mathematicians and actuaries today may not be as familiar with hypergeometric functions as their counter-parts would have been 50 or 100 years ago. Given this lack of familiarity expressions involving ${}_2F_1$ can be a little forbidding. It is important to remember that ${}_2F_1$ is no more mysterious than the other functions built-in to most calculators and spreadsheets. Indeed, it is very easy to program ${}_2F_1$ into a spreadsheet and use it like a built-in function. The properties of ${}_2F_1$ we use, together with pseudo-code to compute it, are given in Appendix A.

The next proposition computes the marginal distribution of $B_1, \dots, B_t$ for $t < n$. Obviously an analogous result would hold for any subset of the B_i. Remember that $v = \sum_{i=1}^{n} v_i$, $v'(t) = \sum_{i=t+1}^{n} v_i$, $b(t) = \sum_{i=1}^{t} b_i$ and $\pi(t) = \sum_{i=1}^{t} \pi_i$.

Proposition 2 *Let $B_1, \dots, B_n \mid \Theta$ have a GPDM distribution. If $t \le n-1$ then the marginal distribution of $(B_1, \dots, B_t \mid \Theta)$ is also GPDM with*

$$\Pr(B_1 = b_1, \dots, B_t = b_t \mid \Theta = (\lambda, \pi_1, \dots, \pi_{n-1})) \tag{18}$$

$$= \Pr(B_1 = b_1, \dots, B_t = b_t \mid (\pi(t)\lambda, \frac{\pi_1}{\pi(t)}, \dots, \frac{\pi_t}{\pi(t)})). \tag{19}$$

The predictive marginal of $(B_1, \ldots, B_t)$ *is*

$$\begin{aligned}\Pr(B_1 = b_1, \ldots, B_t = b_t) &= p^r q^{b(t)} \frac{\Gamma(v)}{\Gamma(b(t)+v)} \frac{\Gamma(b(t)+r)}{\Gamma(r)} \prod_{i=1}^{i=t} \frac{\Gamma(b_i + v_i)}{\Gamma(v_i) b_i!} \\ &\qquad \times {}_2F_1(v'(t), b(t)+r; b(t)+v; q) \quad (20) \\ &= p^r q^{b(t)} \frac{(r)_{b(t)}}{(v)_{b(t)}} \prod_{i=1}^{i=t} \frac{(v_i)_{b_i}}{b_i!}\, {}_2F_1(v'(t), b(t)+r; b(t)+v; q).\end{aligned}$$

Proof: Using Lemma 1 we can sum the unobserved variables $(B_{t+1}, \ldots, B_n)$ and, without loss of generality, assume that $t = n - 1$.

Then $\Pr(B_1 = b_1, \ldots, B_t = b_t \mid \Theta)$

$$\begin{aligned}&= \sum_{b_n \geq 0} \Pr(B_1 = b_1, \ldots, B_n = b_n \mid \Theta) \\ &= \sum_{b_n \geq 0} \left(b(n)! \prod_{i=1}^{n} \frac{\pi_i^{b_i}}{b_i!} \right) \frac{e^{-\lambda} \lambda^{b(n)}}{b(n)!} = \left(\prod_{i-1}^{n-1} \frac{\pi_i^{b_i}}{b_i!} \right) \left(\sum_{b_n \geq 0} \frac{\lambda^{b_n} \pi_n^{b_n}}{b_n!} \right) e^{-\lambda} \lambda^{b(n-1)} \\ &= \frac{\pi_1^{b_1} \ldots \pi_t^{b_t}}{b_1! \ldots b_t!} \lambda^{b(t)} e^{-\pi(t)\lambda} \\ &= b(t)! \frac{(\pi_1/\pi(t))^{b_1} \ldots (\pi_t/\pi(t))^{b_t}}{b_1! \ldots b_t!} \frac{(\pi(t)\lambda)^{b(t)} e^{-\pi(t)\lambda}}{b(t)!} \\ &= \Pr\big(B_1 = b_1, \ldots, B_t = b_t \mid (\pi_n \lambda, \frac{\pi_1}{\pi(t)}, \ldots, \frac{\pi_t}{\pi(t)})\big).\end{aligned}$$

Next, using Proposition 2 and remembering $t = n - 1$, we have $\Pr(B_1 =$

$b_1, \ldots, B_t = b_t)$

$$
\begin{aligned}
&= \sum_{b_n \geq 0} \Pr(B_1 = b_1, \ldots, B_t = b_t, B_n = b_n) \\
&= \sum_{b_n \geq 0} \frac{b(n)!\Gamma(v)}{\Gamma(b(n)+v)} \prod_{i=1}^{n} \frac{\Gamma(b_i + v_i)}{b_i!\Gamma(v_i)} \binom{r + b(n) - 1}{b(n)} p^r q^{b(n)} \\
&= \Gamma(v) p^r q^{b(t)} \prod_{i=1}^{t} \frac{\Gamma(b_i + v_i)}{b_i!\Gamma(v_i)} \sum_{b_n \geq 0} \frac{b(n)!}{b_n!} \frac{\Gamma(b_n + v_n)}{\Gamma(v_n)\Gamma(b(n)+v)} \frac{\Gamma(r + b(n))}{\Gamma(r) b(n)!} q^{b_n} \\
&= p^r q^{b(t)} \frac{\Gamma(v)\Gamma(b(t)+r)}{\Gamma(b(t)+v)\Gamma(r)} \prod_{i=1}^{t} \frac{\Gamma(b_i + v_i)}{b_i!\Gamma(v_i)} \sum_{b_n \geq 0} \frac{(v_n)_{b_n} (b(t)+r)_{b_n}}{(b(t)+v)_{b_n} b_n!} q^{b_n} \\
&= p^r q^{b(t)} \frac{\Gamma(v)\Gamma(b(t)+r)}{\Gamma(b(t)+v)\Gamma(r)} \prod_{i=1}^{t} \frac{\Gamma(b_i + v_i)}{b_i!\Gamma(v_i)} \,{}_2F_1(v_n, b(t)+r; b(t)+v; q)
\end{aligned}
$$

since $b(n) = b(n-1) + b_n = b(t) + b_n$. ■

To evaluate Eqn. (20) use the log-gamma function and convert the product of gamma functions into a sum and difference of log's and then exponentiate. This avoids potential over- or under-flow problems.

It follows that the marginal distribution of B_1 is

$$
\Pr(B_1 = b_1) = p^r q^{b_1} \frac{\Gamma(v)\Gamma(b_1 + r)}{\Gamma(b_1 + v)\Gamma(r)} \frac{\Gamma(b_1 + v_1)}{b_1!\Gamma(v_1)} \,{}_2F_1(v_2, b_1 + r; b_1 + v; q). \tag{21}
$$

Since the two components of a GPDM are *a priori* independent Eqn. (12) implies the mean of B_1 is

$$
\begin{aligned}
\mathrm{E}(B_1) &= \mathrm{E}(\mathrm{E}(B_1 \mid B(n))) \\
&= \mathrm{E}(B(n)\Pi_1) = \mathrm{E}(\Lambda\Pi_1) = \mathrm{E}(\Lambda)\mathrm{E}(\Pi_1) \\
&= v_1 m / v.
\end{aligned} \tag{22}
$$

The variance of B_1 can be computed using Eqn. (13):

$$\begin{aligned}\mathrm{Var}(B_1) &= \mathrm{E}(\mathrm{Var}(B_1 \mid B(n))) + \mathrm{Var}(\mathrm{E}(B_1 \mid B(n))) \\ &= \frac{v_1(v-v_1)}{v^2(1+v)}\mathrm{E}(B(n)^2) + \frac{v_1(v-v_1)}{v(1+v)}\mathrm{E}(B(n)) + \frac{v_1^2 m(1+m/r)}{v^2} \\ &= \frac{mv_1}{v} + \frac{m^2 v_1(v(1+r^{-1}) - v_1(1-r^{-1}v))}{v^2(1+v)} \end{aligned} \tag{23}$$

since $\mathrm{E}(B(n)) = m$, $\mathrm{Var}(B(n)) = m(1+mr^{-1})$ and $\mathrm{E}(B(n)^2) = m + m^2(1+r^{-1})$. Similarly, the covariance of B_1 and B_2 can be computed using Eqn. (14):

$$\begin{aligned}\mathrm{Cov}(B_1, B_2) &= \mathrm{E}(\mathrm{Cov}(B_1, B_2 \mid B(n))) + \mathrm{Cov}(\mathrm{E}(B_1 \mid B(n)), \mathrm{E}(B_2 \mid B(n))) \\ &= -\mathrm{E}\left[\frac{B(n)+v}{1+v}\frac{B(n)v_i v_j}{v^2}\right] + \mathrm{Cov}(v_1 B(n)/v, v_2 B(n)/v) \\ &= m^2 \frac{v_1 v_2 (r^{-1}v - 1)}{v^2(1+v)}. \end{aligned} \tag{24}$$

Eqn. (14) shows that the covariance between two marginals of a Dirichlet-multinomial is always negative. Eqn. (24) shows that the covariance between two marginals of a GPDM is negative if $v < r$, and positive otherwise. It becomes positive because the effect of the common mixing through the gamma prior for λ overwhelms the negative correlation given $B(n)$.

We will show in Section 6 that when $r = v$ the GPDM produces the Borhuetter-Ferguson reserve; when $r > v$, and there is less uncertainty in the prior ultimate than emergence, it favors the peg method; and when $r < v$ it favors the chain-ladder method. Which of these methods is indicated depends on the data being analyzed. Common practice favors the chain-ladder and Borhuetter-Ferguson methods; the peg method is rarely used. Thus we expect to find that $r \le v$ in data.

When $r \to \infty$ the variance of the gamma prior tends to zero and the ultimate claim count distribution tends to a Poisson with mean λ. The the marginal distribution of $B_1, \ldots, B_{n-1}$ becomes

$$\begin{aligned}\Pr(B_1 = b_1, \ldots, B_t = b_t) = e^{-\lambda}\lambda^{b(t)} \frac{\Gamma(v)}{\Gamma(b+v)} \prod_{i=1}^{i=t} \frac{\Gamma(b_i + v_i)}{\Gamma(v_i) b_i!} \\ \times {}_1F_1(v'(t); b(t)+v; q)\end{aligned} \tag{25}$$

where ${}_1F_1$ is a confluent hypergeometric function.

4 Posterior Distributions

In this section we consider the posterior distribution of Θ given observed development data:

$$\Pr(\Theta \mid \text{data}) = \frac{\Pr(\text{data} \mid \Theta)\Pr(\Theta)}{\Pr(\text{data})} \propto \Pr(\text{data} \mid \Theta)\Pr(\Theta). \tag{26}$$

When we are trying to identify the posterior distribution we can ignore any variable which is not a function of the parameters Θ.

Our data consists of multivariate observations of development data $B_1, \ldots, B_n$. However for all but the oldest accident year we only have a partial observation $B_1, \ldots, B_t$ for some $t < n$ with which to update the distribution of Θ. Recall that the prior distribution of $\Theta = (\Lambda, \Pi_1, \ldots, \Pi_n)$ is

$$\Pr(\Theta) = \Gamma(r, p/q) \times \text{Di}(v_1, \ldots, v_n) \tag{27}$$

where Λ has a gamma distribution, the proportions Π_i have a Dirichlet distribution and the two distributions are *a priori* independent. The next proposition shows how to update the prior distribution of Θ given a partial observation of claim counts. Let $\pi(t) = \sum_{i=1}^t \pi_i,\ \pi'(t) = \sum_{i=t+1}^n \pi_i$ and $b(t) = \sum_{i=1}^t b_i$.

Proposition 3 *Let $B_1, \ldots, B_n \mid \Theta$ have a GPDM distribution and let $t \leq n$. Then the posterior distribution of Θ given a partial observation $B_1, \ldots, B_t$ has density*

$$\begin{aligned}\Pr(\Theta = (\lambda, \pi_1, \ldots \pi_{n-1}) \mid B_1 = b_1, \ldots, B_t = b_t) = \\ \kappa\ \lambda^{b(t)+r-1} e^{-\lambda(p/q+\pi(t))} \pi_1^{b_1+v_1-1} \cdots \pi_t^{b_t+v_t-1} \pi_{t+1}^{v_{t+1}-1} \cdots \pi_n^{v_n-1}\end{aligned} \tag{28}$$

where

$$\begin{aligned}\kappa = \Gamma(v + b(t)) \Big(q^{r+b(t)} \Gamma(r + b(t)) \prod_{i=1}^t \Gamma(b_i + v_i) \\ \times \prod_{i=t+1}^n \Gamma(v_i) \ \times \ {}_2F_1(v'(t), b(t) + r; b(t) + v; q) \Big)^{-1}\end{aligned} \tag{29}$$

In Eqn. (28) the distribution of Λ is dependent on the distribution of observed claims through the term $\pi(t)$, so the two have become entangled. This is the reserving conundrum: counts through t periods are higher than expected; is this because we have observed a greater proportion of ultimate claims than expected or because ultimate claims will be higher than expected? Our model will show how to answer this question. When $t = n$, and we have a full observation, the posterior is no longer entangled because $\pi(n) = 1$; the posterior distribution is again a product of independent gamma and Dirichlet distributions.

Proof: Using Eqn. (3), the prior distribution for Θ, and the multinomial expansion in the penultimate step, we have $\Pr(\Theta \mid B_1, \dots, B_t)$

$$
\begin{aligned}
&\propto \Pr(B_1, \dots, B_t \mid \Theta)\Pr(\Theta) \\
&= \sum_{b_{t+1},\dots,b_n} \frac{1}{b_1! \cdots b_n!} \pi_1^{b_1} \cdots \pi_n^{b_n} \lambda^{b(n)} e^{-\lambda} \lambda^{r-1} e^{-p\lambda/q} \pi_1^{v_1-1} \cdots \pi_n^{v_n-1} \\
&\propto \sum_{b\geq 0} \frac{\lambda^b}{b!} \left(\sum_{b_{t+1}+\cdots+b_n=b} \frac{b!\pi_{t+1}^{b_{t+1}} \dots \pi_n^{b_n}}{b_{t+1}! \cdots b_n!} \right) \lambda^{b(t)+r-1} e^{-\lambda(1+p/q)} \\
&\qquad\qquad \times \pi_1^{b_1+v_1-1} \cdots \pi_t^{b_t+v_t-1} \pi_{t+1}^{v_{t+1}-1} \dots \pi_n^{v_n-1} \\
&= \left(\sum_{b\geq 0} \frac{\lambda^b \pi'(t)^b}{b!} \right) \lambda^{b(t)+r-1} e^{-\lambda(1+p/q)} \pi_1^{b_1+v_1-1} \cdots \pi_t^{b_t+v_t-1} \pi_{t+1}^{v_{t+1}-1} \dots \pi_n^{v_n-1} \\
&= \lambda^{b(t)+r-1} e^{-\lambda(\pi(t)+p/q)} \pi_1^{b_1+v_1-1} \cdots \pi_t^{b_t+v_t-1} \pi_{t+1}^{v_{t+1}-1} \dots \pi_n^{v_n-1}.
\end{aligned}
$$

To evaluate the constant κ use the exact form of the conditional and unconditional marginal distributions given in Proposition 2. ■

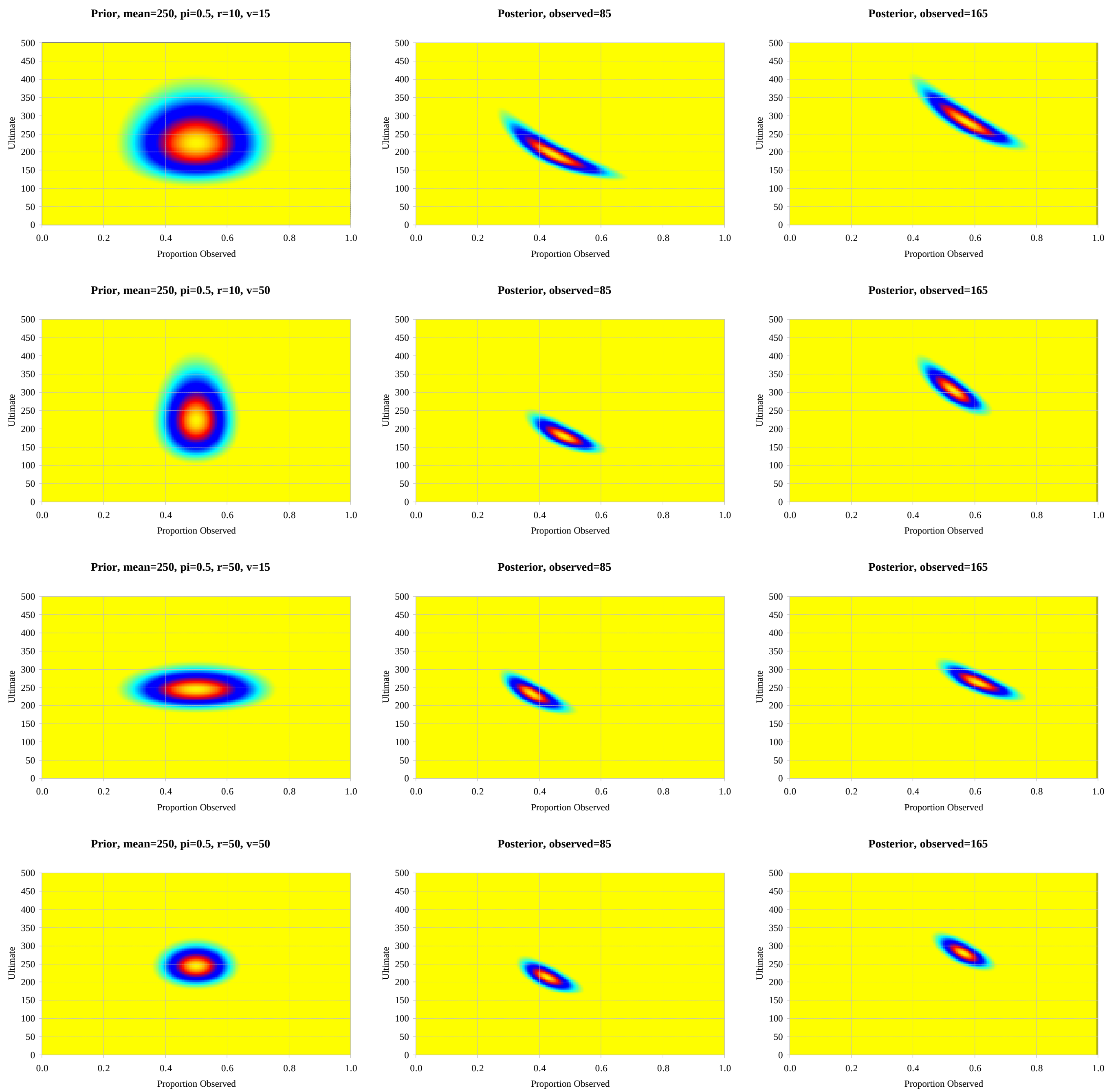

Figure 1: Prior and posterior density of Π_1 vs $\Lambda = \mathrm{E}(B(n))$ for various values of r and v and observed counts. Prior mean equals 250 and $\mathrm{E}(\Pi_1) = 0.5$. Left hand column shows prior density. Middle column shows posterior given observed counts 40 below expected; right hand column posterior given counts 40 above expected.

Figure 1 is a contour plot of the prior and posterior distribution of (Π_1, Λ). The left hand column shows the prior distributions with prior mean 250 and $\mathrm{E}(\Pi_1) = 0.5$. The middle column shows the posterior given an observation 40 below expected and the right hand column the posterior given an observation 40 above expected. The four rows show different degrees of precision in the priors.

Row 1. $r = 10$ and $v = 15$, so both priors have a moderately high uncertainty. Since $r < v$ the model gives weight to the chain-ladder method, so the posterior distributions lie north-east to south-west. Both are still relatively diffuse, reflecting the lack of information in the priors. The correlation between Π_1 and Λ in the posterior densities is very clear.

Row 2. $r = 10$ and $v = 50$, so the emergence is known with more prior certainty than the ultimate. The prior is now stretched along the y-axis, ultimate claims. Since emergence is known more precisely, this method is closer to the chain-ladder method (100% confidence in observed losses). In the picture we see the two posterior distributions lie north-east to south-west, corresponding to the chain-ladder method

Row 3. $r = 50$ and $v = 15$, so the prior ultimate is known with more certainty than the emergence. Now the prior is stretched along the x-axis, emergence. This method is closer to the peg method. The two posterior distribution lie east-west, corresponding to the less weight given to the observed claim information.

Row 4. $r = v = 50$, so both ultimate and emergence are known with more confidence. Compared to row 1 the prior is far more concentrated. Since $r = v$ this method reproduces the Borhuetter-Ferguson—see below.

In the left hand column Λ and Π_1 are uncorrelated in all four examples.

The next corollary computes the exact Bayesian reserve: the expected number of unreported claims given claims to date. It is an important result and we will discuss it further in Section 4. The corollary assumes $n = 2$ and $t = 1$; using Lemma 1 we can reduce any particular reserving problem to this csae.

Corollary 1 *Let* $n = 2$ *and* $t = 1$. *Then*

$$\mathrm{E}(B_2 \mid B_1 = b_1) = q\frac{v_2(b_1+r)}{b_1+v}\frac{{}_2F_1(v_2+1, b_1+r+1; b_1+v+1; q)}{{}_2F_1(v_2, b_1+r; b_1+v; q)}. \tag{30}$$

Proof: By definition

$$\mathrm{E}(B_2 \mid B_1 = b_1) = \mathrm{E}((1-\Pi_1)\Lambda \mid (\Theta \mid B_1)). \tag{31}$$

Now use the explicit form of the posterior distribution of $(\Theta \mid B_1)$ given in the proposition and integrate with respect to λ to get

$$\mathrm{E}(B_2 \mid B_1 = b_1) = \kappa \int_0^1 \int_0^\infty \lambda^{b_1+r} e^{-\lambda(\pi_i+p/q)} \pi_1^{b_1+v_1-1}(1-\pi_1)^{v_2} d\lambda d\pi_1 \tag{32}$$

$$= \kappa \int_0^1 \frac{\Gamma(b_1+r+1)}{(\pi_1+p/q)^{b_1+r+1}} \lambda^{b_1+r} e^{-\lambda(\pi_i+p/q)} \pi_1^{b_1+v_1-1}(1-\pi_1)^{v_2} d\pi_1. \tag{33}$$

Substitute $w = 1 - \pi_1$ and re-arrange to get

$$\kappa q^{b_1+r+1}\Gamma(b_1+r+1)\int_0^1 w^{v_2}(1-w)^{b_1+v_1-1}(1-qw)^{-(b_1+r+1)}dw. \tag{34}$$

The result follows from Euler's integral representation of hypergeometric functions Eqn. (65). ■

We can write Eqn. (30) as

$$(b_1+r)\left(\frac{{}_2F_1(v_2, b_1+r+1; b_1+v; q)}{{}_2F_1(v_2, b_1+r; b_1+v; q)} - 1\right) \tag{35}$$

using Whittaker and Watson [29, Chapter 14, Ex. 1]. Since b_1 is claims observed to date, the Bayesian expected ultimate is

$$\mathrm{E}(\Lambda \mid B_1 = b_1) = b_1 f + r(f-1) \tag{36}$$

where f is the ratio of hypergeometric functions. Thus f is acting like a loss development factor, but one which is a function of b_1. It is interesting that the Bayesian estimate does not go through the origin because of the constant r term.

Using the same approach we can compute all moments of the posterior distribution.

Corollary 2 *Let $n = 2, t = 1$ and let a, b be non-negative integers. Then*

$$\mathrm{E}(\Lambda^a \Pi_1^b) = q^a \frac{(b_1+r)_a (b_1+v_1)_b}{(b_1+v)_b} \frac{{}_2F_1(v_2, b_1+r+a; b_1+v+b; q)}{{}_2F_1(v_2, b_1+r; b_1+v; q)}. \tag{37}$$

5 Predictive Distributions

The next proposition gives an expression for the predictive distribution

$$(B_{t+1}, \dots, B_n \mid B_1, \dots B_t).$$

Remember that $b(t) = \sum_{i=1}^t b_i$, $b'(t) = \sum_{i=t+1}^n b_i$ and $v = \sum_{i=1}^n v_i$.

Proposition 4 *Let $B_1, \dots, B_n \mid \Theta$ have a GPDM distribution and let $1 \le t \le n-1$. Then the conditional distribution of $(B_{t+1}, \dots, B_n)$ given $B_1, \dots, B_t$ is*

$$q^{b'(t)} \frac{(b(t)+r)_{b'(t)}}{(b(t)+v)_{b'(t)}} \prod_{i=t+1}^n \frac{(v_i)_{b_i}}{b_i!} \, {}_2F_1(v'(t), b(t)+r; b(t)+v; q)^{-1}. \tag{38}$$

Proof: Recall that

$$\begin{aligned}
&\Pr(B_{t+1},\ldots,B_n \mid B_1,\ldots B_t) \\
&\qquad = \int \Pr(B_{t+1},\ldots,B_n \mid B_1,\ldots B_t,\Theta) f(\Theta \mid B_1,\ldots B_t) d\Theta \\
&\qquad = \int \frac{\Pr(B_1,\ldots,B_n \mid \Theta)}{\Pr(B_1,\ldots,B_t \mid \Theta)} \frac{\Pr(B_1,\ldots,B_t \mid \Theta) f(\Theta)}{\Pr(B_1,\ldots,B_t)} d\Theta \\
&\qquad = \frac{\Pr(B_1,\ldots,B_n)}{\Pr(B_1,\ldots,B_t)}
\end{aligned}$$

Combine this with Proposition 2 and the definition of the GPDM and then cancel to complete the proof. ∎

Proposition 4 shows the predictive distribution does not depend on the individual observed values $b_1,\ldots,b_t$ but only on their sum $b(t) = b_1 + \cdots b_t$. Thus the GPDM model has a kind of Markov property that the future development depends only on the total number of claims observed to date, and not on how those claims were reported over time.

Considering the probability distribution of the sum $B_{t+1} + \cdots + B_n$ given $B_1 + \cdots + B_t$ gives us the following corollary which we shall need later. This corollary can also be proved using induction and properties of the binomial coefficients.

Corollary 3

$$\sum_{b_1+\cdots+b_n=b} \left(\prod_{i=1}^{n} \frac{(v_i)_{b_i}}{b_i!} \right) = \frac{(v_1 + \cdots + v_n)_b}{b!} \tag{39}$$

Using Lemma 1 we can add $B_{t+1},\ldots,B_n$ and reduce to the case $n = 2, t = 1$. Then Eqn. (38) gives the conditional distribution of unreported claims B_2 given claims reported to date b. This provides a closed form expression for the posterior distribution which is exactly the distribution required for claim count reserving.

Corollary 4

$$\Pr(B_2 = b_2 \mid B_1 = b) = q^{b_2} \frac{(b+r)_{b_2}}{(b+v)_{b_2}} \frac{(v_2)_{b_2}}{b_2!} \; {}_2F_1(v_2, b+r; b+v; q)^{-1}. \tag{40}$$

The probabilities $\Pr(B_2 = j \mid B_1 = b)$ can be computed recursively using

$$\Pr(B_2 = j+1 \mid B_1 = b) = \Pr(B_2 = j \mid B_1 = b) \frac{q}{j+1} \; \frac{(r+b+j)(v_2+j)}{(v+b+j)} \tag{41}$$

for $j \geq 0$ and

$$\Pr(B_2 = 0 \mid B_1 = b) = \; {}_2F_1(v_2, b+r; b+v; q)^{-1}. \tag{42}$$

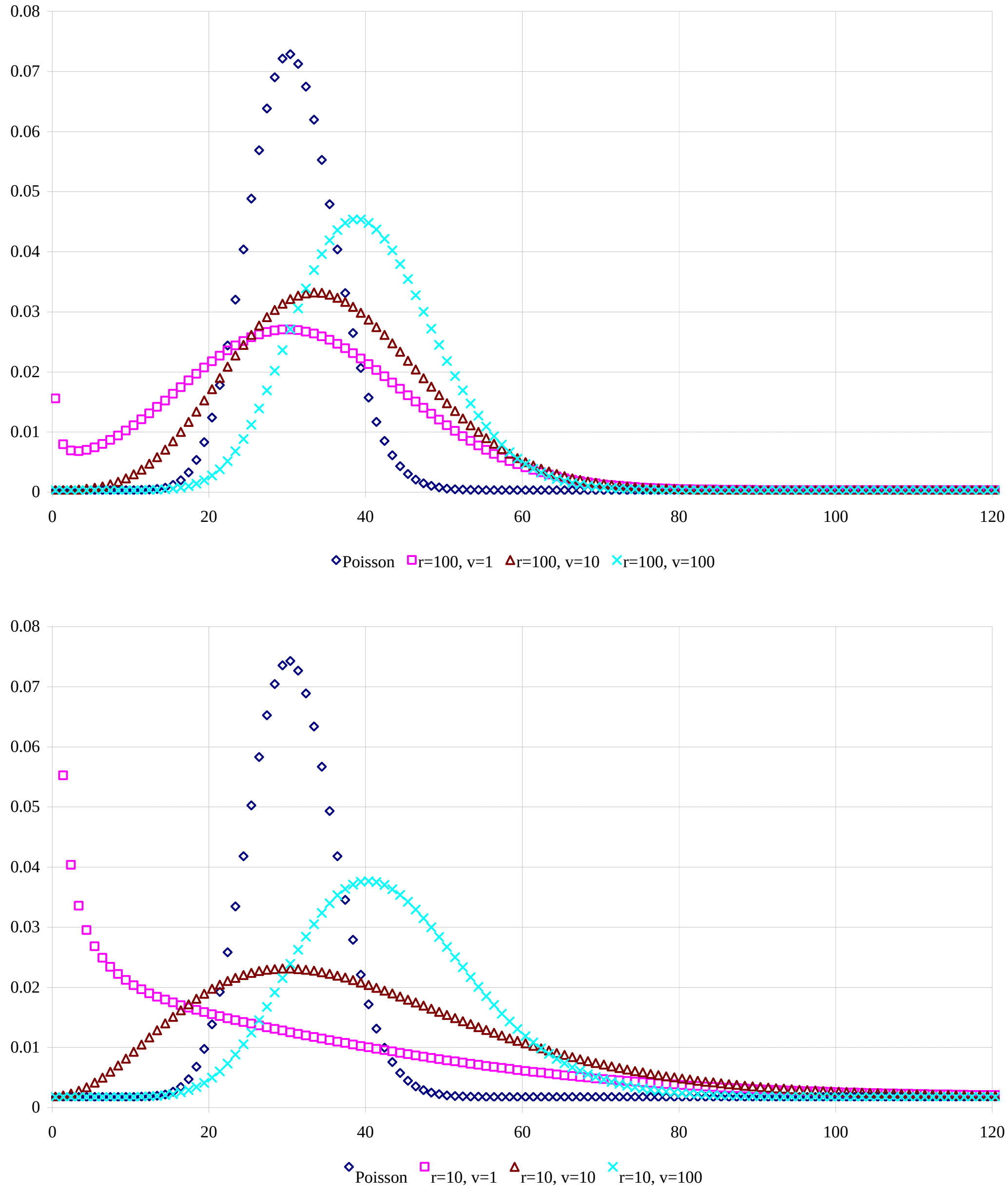

Figure 2: $(B_2 \mid B_1)$ for various values of r and v. $n = 100$, $b = 65$, and $v_1/v = 0.6$.

Figure 2 shows six examples of the density $B_2 \mid B_1$ for various values of v and r. They are the two key shape parameters. For comparison, each plot also has a Poisson with the same mean 30.305 as the $r = 100, v = 1$ frequency.

It follows from Eqn. (40) that the probability generating function[2] of $B_2 \mid B_1 = b$ is

$$G(z) = \frac{{}_2F_1(v_2, b+r; b+v; zq)}{{}_2F_1(v_2, b+r; b+v; q)}. \tag{45}$$

Therefore $B_2 \mid B_1 = b$ is a generalized power series distribution and a generalized hypergeometric probability distribution according to the classification in Johnson et al. [9]. It does not, however, appear in Table 2.4 of [9].

Differentiating G, using Equations 63 and 64 for the derivatives of the hypergeometric function, gives the factorial moments of $B_2 \mid B_1 = b$:

$$\mathrm{E}(B_2 \mid B_1 = b) = \frac{qv_2(b+r)}{b+v} \;\; \frac{{}_2F_1(v_2+1, b+r+1; b+v+1; q)}{{}_2F_1(v_2, b+r; b+v; q)}, \tag{46}$$

which reproduces Corollary 1, and more generally

$$\mu_{(k)}(B_2 \mid B_1 = b) = \frac{q^k (v_2)_k (b+r)_k}{(b+v)_k} \;\; \frac{{}_2F_1(v_2+k, b+r+k; b+v+k; q)}{{}_2F_1(v_2, b+r; b+v; q)}. \tag{47}$$

---

[2]The probability generating function of a nonnegative discrete random variable X is defined as

$$G(z) = \mathrm{E}(z^X).$$

The (descending) kth factorial moment of a random variable X is defined as

$$\mu_{(k)}(X) = E(X(X-1)\cdots(X-k+1)).$$

Factorial moments can be computed from the probability generating function by differentiating:

$$\mu_{(k)} = \frac{d^k G(z)}{dz^k}\Big|_{z=1}. \tag{43}$$

It is easy to compute the central moments and moments about zero from the factorial moments. For example

$$\mathrm{Var}(X) = \mu_{(2)} + \mu - \mu^2. \tag{44}$$

See Johnson et al. [8] for more general relationships.

6 The Continuum of Reserving Methods

Corollary 1 is very important. It provides a Bayesian estimate of unreported claims given claims to date which is exactly the quantity the reserving actuary must estimate. In this section we show that special or limiting cases of the GPDM include the peg, the Borhuetter-Ferguson, Benktander, and the chain-ladder methods, and then compare the GPDM to traditional methods over practical ranges of the parameters r and v. The model confirms the suggestion in Renshaw and Verrall [23] that the chain-ladder is just one of many appropriate methods. A schematic showing how the GPDM interpolates between other reserving methods is shown in Figure 3.

Using Lemma 1 we can reduce each accident year to the case $n = 2$, $t = 1$. B_1 denotes observed claims and B_2 unreported claims. $(B_1, B_2 \mid \Theta)$ has a GPDM distribution; $\Theta = (\Lambda, \Pi_1)$, Λ has a gamma distribution with mean m and variance $m(1 + m/r)$, and Π_1 has a beta distribution with parameters v_1 and v_2. Let $v = v_1 + v_2$ and $\pi := \mathrm{E}(\Pi_1) = v_1/v$. Per Section 2, the parameters of the gamma distribution are r and p/q where $p = r/(r + m)$ and $q = 1 - p = m/(r + m)$. Higher values of r and v correspond to lower variances of Λ and Π_1 respectively. As $r \to \infty$ the claim count distribution tends to a Poisson.

We are going to compare the following six estimates of unreported claims. The estimate of ultimate claims corresponding to each method is simply $b + b'_*(b)$.

1. The GPDM method

$$b'_g(b) = \mathrm{E}(B_2 \mid B_1 = b) = (b + r)\left(\frac{{}_2F_1(v_2, b + r + 1; b + v; q)}{{}_2F_1(v_2, b + r; b + v; q)} - 1\right). \tag{48}$$

2. The peg method

$$b'_p(b) := (m - b)^+. \tag{49}$$

The peg ultimate is insensitive to observed data—until observed claims exceed the peg! The peg is an extreme reserving method. It ignores actual emergence completely.

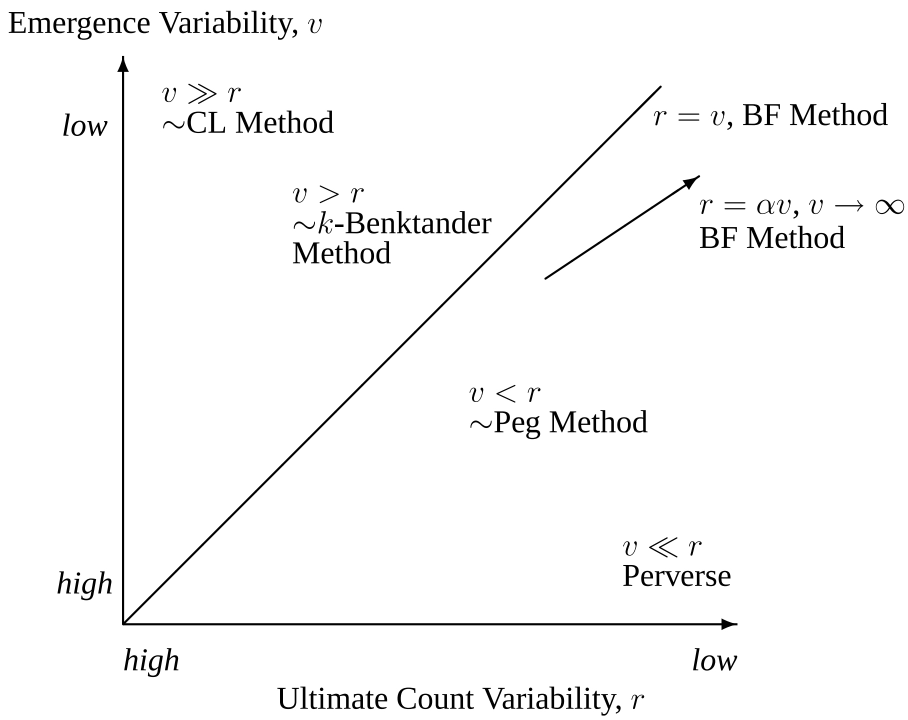

Figure 3: Schematic showing the behaviour of the GPDM reserve as (r, v) vary. Low r (resp. v) corresponds to high uncertainty in ultimate counts (resp. claim emergence). The $r = v$ diagonal is exactly the Borhuetter-Ferguson method.

3. The chain-ladder method

$$b'_c(b) := \frac{(1-\pi)b}{\pi} \tag{50}$$

see Mack [16] or Renshaw and Verrall [23]. π is usually estimated from the data as a product of link ratios. Each link ratio is the weighted average development from one period to the next over all available accident periods. The chain-ladder method is at the opposite extreme to the peg method. It completely ignores prior estimates of ultimate counts.

4. The Borhuetter-Ferguson method estimate

$$b'_b(b) := m(1 - \pi) \tag{51}$$

see Mack [16]. This estimate of unreported claims is completely insensitive to the observation b. The Borhuetter-Ferguson method is sometimes regarded as an extreme, but it is actually a middle-ground method between the chain-ladder and peg methods.

5. The k-Benktander method, $k = 0, 1, 2, \ldots$

$$b'_k(b) := (1 - (1 - \pi)^k) b'_c(b) + (1 - \pi)^k b'_b(b) \tag{52}$$

see Mack [16]. When $k = 0$ this reduces to the Borhuetter-Ferguson. As $k \to \infty$, $b'_k(b) \to b'_c(b)$. The Benktander methods are all linear in b. They are a credibility weighting of the Borhuetter-Ferguson and chain-ladder methods.

6. The linear least squares, or greatest accuracy credibility, estimate

$$b'_l(b) := \alpha + \beta b \tag{53}$$

where α and β are chosen to minimize the expected squared error. This approach is described in Klugman et al. [10, Section 5.4] from a credibility perspective and in Murphy [19] from a linear least squares loss development perspective. Solving by differentiating $\mathrm{E}((B_2 - \alpha - \beta B_1)^2)$ with respect to α and β and setting to zero gives

$$\alpha = \mathrm{E}(B_2) - \beta \mathrm{E}(B_1) \qquad \beta = \frac{\mathrm{Cov}(B_1, B_2)}{\mathrm{Var}(B_1)}. \tag{54}$$

In order to actually compute α and β we need a bivariate distribution for B_1 and B_2; we use the GPDM. The variance and covariance are computed in Eqn. (23) and Eqn. (24). By construction b'_l will be the least squares line through b'_g. When $r = v$, and B_1 and B_2 are uncorrelated, b'_l reduces to the Borhuetter-Ferguson.

Neither the chain-ladder nor the Borhuetter-Ferguson method is sensitive to the relative variance of ultimate losses $B_1 + B_2$ and the proportion of claims observed Π_1. This is a weakness that can be illustrated by considering two hypothetical situations. In the first, the ultimate is estimated with low confidence but the claim reporting pattern is very predictable, so $r < v$. We would favor the chain-ladder estimate over the prior m. This corresponds to the second row in Figure 1. In the second situation, the ultimate claim count distribution is known with a high confidence, but the reporting pattern is estimated with less confidence, so $r > v$. Here we would weigh the prior estimate m more than the chain-ladder which relies on the proportion reported. This corresponds to the third row in Figure 1. Corollary 1 provides a probabilistic model of these intuitions that continuously interpolates from one extreme to the other. The GPDM captures and models the process behind the actuarial judgment of selecting appropriate reserves. By providing a quantification of what is currently a judgmental process the model should be of great value to the practicing actuary.

Here are six examples of how the GPDM behaves for different values of r and v. They are illustrated in Figure 3.

1. For fixed r, $b_g'(b) \to m(1-\pi)(b+r)/(m\pi + r)$ as $v \to \infty$. *Proof:* As $v \to \infty$

$$\begin{aligned} {}_2F_1(v_2, b+r; b+v; q) &= {}_2F_1((1-\pi)v, b+r; b+v; q) \\ &\to {}_2F_1(1, b+r; 1; (1-\pi)q) \\ &= (1-(1-\pi)q)^{-(b+r)} \end{aligned}$$

by Eqn. (66). Therefore

$$b_g'(b) \to \frac{q(1-\pi)(b+r)}{1-(1-\pi)q} = \frac{m(1-\pi)(b+r)}{m\pi + r}. \tag{55}$$

We can write this limit as a credibility weighting of the chain-ladder and Borhuetter-Ferguson with credibility $z = m\pi/(m\pi + r)$ given to the chain ladder:

$$\frac{m(1-\pi)(b+r)}{m\pi + r} = \left(\frac{m\pi}{m\pi + r}\right)\frac{(1-\pi)b}{\pi} + \left(\frac{r}{m\pi + r}\right)((1-\pi)m). \quad (56)$$

This equation corresponds to the k-Benktander method with

$$k = \frac{\log(r/(m\pi + r))}{\log(1-\pi)}. \quad (57)$$

2. As $r \to 0$ and $v \to \infty$ the GPDM reserve tends to the chain ladder reserve, $b'_g(b) \to b'_c(b)$. ***Proof:*** This follows letting $r \to 0$ in Eqn. (55) since $q = m/(m+r) \to 1$.

3. If $r = v$ then the GPDM reserve equals the Borhuetter-Ferguson reserve, $b'_g(b) = b'_b(b)$. ***Proof:*** Applying Eqn. (66) to Eqn. (46) gives

$$b'_g(b) = \frac{qv_2}{1-q} = \frac{qv_2}{p}. \quad (58)$$

Since $r = v$, $q/p = m/v$ and so

$$b'_g(b) = \frac{mv_2}{v} = m(1-\pi) \quad (59)$$

as required. The case $r = v$ represents an exact balance between the uncertainty in ultimate losses and claim count emergence which reproduces the Borhuetter-Ferguson. By Eqn. (24) it also represents the case when B_1 and B_2 are uncorrelated.

4. If $r = \alpha v$ for a constant α then as $v \to \infty$ the GPDM reserve converges to the Borhuetter-Ferguson: $b'_g(b) \to b'_b(b)$.

5. For fixed small $v > 0$ the GPDM reserve is close to the peg reserve as $r \to \infty$. See the figures below.

6. As $v \to 0$ and $r \to \infty$ the GPDM reserve tends to zero if $b > 0$ and m if $b = 0$. This is a perverse kind of reserve! It is possible to prove this analytically, but heuristically the reason is that as $v \to 0$ the Dirichlet distribution becomes concentrated at the corners. Thus the claims all tend to be reported at once. So if any claims have been reported then no more are expected. On the other hand, if none have been reported they should all still be held in reserve.

The fourth point needs elaborating because it appears to contradict the main result of Renshaw and Verrall [23]. Their model assumes incremental claims B_i have a Poisson distribution and hence emergence is modeled with a multinomial distribution. They prove their model reproduces the chain-ladder reserve when the parameters are determined using maximum likelihood. As $r \to \infty$ the GPDM ultimate has a Poisson distribution. As $v \to \infty$ the Dirichlet prior becomes a degenerate distribution, so the DM becomes a multinomial distribution conditional on ultimate counts $B(n)$. In this situation B_i will also have a Poisson distribution and so in the limit the GPDM model appears to be the same as Renshaw and Verrall's, and yet it gives to the Borhuetter-Ferguson reserve and not the chain-ladder. The reconciliation of this apparent contradiction is that Renshaw and Verrall fit the emergence pattern (means of the multinomial) and the prior accident year means *a posteriori* from the data. If we interpret these parameters as prior estimates then their model produces exactly expected emergence—see [23, Eqn. (2.4)]. In the GPDM model the emergence pattern and accident year means are given *a priori*. As $v, r \to \infty$ both parameters become certain. If losses emerge exactly as expected then the chain-ladder and Borhuetter-Ferguson methods agree and so the GPDM would also give the chain-ladder reserve. However, actual emergence from the GPDM need not be exactly equal to expected because the means and emergence are specified *a priori*. Note that in the Poisson-multinomial model

$(r, v \to \infty)$ B_1 and B_2 are independent so the linear least squares method also reproduces the Borhuetter-Ferguson.

These mathematical limits of the GPDM method are mainly of academic interest. However, the way the GPDM interpolates between the common reserving methods for realistic values of r and v is of practical interest because it provides analytical guidance to supplement actuarial judgment. We now explore that interpolation.

Figure 4 shows a plot of $b'_g(b)$ against b for $r = 25$, $\pi = 0.45$, $m = 110$ and $v = 0.1, 1, 10, 25, 100, 1000$. Each plot also shows the peg, chain-ladder, Borhuetter-Ferguson, k-Benktander and linear least squares reserves. The value of k is determined by Eqn. (57). Figure 4 ties back to the six points we made about b'_g.

- The four standard methods do not change with v. The linear least squares method is sensitive to v and is a line through b'_g as expected.
- Point 1 is illustrated by $v = 100$ and $v = 1000$. The PGDM method tends to the predicted k-Benktander method line for larger k. If we had plotted large v and small r the PGDM line would eventually convert up to the chain-ladder line, per Point 2.
- Point 3 is illustrated by $v = 25 = r$; the PGDM line lies underneath the Borhuetter-Ferguson line.
- The fact that the PGDM favors the peg method when $v < r$ and the chain-ladder method when $v > r$ is shown in the increasing slope of the PGDM line from the first plot to the last.
- Point 5 is illustrated by $v = 1$: b'_g is close to the peg method.
- Point 6 is illustrated by $v = 0.1$ which shows $b'_g \to 0$ for larger b.

Figures 5 and 6 are two views of the bivariate density of (B_1, B_2) computed with $m = 110$ claims, $r = 25$ and $\pi = 0.45$, so $\mathrm{E}(B_1) = 49.5$ and $\mathrm{E}(B_2) = 109.5$. The nine contour plots correspond to $v = 0.1, 1, 2.5, 5, 10$, 25, 100, 1000, 10000. As expected, when $v < 25$ B_1 and B_2 are negatively correlated. When $v = 25$ they are uncorrelated and when $v > 25$ they are positively correlated. The posterior distribution of $B_2 \mid B_1 = b_1$ is simply a re-scaled vertical slice through these distributions, so the reader should be able to connect these plots with the plots of $b'_g(b)$. The cases $v = 0.1, 1$ help explain how the GPDM reacts given extreme uncertainty in the payout pattern. The 3-d plot explains why the contour plot seems to disappear: the probability becomes concentrated along the axes.

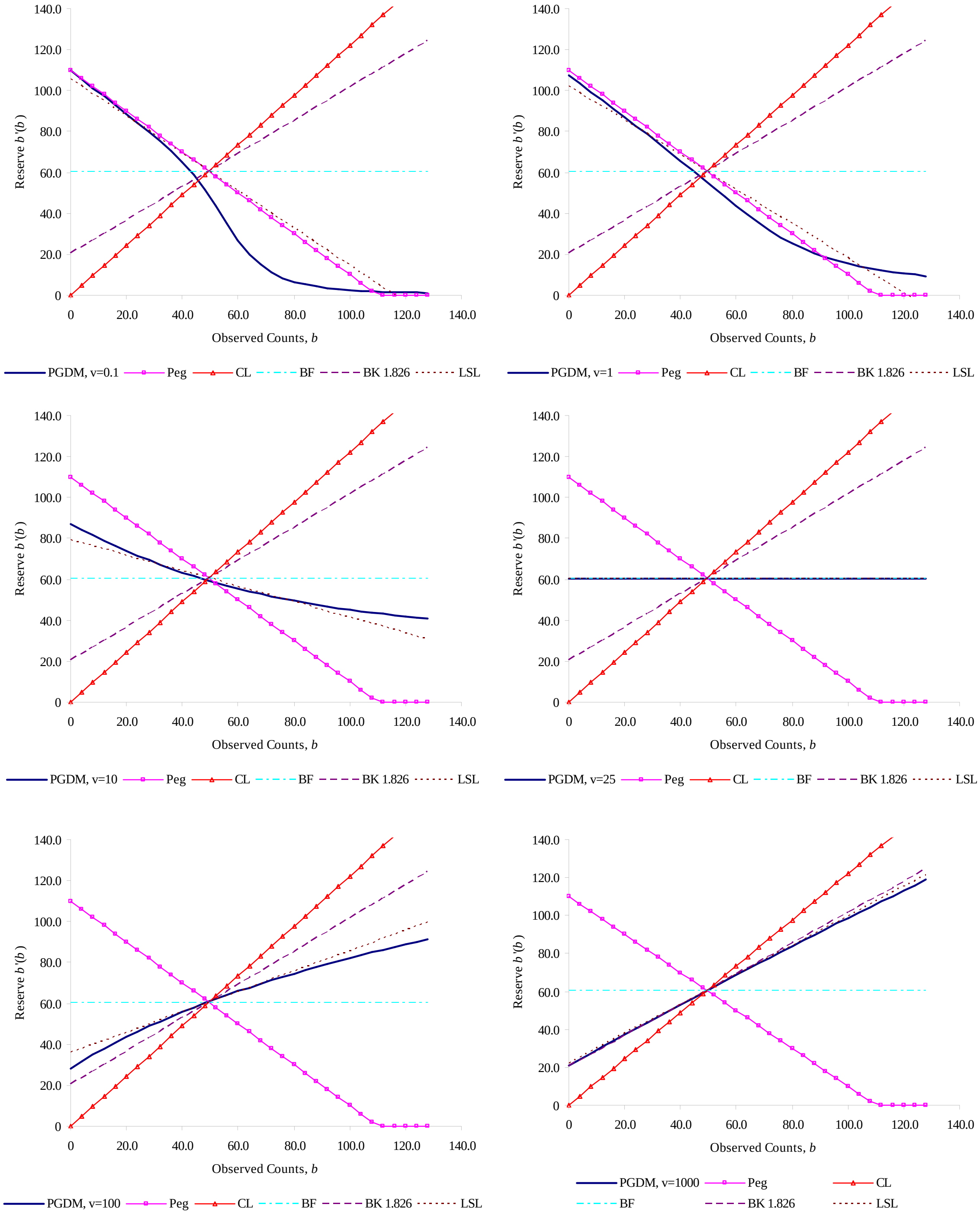

Figure 4: $b'_p(b)$ for $r = 25$, $\pi = 0.45$, $\mathrm{E}(\Lambda) = 110$ and $v = 0.1, 1, 10, 25, 100, 1000$, compared with the peg, chain-ladder, Borhuetter-Ferguson, Benktander k and linear least squares methods. $k = 1.826$ determined by Eqn. (57).

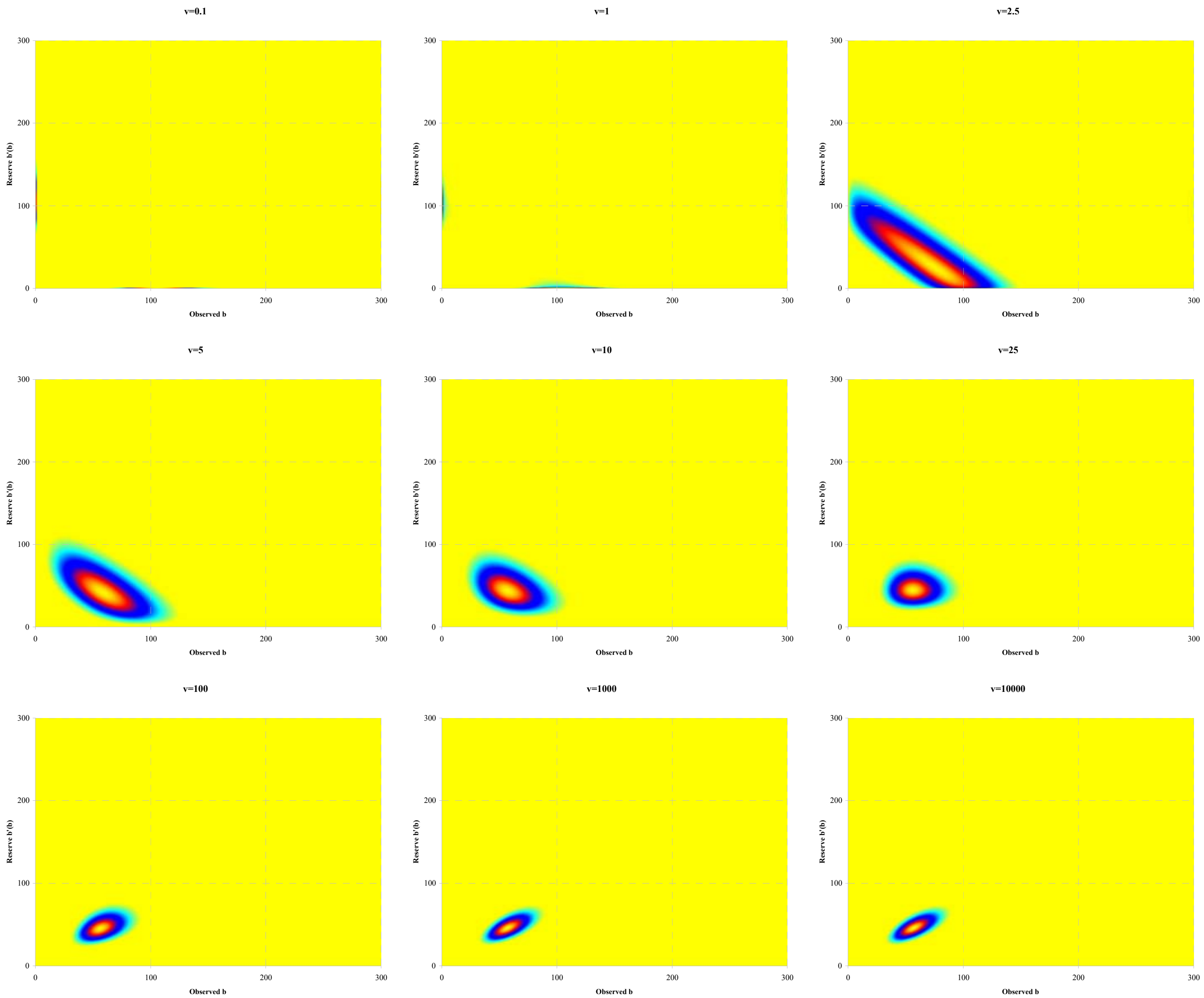

Figure 5: Contour plots of the bivariate density (B_1, B_2) with $r = 25$, $m = 110$, $\pi = 0.45$ shown for $v = 0.1, 1, 2.5, 5, 10, 25, 100, 1000, 10000$.

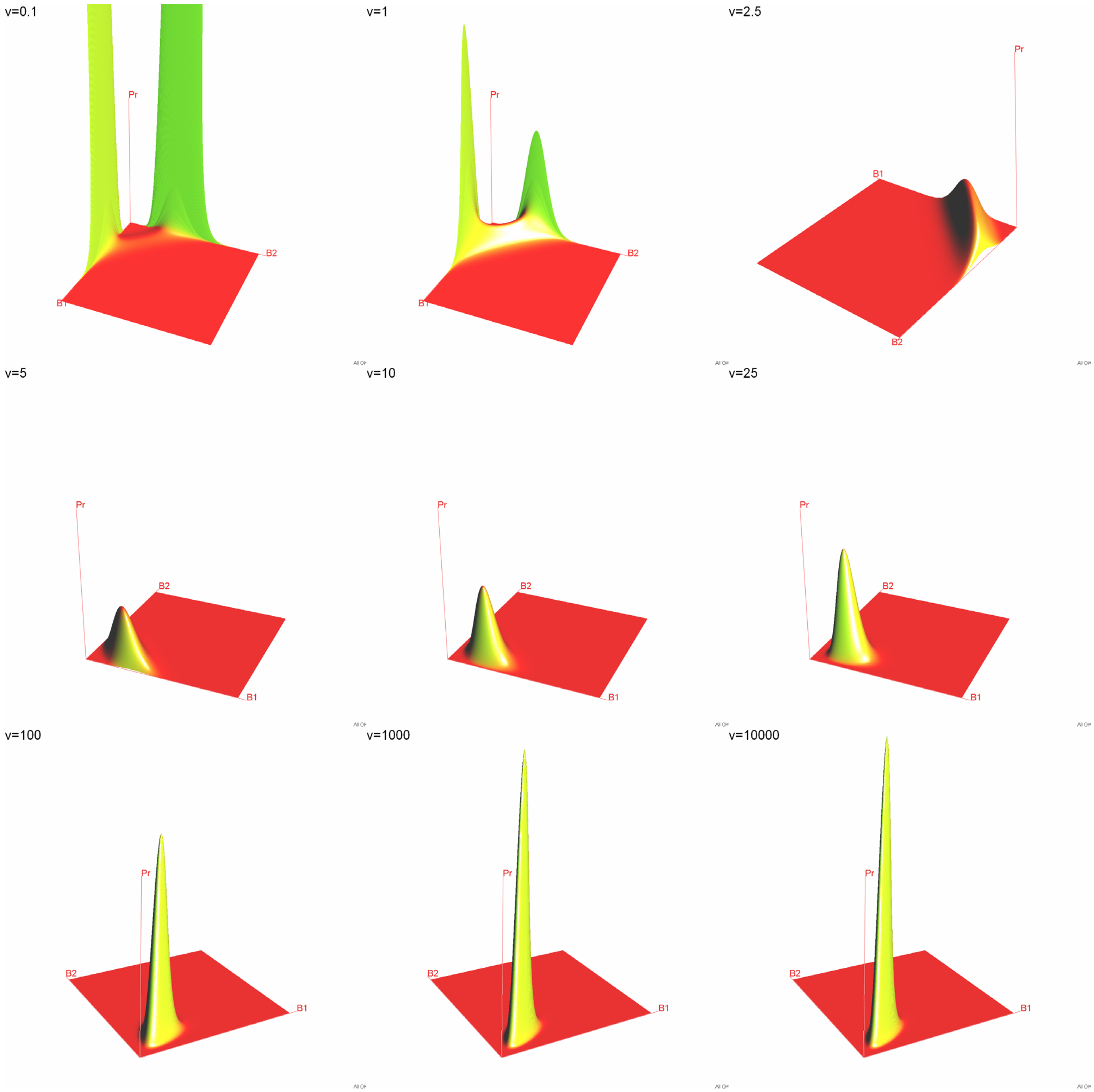

Figure 6: Three dimensional density plots of the bivariate density (B_1, B_2) with $r = 25$, $m = 110$, $\pi = 0.45$ shown for $v = 0.1, 1, 2.5, 5, 10, 25, 100, 1000, 10000$. The z-scales are all the same. The orientations vary by plot.

This completes our theoretical investigation of the properties of the GPDM distribution. We have produced easy-to-compute expressions for the marginal and conditional distributions and written down the mean of the posterior distribution of unreported claims given claims observed to date. Next we show how the GPDM can be used in practice by applying it to a particular claim count development triangle.

7 Parameter Estimation

The GPDM model for n periods of development uses $n+2$ parameters; of these the n development-related parameters $v_1, \ldots, v_n$ would usually be shared across multiple accident years. The prior mean m of the ultimate distribution would vary by accident year and r would generally be considered common. Thus to model a development triangle with n accident years and development periods there are $2n+1$ parameters. If there is a good exposure measure then prior mean could be modeled as a common frequency times exposure and that would reduce the number of parameters to $n+2$.

Reasonable initial estimates for m should be available from the pricing department. A view of r could be driven by a macro line-of-business level study. Alternatively we could take r to be very small corresponding to a non-informative prior for the ultimate.

Kotz et al. [11] discuss using sample moments to estimate the parameters v_t of a Dirichlet-multinomial. Let M'_{1t} be the sample mean of the proportion of claims observed in the tth period (computed with respect to the chain-ladder, for example), and let M'_{21} be the mean of the square of the proportion of claims observed in the first period. Then reasonable starting parameters are

$$\widehat{v}_t = \frac{(M'_{11} - M'_{21})M'_{1t}}{M'_{21} - (M'_{11})^2}, \quad i = 1, \ldots, n-1 \tag{60}$$

$$\widehat{v}_n = \frac{(M'_{11} - M'_{21})(1 - \sum_{t=1}^{n-1} M'_{1t})}{M'_{21} - (M'_{11})^2}. \tag{61}$$

Alternatively taking $v_1 = \cdots = v_n = 1$ gives a prior emergence distribution equal over all periods, which could be regarded as a non-informative prior.

Table 1: Incremental Claim Count Data

Year	1	2	3	4	5	6	7	8	9	10	b
1990	40	124	157	93	141	22	14	10	3	2	606
1991	37	186	130	239	61	26	23	6	6		714
1992	35	158	243	153	48	26	14	5			682
1993	41	155	218	100	67	17	6				604
1994	30	187	166	120	55	13					571
1995	33	121	204	87	37						482
1996	32	115	146	103							396
1997	43	111	83								237
1998	17	92									109
1999	22										22

8 Example

We now give an example to illustrate the use of the GPDM to estimate the distribution of unreported claims. Further examples and a more detailed discussion of applying this method in practice is given in Mildenhall [18].

The incremental claim count data is shown in Table 1 and the claim count development factors are shown in Table 2. The right hand column shows total counts observed to date b. This data was analyzed by de Alba [5]. Using a Bayesian model he found a mean of 919 outstanding claims with a standard deviation of 79.51.

We use Eqn. (20) to compute the likelihood of each row of the development triangle and then determine the maximum likelihood estimate parameters. Initial parameter values were $r = 25$, the chain ladder estimates for the prior means m_i by accident year, and the estimates of v_t given in the previous section. The starting values and maximum likelihood estimates for m_i are shown in Table 3. Table 4 shows the same thing for v_t along with the incremental reporting patterns for both estimates. The maximum likelihood estimator for r is 1625458.8 which is much closer to Poisson than the starting value and $v = 129.018$ so the model has $r > v$.

Table 2: Loss Development Factors

AY	1 : 2	2 : 3	3 : 4	4 : 5	5 : 6	6 : 7	7 : 8	8 : 9	9 : 10
1990	4.100	1.957	1.290	1.341	1.040	1.024	1.017	1.005	1.003
1991	6.027	1.583	1.677	1.103	1.040	1.034	1.009	1.008	
1992	5.514	2.259	1.351	1.081	1.041	1.021	1.007		
1993	4.780	2.112	1.242	1.130	1.029	1.010			
1994	7.233	1.765	1.313	1.109	1.023				
1995	4.667	2.325	1.243	1.083					
1996	4.594	1.993	1.352						
1997	3.581	1.539							
1998	6.412								
Wtd. Avg.	5.055	1.930	1.350	1.134	1.035	1.023	1.011	1.007	1.003

Clearly the development pattern for this triangle is quite erratic, and so a low v is expected. One reason that r is so large is the use of a different variable m_i for each accident year. These parameters absorb some of the claim count variability and increase r.[3]

Table 5 shows the GPDM, chain ladder and Borhuetter-Ferguson reserves, and the standard deviation and coefficient of variation of the GPDM reserve. The overall reserve is slightly lower than the chain ladder. It is interesting that the reserves are actually higher for the older years and lower for the more recent years.

Figure 7 shows the distribution of the GPDM reserve. This distribution is the sum of the reserve distributions for each accident year, assuming they are independent. Figure 8 shows the evolution of the predictive distribution of ultimate claims for the oldest accident year, as more and more claim information becomes available.

[3] Exposure information was not available for this triangle, but estimating an exposure base produced a modeled $\hat{r} = 196.1$, lowered the reserve to 889 from 895 and increased the standard deviation of the reserve to 55.4 from 40.5. The estimate of v declined slightly.

Table 3: Starting and maximum likelihood estimates of m for each accident year.

Year	CL Mean	Prior Mean m
1990	606.0	606.0
1991	716.4	718.2
1992	689.0	692.5
1993	616.7	621.6
1994	596.2	601.8
1995	520.8	527.1
1996	485.1	487.9
1997	391.9	390.0
1998	347.9	339.8
1999	355.0	333.0

Table 4: Starting and Maximum Likelihood Estimates for v_i with implied incremental and cumulative proportion of claims reported

t	1	2	3	4	5	
Initial v_t	13.195	52.531	60.500	43.094	22.834	
Incremental	0.064	0.253	0.292	0.208	0.110	
Cumulative	0.064	0.317	0.608	0.816	0.926	
MLE v_t	8.477	32.702	36.891	26.322	13.367	
Incremental	0.066	0.253	0.286	0.204	0.104	
Cumulative	0.066	0.319	0.605	0.809	0.913	
t	6	7	8	9	10	$v = \sum v_i$
Initial v_t	6.636	4.428	2.225	1.384	0.686	207.514
Incremental	0.032	0.021	0.011	0.007	0.003	
Cumulative	0.958	0.979	0.990	0.997	1.000	
MLE v_t	4.488	3.010	1.729	1.246	0.786	129.018
Incremental	0.035	0.023	0.013	0.010	0.006	
Cumulative	0.948	0.971	0.984	0.994	1.000	

Table 5: Comparison of Reserve Estimates ($r = 1625458.8$, $LL = -221.42$)

Year	b	m	$b'_g(b)$	Std. Dev.	CV	$b'_c(b)$	$b'_b(b)$
1991	714	718	4	5.018	1.191	2	2
1992	682	693	11	7.671	0.728	7	7
1993	604	622	18	9.300	0.529	13	13
1994	571	602	31	11.688	0.380	25	25
1995	482	527	45	12.910	0.287	39	39
1996	396	488	92	15.965	0.174	89	90
1997	237	390	153	16.774	0.110	155	154
1998	109	340	231	17.344	0.075	239	233
1999	22	333	311	18.072	0.058	333	312
Total	4423		895	40.465	0.045	902	876

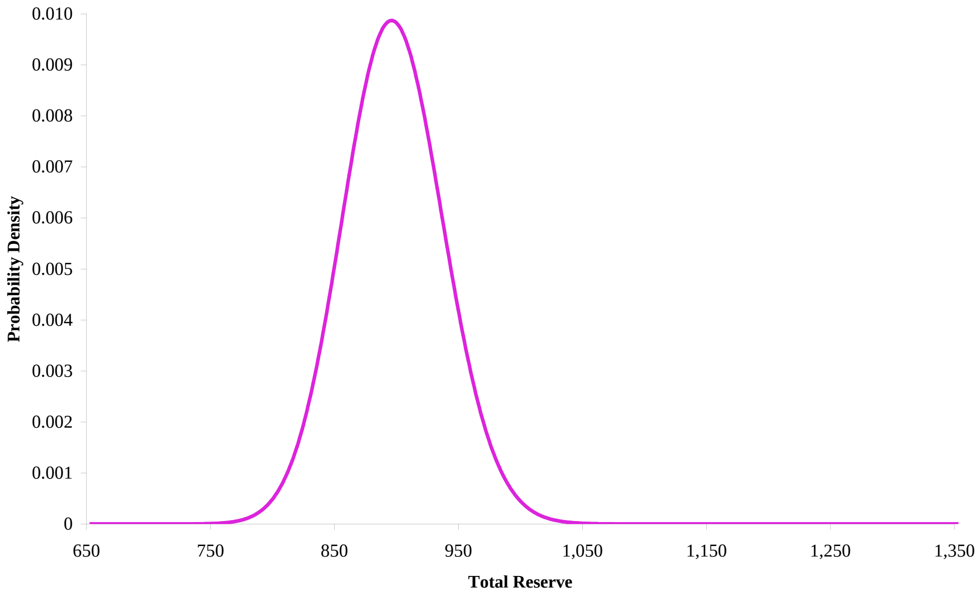

Figure 7: Distribution of total reserve.

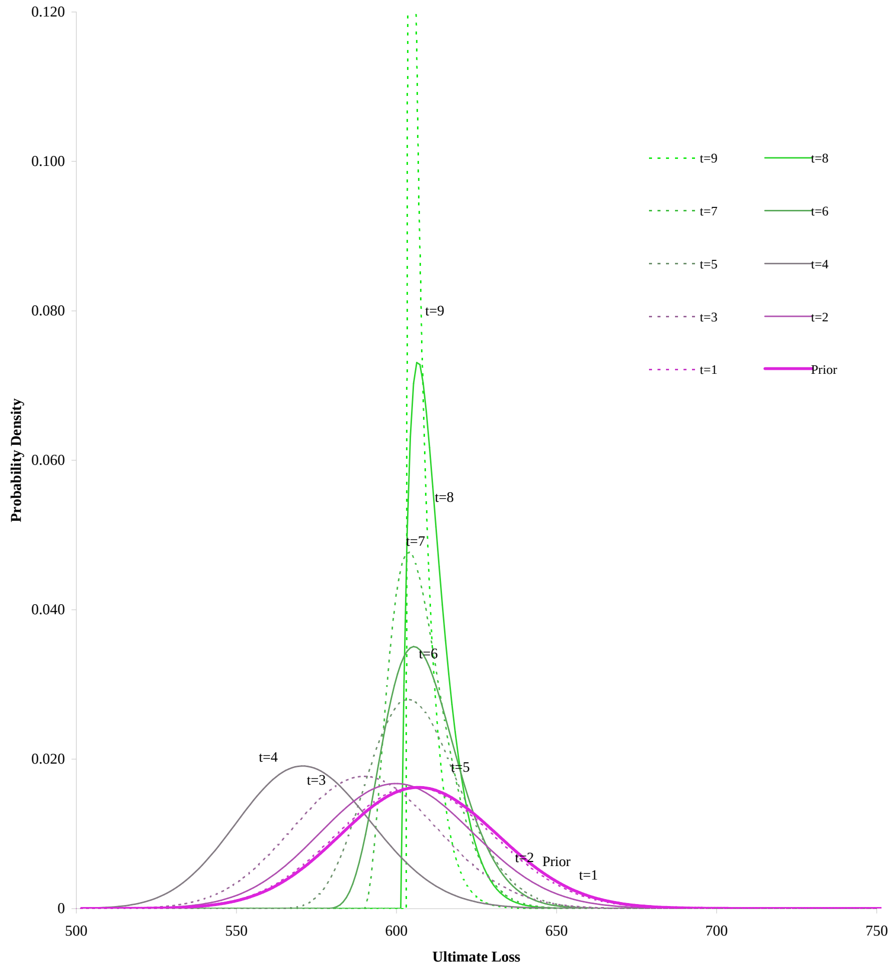

Figure 8: Predictive distribution of ultimate losses for oldest accident year starting with prior and adding observed losses for each development period.

9 Extension to Loss Development

The GPDM model applies to claim counts. Understanding claim counts can be a hard problem and the power of the model for working with counts should not be discounted. Nonetheless an extension to loss development is desirable. There will not be a similarly tractable model for loss development—just as there is no analog of the Poisson-gamma model for aggregate loss distributions. However, the general philosophy of the GPDM model, that the appropriate reserve depends on the relative variance of ultimate losses and loss emergence, carries over intact to losses and the problem is to determine a suitable bivariate distribution for observed and unobserved claims. Once that bivariate distribution is in hand numerical methods can be used to produce predictive reserve distributions. There are at least two approaches we could take.

Firstly, like Renshaw and Verrall [23], we can just use the GPDM directly to model losses. This is actually a more rational assumption than it seems. For a large book of business with a "tame" severity distribution (for example, where all policies have a low limit) the severity quickly diversifies and the normalized distribution of ultimate losses converges in distribution to the distribution of Λ as the book gets larger, see Daykin et al. [4, Appendix C] or Mildenhall [17, Section 2.10]. This method would be particularly appropriate when the maximum severity is of the same order of magnitude as the average severity because the diversification would occur more quickly. Working layer excess of loss reinsurance is an example.

The second approach is to try and determine a bivariate distribution for B_1 and B_2 and work with it numerically. Here we need a distribution of severity at tth report and ultimate. This could be estimated directly from a transactional loss database. The severity component would be combined with a mixed count emergence model like the GPDM. The aggregate distributions could be computed numerically using Fourier or fast Fourier transforms, or simulation. Alternatively, given the model specification and conditional severity distribution, we could use WinBUGs and MCMC techniques—see Verrall [28]. Understanding individual claim severity development is a great opportunity for further actuarial research in loss development. Since claim databases for most lines (except workers compensation) are much smaller than exposure databases this is also a practical thing to do.

10 Conclusions

We have introduced the GPDM model of claim count development and computed many of its important actuarial properties. The GPDM model incorporates estimates of the variability of ultimate claims and the claim emergence pattern into its estimates of reserves. Selecting between different reserve estimates is something usually done via actuarial judgment. The GPDM model can help bolster actuarial judgment by supplying a well-defined analytic selection framework.

The model includes the chain-ladder and Borhuetter-Ferguson methods as special cases, and also closely approximates the peg method and k-Benktander methods. Thus it provides a rich modeling framework for the practitioner.

The GPDM is a statistical model of claim development which can be fit using maximum likelihood. Given an exposure base, it can also be used to fit ultimates in the presence of covariates, again also using maximum likelihood. The model is easy to use and provides full posterior distributions rather than just a point estimate and standard deviation.

A Appendix: Hypergeometric Functions

The hypergeometric function ${}_2F_1$ is defined as

$$ {}_2F_1(a,b;c;q) = \sum_{k\geq 0} \frac{(a)_k(b)_k}{(c)_k k!} q^k. \tag{62} $$

The notation $(a, b; c; q)$ indicates there are two variables in the numerator, one in the denominator and one argument (there are generalizations the reader can readily imagine). The series is absolutely convergent for $|q| < 1$ and conditionally convergent for $|q| = 1$. In our applications q is real and $0 < q < 1$, so convergence is not an issue. Hypergeometric functions have been described as a staple of nineteenth century math; a glance at any table of mathematical equations will explain why. The facts we use are gathered from Abramowitz and Stegun [1, Chapter 15] and Lebedev [13].

The hypergeometric function is very easy to compute for $|q| < 1$. The following algorithm, taken from Press et al. [22], will compute ${}_2F_1(a, b; c; q)$ for $a > 0$, $b > 0$, $c > 0$ and $0 < q < 1$ to machine accuracy.

```
Initialize: f = 1,  g = 1,  i = 1
do
      g = g *  q * a * b / c / i
      f = f + g
      a = a + 1
      b = b + 1
      c = c + 1
      i = i + 1
while g > 0
return f
```

Because the series defining ${}_2F_1$ is absolutely convergent it can be differentiated term by term, giving

$$ \frac{dF}{dq} = \frac{ab}{c}\, {}_2F_1(a+1, b+1; c+1; q) \tag{63} $$

and more generally

$$\frac{d^nF}{dq^n} = \frac{(a)_n(b)_n}{(c)_n}\,{}_2F_1(a+n, b+n; c+n; q). \tag{64}$$

Euler's integral representation of ${}_2F_1$ is

$${}_2F_1(a, b; c; q) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tq)^{-a} dt \tag{65}$$

[1, Chapter 15.3]. We will use the result

$${}_2F_1(a, b; b; q) = {}_2F_1(b, a; b; q) = (1-q)^{-a} \tag{66}$$

from [1, Chapter 15.1]. This can be seen by considering the sum of the probabilities of a negative binomial distribution.

References

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions*. Dover, ninth edition, 1970.

[2] Vijay Balasubramanian. A geometric formulation of Occam's razor for inference of parametric distributions. *Princeton Preprints*, arxiv:adap-org/9601001 v1, 1996.

[3] José M. Bernardo and Adrian F. M. Smith. *Bayesian Statistical Decision Theory*. J. Wiley and Sons, 1994.

[4] C. D. Daykin, T. Pentikainen, and M. Pesonen. *Practical Risk Theory for Actuaries*. Chapman and Hall, 1994.

[5] Enrique De Alba. Bayesian estimation of outstanding claims reserves. *NAAJ*, 6 No. 4:1–20, 2002.

[6] Pedro Domingos. The role of Occam's razor in knowledge discovery. *Data Mining and Knowledge Discovery*, 3:409–425, 1999.

[7] Peter D. England and Richard J. Verrall. A flexible framework for stochastic claims reserving. *Proceedings of the Casualty Actuarial Society*, LXXXVIII:1–38, 2001.

[8] Norman L. Johnson, Samuel Kotz, and Adrienne W. Kemp. *Univariate Discrete Distributions*. J. Wiley and Sons, second edition, 1992.

[9] Samuel Kotz Johnson, Norman L. and N. Balakrishnan. *Discrete Multivariate Distributions*. J. Wiley and Sons, 1996.

[10] Harry H. Panjer Klugman, Stuart A. and Gordon E. Willmot. *Loss Models, From Data to Decisions*. John Wiley and Sons, 1998.

[11] Samual Kotz, Norman L. Johnson, and N. Balakrishnan. *Continuous Multivaritate Distributions*, volume 1. John Wiley and Sons, second edition, 2000.

[12] Michael Kunkler. Modelling zeros in stochastic reserve methods. *Insurance Mathematics & Economics*, 34:23–35, 2004.

[13] N. N. Lebedev. *Special Functions and Their Applications*. Dover, New York, NY, 1972.

[14] Thomas Mack. The stochastic model underlying the chain ladder method. *ASTIN Bulletin*, Vol. REF:22–61, 1991.

[15] Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, Vol. 23 No. 2:213–25, 1993.

[16] Thomas Mack. Credible claims reserves: The Benktander method. *ASTIN Bulletin*, Vol. 20 No. 2:333–347, 2000.

[17] Stephen J. Mildenhall. Correlation and aggregate loss distributions with an emphasis on the Iman-Conover method. *CAS Correlation Working Party draft, www.mynl.com/wp/ic.pdf*, 2004.

[18] Stephen J. Mildenhall. A practical Bayesian model of claim count development. *forthcoming*, 2005.

[19] Daniel M. Murphy. Unbiased loss development factors. *Proceedings of the Casualty Actuarial Society*, LXXXI:154–222, 1994.

[20] Ragnar Norberg. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin*, Vol. 23, No. 1:95–115, 1993.

[21] Ioannis Ntzoufras and Petros Dellaportas. Bayesian modelling of outstanding liabilities incorporating claim count uncertainty. *NAAJ*, 6 No. 1:113–128, 2002.

[22] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numercial Recipies in C, The Art of Scientific Computing*. Cambridge University Press, second edition, 1992.

[23] A. E. Renshaw and R. J. Verrall. A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, 4, IV:903–923, 1998.

[24] Ira Robbins. A Bayesian credibility formula for IBNR claims. *Proceedings of the Casualty Actuarial Society*, LXXIII:129–164, 1986.

[25] David A. Stephens, Martin J. Crowder, and Petros Dellaportas. Quantification of automobile insurance liability: a Bayesian failure time approach. *Insurance Mathematics & Economics*, 34:1–21, 2004.

[26] Gary G. Venter. Testing the assumptions of age-to-age factors. *Proceedings of the Casualty Actuarial Society*, LXXXIV:807–847, 1998.

[27] R. J. Verrall. Bayes and empirical Bayes estimation for the chain ladder method. *ASTIN Bulletin*, 20 No. 2:218–243, 1990.

[28] R. J. Verrall. A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims resering. *NAAJ*, 8 No. 3:67–89, 2004.

[29] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis*. Cambridge University Press, Cambridge, fourth edition, 1927.

[30] Thomas S. Wright. Stochastic claims reserving when past claim numbers are known. *Proceedings of the Casualty Actuarial Society*, LXXIX:255–361, 1992.

[31] Ben Zehnwirth. Probabilistic development factor models with application to loss reserve variability, prediction intervals and risk based capital. *Casualty Actuarial Society Forum*, Spring:447–605, 1994.

[32] Ben Zehnwirth and Glen Barnett. Best estimates for reserves. *Proceedings of the Casualty Actuarial Society*, LXXXVII:245–321, 2000.