A MULTIVARIATE MODEL OF THE TOTAL CLAIMS PROCESS

BY J. DAVID CUMMINS AND LAUREL J. WILTBANK

University of Pennsylvania, Philadelphia

KEYWORDS

Compound distributions, aggregate claim distributions.

Much of the risk theory literature deals with the total claims distribution $F(x) = \sum_{k=0}^{\infty} p_k S^{k^*}(x)$, where p_k = the probability of k claims and $S(x)$ is the distribution function of severity. Both p_k and $S(x)$ are univariate probability distributions. Thus, $F(x)$ can be interpreted as a model of claims from one class of policies or as an aggregate model where p_k and $S(x)$ represent mixed probability distributions from a heterogeneous portfolio of policies. An alternative approach to modelling total claims in the latter case would be to recognize explicitly that total claims are the result of the interaction of multivariate processes. In the most general case, total claims arise from a multivariate accident process where each accident triggers multivariate claims frequency and severity processes.

The purpose of this article is to present a multivariate model of total claims and to develop the cumulant generating function of this distribution. Such a model is superior to the traditional model in two respects: (1) It permits explicit recognition of shifts in the overall portfolio composition. Applications of the traditional model, in contrast, rely on the assumption that the portfolio composition is relatively constant over time. (2) It facilitates the evaluation of the effects of reinsurance on the total claims distribution when the reinsurance arrangements are not the same in different segments of the portfolio.

THE TOTAL CLAIMS DISTRIBUTION

As indicated, the total claims distribution involves three multivariate processes: the accident process, claims frequency processes, and claims severity processes. Each type of accident can be assigned unique multivariate claims frequency and severity processes. For example, automobile accidents can give rise to bodily injury liability, property damage liability, and physical damage claims; workers' compensation accidents can give rise to wage loss and medical claims.

Dependencies can arise at various stages of the process. For example, bodily injury and property damage liability claims severity from a given accident may be dependent. The model presented in this article recognizes dependencies of three types: dependencies among different types of accident frequencies, among different types of claims frequencies for a given accident of a particular type, and among claims severities for a given accident of a particular type. The authors believe that these are the types of dependencies most likely to arise in practice. Dependencies may exist among different types of accidents due to weather conditions, business cycles, or other factors. Further, accidents with more (or more severe) claims of one type also may be likely to have more (or more severe) claims of other types and vice versa. Assuming independence among the accident frequency, claims frequency, and claims severity distributions is analogous to the usual assumption of independence between frequency and severity. Dependencies may exist among severities from different types of accidents but this is likely to be attributable to a common inflationary effect, which can be handled more satisfactorily through the use of forecasting models.

The Accident Process

Let $i = 1, 2, \ldots, A$ index the *types of accidents*, and let N_i be the random total number of *accidents of type i* in a given period. The N_i may be statistically dependent, with joint density:

(1) $$\Pr\{N_1 = n_1; N_2 = n_2; \ldots, N_A = n_A\} = q(n_1, n_2, \ldots, n_A), \qquad n_i = 0, 1, 2, \ldots.$$

The Claim Frequency Process

For a single accident of type i, claims of B_i different *claim types* can arise. Let K_{ij} be the random variable, number of *claims* of claim-type j from a single accident of type i. The K_{ij} may be dependent for a given accident of type i, with joint density:

(2) $$\Pr\{K_{i1} = k_1; K_{i2} = k_2; \ldots, K_{iB_i} = k_{B_i}\} = p_i(k_1, k_2, \ldots, k_{B_i}),$$
$$(k_j = 0, 1, 2, \ldots; i = 1, 2, \ldots, A).$$

The total *number of claims* of all types from a *single* accident of type i is $K_i = K_{i1} + K_{i2} + \cdots + K_{iB_i}$. The numbers of claims from *successive* accidents of the same type or from accidents of different types are assumed to be independent.

The Claim Severity Process

Each claim of claim-type j in an accident of accident-type i is assumed to have a random severity X_{ijl}, where l indexes the individual claim, $l = 1, 2, \ldots, K_{ij}$, in a different accident and claim type. Thus, the total severity in claim category j in a single accident of type i is the random sum:

(3) $$X_{ij} = \begin{cases} 0, & K_{ij} = 0 \\ X_{ij1} + X_{ij2} + \cdots + X_{ijK_{ij}}, & K_{ij} \neq 0 \end{cases}$$

where $X_{ijl}, X_{ij} \geq 0$; $i = 1, 2, \ldots, A$; and $j = 1, 2, \ldots, B_i$. Then, the *total severity* of a *single accident* of type i over all claim categories is:

(4) $$X_i = X_{i1} + X_{i2} + \cdots + X_{iB_i}.$$

The random variable X_i is clearly conditional on the outcome of the random vector of claim frequencies, $\mathbf{K}_i = (K_{i1}, K_{i2}, \ldots, K_{iB_i})$, associated with a single accident of type i. The individual claim severities from a given accident of type

i, $X_{ijl}(j=1,2,\ldots,B_i;\ l=1,2,\ldots,K_{ij})$, can be treated as mutually statistically dependent. The joint severity distribution function can be written as:

(5) $$S_i(x_{i11},\ldots,x_{i1K_{i1}};\ x_{i21},\ldots,x_{i2K_{i2}};\ldots;\ x_{iB_i1},\ldots,x_{iB_iK_{iB_i}}|\boldsymbol{k}_i).$$

The marginal distributions of (5) can be written as:

(6) $$S_{ijl}(x_{ijl}|\boldsymbol{k}_i)=S_{ij\cdot}(x_{ijl}|\boldsymbol{k}_i),\quad \text{all } l.$$

This permits the distributions to vary depending upon the claims frequency vector $\boldsymbol{k}_i$, e.g., more dangerous distributions may characterize accidents with larger numbers of claims. The notation $S_{ij\cdot}(x_{ijl}|\boldsymbol{k}_i)$ reflects the assumption that the marginals are identical (but not necessarily independent) for different claims of the same type arising out of an accident with a given claim vector $\boldsymbol{k}_i$.

One can also define the conditional distribution of the sum of claims from an accident of type i (equation (4)):

(7) $$S_i(x_i|\boldsymbol{k}_i)=\Pr\{X_i\leqslant x_i|\boldsymbol{K}_i=\boldsymbol{k}_i\}.$$

This distribution is a convolution of simpler distributions *only* in the special case where the X_{ijl} are statistically mutually independent for all (j, l) with i fixed. It is assumed that the X_i are independent between different accident types. Independence is also assumed among claim severities for different accidents of the same type.

Distribution of Accidents Among Claim Categories

Given the foregoing, the next step is to obtain the distribution of the total severity of all accidents of accident-type i. First, note that the vector of outcomes of $\boldsymbol{K}_i$ can be thought of as a selection of one of a countable number of *patterns*: $\boldsymbol{k}_i(0)=(0,0,\ldots,0)$; $\boldsymbol{k}_i(1)=(1,0,\ldots,0)$; $\boldsymbol{k}_i(2)=(0,1,\ldots,0);\ldots;$ $\boldsymbol{k}_i(B_i)=(0,0,\ldots,1)$; $\boldsymbol{k}_i(B_i+1)=(1,1,\ldots,0);\ldots$, etc. Indexing this set by $\pi=0,1,2,\ldots,\Pi$, where Π may be infinite, we observe that (2) provides the probability distribution of these patterns:

(8) $$p_i(\boldsymbol{k}_i(\pi))=\Pr\{\boldsymbol{K}_i=\boldsymbol{k}_i(\pi)\}=p_i(\pi).$$

These are the probabilities of patterns of claim numbers for a *single* accident. If the patterns generated by each of the n_i accidents of this type are mutually independent and independent of all patterns of other accident types, the distribution of patterns over all accidents of the ith type follows a multinomial law.

Let $\boldsymbol{N}_i=(N_i(\pi);\ \pi=0,1,\ldots,\Pi)$ be the random vector describing the distribution of the N_i accidents of type i over the various claim category patterns where $N_i(\pi)$ is the number of accidents with claim pattern $\boldsymbol{k}_i(\pi)$. Then

(9) $$\begin{aligned}\Pr\{\boldsymbol{N}_i=\boldsymbol{n}_i|N_i=n_i\}&=p_i(n_i(0),\,n_i(1),\ldots,n_i(\Pi)|n_i)\\&=\binom{n_i}{n_i(0)n_i(1)\cdots n_i(\Pi)}[p_i(0)]^{n_i(0)}\\&\quad\times[p_i(1)]^{n_i(1)}\cdots[p_i(\Pi)]^{n_i(\Pi)}\end{aligned}$$

where

$$p_i(\pi) \geq 0; \qquad \sum_{\pi=0}^{\Pi} p_i(\pi) = 1 \quad \text{and} \quad \sum_{\pi=0}^{\Pi} n_i(\pi) = n_i.$$

Note that the $p_i(\pi)$ may come from any appropriate probability distribution. It is the *allocation* of accidents among claim category patterns and not the probabilities of patterns which is multinomial.

We can now find the distribution of Y_i, the total value of claims in accident class i, conditional on the realized number of accidents, $N_i = n_i$. Note that the single-accident conditional severity distribution function, (7), can be rewritten as $S_i(x_i|\pi)$, $\pi = 0, 1, \ldots \Pi$. It follows that:

$$\text{(10)} \qquad \Pr\{Y_i \leq y_i|n_i\} = H_i(y_i|n_i) = \sum_{\mathbf{n}_i} p_i(n_i(0), n_i(1), \ldots, n_i(\Pi)|n_i) \\ \times [S_i(y_i|0)]^{n_i(0)*} * [S_i(y_i|1)]^{n_i(1)*} * \cdots * [S_i(y_i|\Pi)]^{n_i(\Pi)*}$$

where $\sum_{\mathbf{n}_i}$ indicates summation over all possible realizations of $\mathbf{N}_i$ such that $\sum_{\pi=0}^{\Pi} n_i(\pi) = n_i$.

The Total Claims Distribution

The unconditional grand total value of claims over all accident classes can be written as:

$$\text{(11)} \qquad Y = Y_1 + Y_2 + \cdots + Y_A.$$

The distribution function of Y is easy to specify because the severities of different accident classes are assumed to be independent. From (1) and (10),

$$\text{(12)} \qquad \Pr[Y \leq y] = F(y) = \sum_{n_1} \sum_{n_2} \cdots \sum_{n_A} q(n_1, n_2, \ldots, n_A) \\ \times [H_1(y|n_1)] * [H_2(y|n_2)] * \cdots * [H_A(y|n_A)].$$

The Cumulant Generating Function

The formula for $F(y)$ is mathematically intractable for most probability distributions encountered in practice. However, the cumulant generating function of $F(y)$ can be written quite compactly, facilitating the derivation of cumulants for use in the Normal-Power or Gamma approximations. The cumulant generating function is preferable to the moment generating function since moments and cumulants can be obtained much more simply using the former function. The function is analogous to that developed by Brown (1977) for the univariate accident frequency-claims frequency-claims severity case.

To obtain the cumulant generating function, we first derive the moment generating function. Let

$$\text{(13)} \qquad M_Y(t) = \int e^{ty}\, dF(y) \quad \text{and} \quad \Upsilon_i(t|n_i) = \int e^{ty_i}\, dH_i(y_i|n_i).$$

From (12), one obtains:

(14) $$M_Y(t)=\sum_{n_1}\sum_{n_2}\cdots\sum_{n_A} q(n_1,n_2,\ldots,n_A)\prod_{i=1}^{A} \Upsilon_i(t|n_i).$$

But if

(15) $$\Psi_i(t|\pi)=\int e^{tx_i}\,dS_i(x_i|\pi)$$

we find from (9), (10), and the expansion of a multinomial:

(16) $$\begin{aligned}\Upsilon_i(t|n_i)&=\sum_{n_i} p_i(n_i(0),\ldots,n_i(\pi)|n_i)[\Psi_i(t|0)]^{n_i(0)}\\&\quad\times[\Psi_i(t|1)]^{n_i(1)}\cdots[\Psi_i(t|\Pi)]^{n_i(\Pi)}\\&=[p_i(0)\Psi_i(t|0)+p_i(1)\Psi_i(t|1)+\cdots+p_i(\Pi)\Psi_i(t|\Pi)]^{n_i}.\end{aligned}$$

Using (16), the moment generating function (14) can be written:

(17) $$M_Y(t)=\sum_{n_1}\sum_{n_2}\cdots\sum_{n_A} q(n_1,n_2,\ldots,n_A)\prod_{i=1}^{A}\left[\sum_{\pi=0}^{\Pi} p_i(\pi)\Psi_i(t|\pi)\right]^{n_i}.$$

This is the moment generating function of the multivariate accident frequency distribution $q(n_1,n_2,\ldots,n_A)$ with auxiliary parameters $\log[\sum_{\pi=0}^{\Pi} p_i(\pi)\Psi_i(t|\pi)]$, $i=1,\ldots,A$. Thus, (17) can be written as:

(18) $$M_Y(t)=M_{N_1,N_2,\ldots,N_A}\left\{\log\left[\sum_{\pi=0}^{\Pi} p_1(\pi)\Psi_1(t|\pi)\right],\ldots,\log\left[\sum_{\pi=0}^{\Pi} p_A(\pi)\Psi_A(t|\pi)\right]\right\}.$$

The definition of the cumulant generating function is $C_Y(t)=\log M_Y(t)$. Using this definition and (18), one can write

(19) $$C_Y(t)=C_{N_1,N_2,\ldots,N_A}\left\{\log\left[\sum_{\pi=0}^{\Pi} p_1(\pi)\Psi_1(t|\pi)\right],\ldots,\log\left[\sum_{\pi=0}^{\Pi} p_A(\pi)\Psi_A(t|\pi)\right]\right\}.$$

Next, notice that $\Psi_i(t)=\sum_{\pi=0}^{\Pi} p_i(\pi)\Psi_i(t|\pi)$, $i=1,2,\ldots,A$, is the moment generating function of the mixed severity distribution:

(20) $$S_i(x_i)=\sum_{\pi=0}^{\Pi} p_i(\pi)S_i(x_i|\pi).$$

Hence, (19) can be rewritten as:

(21) $$C_Y(t)=C_{N_1,N_2,\ldots,N_A}\{C_{X_1}(t),\ldots,C_{X_A}(t)\}$$

where $C_{X_i}(t)=$ the cumulant generating function of $S_i(x_i)$.

An interesting special case occurs when the claim severities within each accident type are mutually independent. Recalling (3) and (6), one obtains:

(22) $$\Psi_i(t)=\sum_{\pi=0}^{\Pi} p_i(\pi)[\Psi_{i1\cdot}(t|\pi)]^{k_{1i}}\cdots[\Psi_{iB_i}(t|\pi)]^{k_{B_i}}$$

where $\Psi_{ij\cdot}(t)=$ the moment generating function of $S_{ij\cdot}(x_{ijl})$. Equation (22) is recognizable as the moment generating function of a multivariate claims frequency process with auxiliary parameters $\log \Psi_{ij\cdot}(t)$, $j = 1, 2, \ldots, B_i$. Using (22), equation (21) becomes:

$$C_Y(t) = C_{N_1,\ldots,N_A}\{C_{K_{11},\ldots,K_{1B_1}}[C_{X_{11}}(t), \ldots, C_{X_{1B_1}}(t)], \ldots, C_{K_{A1},\ldots,K_{AB_A}}[C_{X_{A1}}(t), \ldots, C_{X_{AB_A}}(t)]\} \tag{23}$$

where $C_{X_{ij}}(t)$ = the cumulant generating function of $S_{ij\cdot}(x_{ijl})$, and $C_{K_{i1},\ldots,K_{iB_i}}[\cdot]$ = the cumulant generating function of the multivariate claims frequency distribution applicable to accident type i.

Examples of Cumulants

The first and second cumulants of $F(y)$ are straightforward generalizations of the usual formulas for the first two cumulants of sums of random variables. The third cumulant, while also a generalization, is more interesting and is shown below:

$$\kappa_{3Y} = \sum_{i=1}^{A} [\kappa_{3N_i}\kappa_{1X_i}^3 + \kappa_{1N_i}\kappa_{3X_i} + 3\kappa_{2N_i}\kappa_{1X_i}\kappa_{2X_i}] + 3\sum\sum_{i\neq g} [\kappa_{1N_i1N_g}\kappa_{2X_i}\kappa_{1X_g}]$$
$$+3\sum\sum_{i\neq g}[\kappa_{2N_i1N_g}\kappa_{1X_i}^2\kappa_{1X_g}] + 6\varepsilon \sum_{i\neq g\neq h} \kappa_{1N_i1N_g1N_h}\kappa_{1X_i}\kappa_{1X_g}\kappa_{1X_h} \tag{24}$$

where $i, g = 1, \ldots, A$; and $\varepsilon = 1$ for $A \geq 3$, 0 otherwise. The double summation $\sum\sum_{i\neq g}$ means the summation over both subscripts, omitting terms where the subscripts are equal. The summation $\sum_{i\neq g\neq h}$ means the summation over all combinations $i \neq g \neq h$, where $i, g, h = 1, 2, \ldots, A$.

The κ's are cumulants. Numerical subscripts refer to the cumulant number, while letter subscripts refer to random variables. Symbols with more than one of each type of subscript are cross-cumulants. For example, $\kappa_{1N_11N_2}$ is the first cross-cumulant (covariance) of the accident frequency random variables N_1 and N_2.

Cumulants of the mixed severity distributions $S_i(x_i)$, $i = 1, 2, \ldots, A$, can be obtained directly using (2), (3), (4), (7), and (20). The formulas for the first three cumulants are as follows:

$$\kappa_{1X_i} = \sum_{\pi} p_i(\pi)\mu_i(X_i|\pi) \tag{25}$$

$$\kappa_{2X_i} = \sum_{\pi} p_i(\pi)\alpha_{2i}(X_i|\pi) - \left[\sum_{\pi} p_i(\pi)\mu_i(X_i|\pi)\right]^2 \tag{26}$$

$$\kappa_{3X_i} = \sum_{\pi} p_i(\pi)\alpha_{3i}(X_i|\pi) - 3\left[\sum_{\pi} p_i(\pi)\mu_i(X_i|\pi)\right]\cdot\left[\sum_{\pi} p_i(\pi)\alpha_{2i}(X_i|\pi)\right]$$
$$+2\left[\sum_{\pi} p_i(\pi)\mu_i(X_i|\pi)\right]^3 \tag{27}$$

where

$$\mu_i(X_i|\pi) = \int x_i \, dS_i(x_i|\pi) \quad \text{and} \quad \alpha_{mi}(x_i|\pi) = \int x_i^m \, dS_i(x_i|\pi).$$

The formulas for the moments of $S_i(x_i|\pi)$ are straightforward but cumbersome. The moments of individual claim severities are permitted to vary according to the claim pattern, π, allowing for the possibility of larger claim severity means, variances, etc. for accidents involving large numbers of claims.

If the individual claim severity moments are assumed to be equal within a claim type (regardless of π), the cumulant formulas are simplified. The second cumulant, for example, is:

$$\kappa_{2X_i} = \sum_j [\kappa_{1K_{ij}}\kappa_{2X_{ij}} + \kappa_{2K_{ij}}\kappa^2_{1X_{ij}} + (\kappa_{2K_{ij}} + \kappa^2_{1K_{ij}} - \kappa_{1K_{ij}})\kappa_{1X_{ijl}1X_{ijg}}]$$
$$+ \sum\sum_{j \neq h} [\kappa_{1K_{ij}1K_{ih}}\kappa_{1X_{ij}}\kappa_{1X_{ih}} + (\kappa_{1K_{ij}1K_{ih}} + \kappa_{1K_{ij}}\kappa_{1K_{ih}})\kappa_{1X_{ij}1X_{ih}}] \tag{28}$$

where j, $h = 1, 2, \ldots, B_i$; $\kappa_{mK_{ij}}$ = the cumulants of the marginals of (2), here $m = 1, 2$; $\kappa_{1K_{ij}1K_{ih}}$ = first cross-cumulant of (2), $j \neq h$; $\kappa_{mX_{ij}}$ = cumulants of $S_{ij}(x_{ijl})$, assumed identical for all l within claim type j, $m = 1, 2$, where $S_{ij}(x_{ijl})$ is (6) without the condition; $\kappa_{1X_{ijl}1X_{ijg}}$ = first cross-cumulant of claim severities of the same type, $l \neq g$, assumed identical for all l, g; and $\kappa_{1X_{ij}1X_{ih}}$ = first cross-cumulant of claim severities of different types, $j \neq h$.

In practical applications, it generally will be necessary to combine the higher order claim patterns to permit the estimation of severity distributions. For example, with two claim types, analysis might be confined to the following patterns: $k_i'(0) = (0, 0)$; $k_i'(1) = \{(K_{i1}, 0); K_{i1} = 1, 2, \ldots\}$; $k_i'(2) = \{(0, K_{i2}), K_{i2} = 1, 2, \ldots\}$; and $k_i'(3) = \{(K_{i1}, K_{i2}); K_{i1}, K_{i2} \geq 1\}$. The probabilities of each revised pattern, $k_i'(\pi)$, $\pi = 1, 2, 3$, can be obtained by summing the appropriate probabilities from (8). The severity distributions $S_i(x_i|1)$ and $S_i(x_i|2)$ are univariate distributions estimated on a set of observations on X_{i1} and X_{i2}, respectively, from accidents with the designated claim patterns $\pi = 1$ and $\pi = 2$, respectively. (Recall (3).) Thus, in estimating the severity distributions, no distinction is made among accidents which have different numbers of claims, K_{ij}. Rather, the sum of claim severities of a particular type from a given accident is considered a single observation from the appropriate severity distribution. Cumulants of $S_i(x_i|3)$ are obtained from a bivariate severity distribution estimated on a set of observations on (X_{i1}, X_{i2}), where $K_{i1}, K_{i2} \geq 1$. $S_i(x_i|0)$ is a degenerate distribution.

Conclusion

This article has presented a multivariate model of the total claims distribution. The model could be used in conjunction with the Normal-Power or Gamma approximations to model the total claims of an insurance company by estimating cumulants for each segment of the portfolio and combining the cumulants according to the appropriate formulas. This approach should be superior to the traditional $F(x)$ model for some applications because it focuses directly on individual segments of the portfolio and clarifies the interactions among the segments. Empirical research is needed on the types of distributions that are appropriate for modeling the claims process in the multivariate context and about the nature and magnitude of the dependencies among the variables comprising the process.

REFERENCES

BEARD, R E., PENTIKAINEN, T and PESONEN, E (1977) *Risk Theory*, 2nd ed Chapman and Hall London.

BROWN, A (1977) Cumulants of Convolution-Mixed Distributions. *ASTIN Bulletin* **9**, 59–63

JOHNSON, N L and KOTZ, S (1969) *Discrete Distributions* John Wiley & Sons New York.