An Alternative Frequency Dependence Model and its Applications*

Shubiao Li
Comerica Loan Center
Comerica Bank
Auburn Hills, Mich., USA

2011 Enterprise Risk Management Symposium
Society of Actuaries
March 14-16, 2011

* The ideas in this paper represent those of the author and not of any institution. Contact email: Shubiaoli2009@yahoo.com.

Abstract

In this paper, a multivariate quasi-negative binomial distribution is proposed to model frequency dependence among different risk types. The operational risk diversification effect is illustrated through frequency dependency modeled by the bivariate quasi-negative binomial distribution under a framework of Monte Carlo simulation.

Keywords and phrases: Bivariate quasi-negative binomial distribution; marginal distribution; dependence; operational risk.

1. Introduction

In the banking industry, three types of dependence—loss (severity), frequency and aggregate loss—can be observed in operational risk loss data (Chernobai et al. 2007). Under the advanced measurement approaches (AMA) of Basel II guidelines, operational risk capital charge calculations may be allowed to take these types of dependence into account.

> "Banks may use internal estimates of dependence among operational losses across and within units of measure if the [bank] can demonstrate to the satisfaction of the [AGENCY] that its process for estimating dependence is sound, robust to a variety of scenarios, and implemented with integrity, and allows for the uncertainty surrounding the estimates. If the [bank] has not made such a demonstration, it must sum operational risk exposure estimates across units of measure to calculate its total operational risk exposure." (U.S. Office of the Federal Register, National Archives and Records Administration 2007)

There are many studies on risk dependence modeling. Copula theory is the most popular and extremely useful in modeling severity dependence structure.

Let $(X_1, \cdots, X_n)$ be an n-dimension random vector with marginal distributions $F_1, ..., F_n$ and the associated uniform random variables $U_i = F_i(X_i)$, $i = 1, ..., n$. If $C(u_1, ..., u_n)$ is the distribution function of $(U_1, \cdots, U_n)$, then the distribution function of $(X_1, \cdots, X_n)$ is given by

$$H(x_1, \cdots, x_n) = C(F_1(x_1), \cdots, F_n(x_n))$$

and $C(u_1, ..., u_n)$ is the copula function. So, the copula can be interpreted as a function that links the marginal distributions of a random vector to form their joint distribution. Conversely, if C is a copula and F_i, $i = 1, \cdots, n$ are distribution functions, then the function H above is a joint distribution function with marginals F_i, $i = 1, \cdots, n$.

Copula theory captures the dependence structure among continuous random variables and hence offers great flexibility in building multivariate statistical models. However, copula theory cannot fully capture the dependence structure of discrete random variables (Genest and Neslehova 2007). More detailed review on copulas and dependence can be found in McNeil et al. (2005) and Nelsen (2006).

Frequency dependence occurs when different types of risks share some common risk driving factors such as the business size or economic cycle. These risk factors may or may not be observable. Wang (1998) provided examples where individual risks are correlated because the risks are subject to the same claim generating mechanisms. Examples include property insurance where risk portfolios in the same geographic region are correlated because claims may be contingent upon the occurrence of a natural disaster. Powojowski et al. (2002) assumed that the number of events of the operational processes follows a Poisson distribution and obtained a multivariate Poisson distribution by assuming all these number of events share an underlying common Poisson process. The special case, a bivariate Poisson model, is the same as the model obtained through trivariate reduction (Johnson et al. 1997, chap. 37). Lindskog (2003) derived a more general common Poisson shock processes model using the same logic. A limitation of the common Poisson shock model is that it can only model positive correlations.

Aggregate loss dependence is the joint effect of frequency dependence and loss dependence. Due to the complexity of the dependence structure, aggregate loss dependence modeling is often achieved by Monte Carlo simulation.

The central idea for both copula and common Poisson shock models is to construct a joint distribution for a random vector of risk types. The joint distribution must be able to describe the marginal behavior of individual risk types and their dependence structure as well. Li et al. (2010) proposed a new method to construct multivariable discrete distributions by using the generalized Lagrangian distribution of the first type. With the method, one can derive numerous discrete multivariate distributions including bivariate quasi-negative binomial distribution (BQNBD). Section 2 describes the BQNBD and its probabilistic structure. The risk diversification effect is illustrated through frequency dependency modeled by the BQNBD under a framework of Monte Carlo simulation in Section 3.

2. Bivariate Quasi-Negative Binomial Distribution and its Probabilistic Structure

Li et al. (2006) and Li (2007) defined the class of generalized Lagrangian distributions, in which an extra parameter (the Lagrangian expansion point) was brought into its probability mass function (pmf). Let $f(z)$ and $g(z)$ be analytic functions, $\left\{D^{x-1}\left[\left(g(z)\right)^x f'(z)\right]\right\}_{z=0} \geq 0$ and $g(0)$ > 0, where $D = \partial / \partial z$. If there is a point t > 0, such that $f(t)$ > 0 and $g(t)$ > 0, for $x \in N$, where N is the set of natural numbers, then the generalized Lagrangian probability distribution of the first kind ($GL_1(f, g, t; x)$) is defined as

$$p(x \mid t) = P(X = x) = \begin{cases} f(0) / f(t), & x = 0 \\ \dfrac{\left(t / g(t)\right)^x}{x!\, f(t)} \left\{D^{x-1}\left[\left(g(z)\right)^x f'(z)\right]\right\}_{z=0}, & x \geq 1. \end{cases} \quad (1)$$

Many discrete probability distributions can be derived by using specific $f(z)$ and $g(z)$ functions. For example, if $g(z)$ = $e^{\lambda z}$ and $f(z)$ $= e^{\theta z}$, then, the pmf of the generalized Lagrangian distribution is

$$P(X(t) = x) = \theta t \left(\theta t + \lambda t x\right)^{x-1} e^{-\theta t - \lambda t x} / x!, \text{ for } x = 0, 1, 2, \ldots, \quad (2)$$

which is the pmf of the generalized Poisson distribution (GPD) when t = 1 and is the pmf of the Poisson distribution when $\lambda = 0$ (Consul 1989). In the pmf in (2), we have 0 < λ < 1, θ > 0 and t > 0.

Let $X_i(t)$ ($i = 1, 2, \ldots, m$) be discrete random variables following the generalized Lagrangian distributions $GL_1(f_i, g_i, t; x_i)$ with probability mass function $p_i(x_i \mid t)$ and the variable t be a random variable with density $s(t)$. Li et al. (2010) showed that under certain conditions, the function $p(x_1,...,x_m,t)=s(t)\prod_{i=1}^{m} p_i(x_i\mid t)$ is a joint probability distribution of $(X_1,X_2,...,X_m,T)$ for random variables $X_1,X_2,...,X_m$ and T, where T is a random variable with the density function $s(t)$. With the joint probability function, Li et al. (2010) proposed a new method to generate multivariate discrete distributions and their marginal distributions. By choosing $f_i(z)=e^{\theta_i z}$, $g_i(z)=e^{\lambda_i z}$ and $i=1,2$ and supposing the random variable t follows a gamma distribution with the density function $s(t)=\beta^{\alpha}t^{\alpha-1}e^{-\beta t}/\Gamma(\alpha),\ t>0$, Li et al. (2010) obtained a new bivariate discrete distribution: the bivariate quasi-negative binomial distribution. For more details in generating multivariate distributions including multivariate quasi-negative binomial distribution (MQNBD), refer to Li et al. (2010).

Bivariate random variables *(X, Y)* are said to have a BQNBD if their pmf (Li et al. 2010) is given by:

$$P(X=x,Y=y)=\frac{\Gamma(x+y+\alpha)}{\Gamma(\alpha)y!x!}\frac{\delta_1^{y+\alpha}\delta_2^{x+\alpha}\left(1+\varepsilon_2 y\right)^{y-1}\left(1+\varepsilon_1 x\right)^{x-1}}{\left[\delta_1+\delta_2+\varepsilon_1\delta_2 x+\varepsilon_2\delta_1 y+\delta_1\delta_2\right]^{x+y+\alpha}}\ ,\ x,y=0,1,2,\ldots \qquad (3)$$

where constants $\varepsilon_i\ge 0,\delta_i\ge 0,\alpha>0$ and $i=1,2$.

The marginal distributions of X and Y are, respectively, given by,

$$P(X=x)=\frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)}\frac{1}{1+\varepsilon_1 x}\left(\frac{1+\varepsilon_1 x}{1+\delta_1+\varepsilon_1 x}\right)^{x}\left(\frac{\delta_1}{1+\delta_1+\varepsilon_1 x}\right)^{\alpha},\ x=0,1,2,\ldots \qquad (4)$$

$$P(Y=y)=\frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)}\frac{1}{1+\varepsilon_2 y}\left(\frac{1+\varepsilon_2 y}{1+\delta_2+\varepsilon_2 y}\right)^{y}\left(\frac{\delta_2}{1+\delta_2+\varepsilon_2 y}\right)^{\alpha},\ y=0,1,2,\ldots \qquad (5)$$

The above distribution, quasi-negative binomial distribution (QNBD), was also derived through a gamma mixture of generalized Poisson distribution by Li et al. (2008).

The covariance of $(X,\ Y)$ is

$$\operatorname{cov}(X,Y) = \frac{1}{\Gamma(\alpha)}\int_0^\infty \frac{t^{\alpha+1}e^{-t}}{(\delta_1-\varepsilon_1 t)(\delta_2-\varepsilon_2 t)}dt - \frac{1}{\left(\Gamma(\alpha)\right)^2}\int_0^\infty \frac{t^{\alpha}e^{-t}}{\delta_1-\varepsilon_1 t}dt\int_0^\infty \frac{t^{\alpha}e^{-t}}{\delta_2-\varepsilon_2 t}dt\,. \tag{6}$$

It has been shown that the model (3) can model both positive and negative correlation of *X* and *Y* (Li et al. 2010).

The conditional distribution of *X* given *Y* is

$$P(X=x|Y=y) = \frac{\Gamma(x+y+\alpha)}{x!\Gamma(\alpha+y)}\frac{\delta_2^x(\delta_1+\varepsilon_2\delta_1 y+\delta_1\delta_2)^{\alpha+y}\left(1+\varepsilon_1 x\right)^{x-1}}{\left(\delta_1+\delta_2+\varepsilon_1\delta_2 x+\varepsilon_2\delta_1 y+\delta_1\delta_2\right)^{x+y+\alpha}}\,. \tag{7}$$

This also is a QNBD.

Equations (3)-(7) describe the probabilistic structure of the BQNBD. Noting that if we let $\varepsilon_i = 0, i = 1,2$, then the equation (3) reduces to

$$p(X=x,Y=y) = \frac{\Gamma(x+y+\alpha)}{x!y!\Gamma(\alpha)}\frac{\delta_2^{x+\alpha}\delta_1^{y+\alpha}}{\left(\delta_1+\delta_2+\delta_1\delta_2\right)^{x+y+\alpha}},\ x,y = 0,1,2,\ldots, \tag{8}$$

The marginal distributions (4) and (5) reduce to the negative binomial distributions (NBD) and the covariance of $(X,\ Y)$ is given by

$$\operatorname{cov}(X,Y) = \alpha\delta_1^{-1}\delta_2^{-1}. \tag{9}$$

Other special cases can be obtained in a similar way. For example, one can assume one of the $\varepsilon_i, i = 1, 2$ to be zero and the other parameters to be certain positive numbers.

3. Frequency Dependence and Diversification Effect of Operational Risk

The bivariate quasi-negative binomial distribution provides a new distribution for studying frequency dependency in practice. Li et al. (2011) investigated the properties of the marginal distribution of BQNBD and showed that the moments of the marginal distribution do not exist in some situations and the limiting distribution of the marginal distribution is the generalized Poisson distribution under certain conditions. Various application results in different fields showed that the marginal distribution and its zero-inflated model is extremely suitable for operational risk or insurance claims where the data is highly skewed, has heavy tails or excessive numbers of zeros. More detailed review on QNBD can be found in Li et al. (2011). In this section, we apply the bivariate distribution BQNBD to an operational risk data to demonstrate its usefulness in studying diversification effects of risk.

Li et al. (2010) applied the BQNBD to an operational dataset to model the frequency dependence of two risk types. The frequency data was from American Banking Association (ABA) and on monthly basis. The sample size is 50. Let variable *X* be the number of loss events that occurred for the risk type of employment practices and workplace safety per month and *Y* be the number of loss events that occurred for the risk type of client, products and business practices per month. The sample means and standard deviations are, respectively, $\overline{x}$ = 14.48, s_x = 7.88 and $\overline{y}$ = 13.52 and s_y = 7.80. The Pearson correlation coefficient between the two risk types is 0.45. Clearly, both *X* and *Y* exhibit over-dispersion property. Li et al. (2010) used the BQNBD to model the dependence of *X* and *Y* and justified that the BQNBD provided a good fit only when all parameters in the BQNBD are positive. The parameter estimations are $\hat{\delta}_1$ = 2.3511(1.4836), $\hat{\delta}_2$ = 2.6027(1.6284), $\hat{\varepsilon}_1$ = 0.0377(.0143), $\hat{\varepsilon}_2$ = .0440(.0154) and $\hat{\alpha}$ = 21.4488(11.5390). The numbers in the parentheses are the standard error of the estimation. These parameters specify the pmf of the marginal distribution of Y in equation (5) and the pmf of the conditional distribution of X given Y in equation (7). For more details on parameter estimation and other properties of the BQNBD, refer to Li at el. (2010).

Value at risk (VaR) is a standard risk measurement in the banking industry. It is defined as a $(1-\alpha)\%$ quintile of an aggregate loss distribution. In this paper, the aggregate loss distribution is obtained by Monte Carlo simulation. Mathematically, the aggregate loss (AL) may be written as

$$AL = \sum_{i=1}^{N} X_i$$

where N is a random number measuring the frequency of losses and X_i are loss severities. If the probability of density functions (PDF) of the severity distributions of X and Y are f_x, f_y, the procedure for generating the aggregate loss distribution of X and Y under the BQNBD dependency model is below:

1. Generate a random number from equation (5), denoted by N_y. (To obtain annual frequency from monthly frequency, one needs to repeat step 1 to generate 12 random numbers from equation [5] and add them up.)
2. Generate a random number, denoted by N_x, from equation (7) given $y = N_y$. (To obtain annual frequency from monthly frequency, one needs to repeat step 2 to generate 12 random numbers from equation [7] and add them up.)

3. Generate number of N_y random numbers from f_y and N_x random numbers from f_x. The summation of all these random numbers represents the total loss of *X* and *Y* in a given time period.
4. Repeat step 1 to 3 a certain number of times (for example, 1 million times) to form the aggregate loss distribution of *X* and *Y*.

Generating a random number from equation (5) can be done with a lookup table. Specifically, given estimated parameters, one can calculate the probability of $P(X = i), i = 1, 2, ...,$ and accumulative probability of $P(X \le i) = \sum_{j=0}^{i} P(X = j)$ from equation (5). Then for any random number *p* generated from a uniform distribution, the random number corresponding to the number *p* from equation (5) is the number *x* such that $\sum_{j=0}^{x} P(X = j) \le p < \sum_{j=0}^{x+1} P(X = j)$. A random number from equation (7) can be generated in a similar way.

The random number generators for most commonly used distributions including normal, lognormal, uniform, Weibull, gamma, Poisson, and binomial and negative binomial distributions are embedded in commonly used statistical software such as SAS, SPSS and R.

The severity of *X* and *Y* could be empirical distributions or of any analytic form estimated from data. For illustration purposes and for the sake of confidentiality and simplicity, we arbitrarily choose the severity distributions of *X* and *Y*, which are presented in table 1. The distributions include lognormal, exponential and Weibull for both *X* and *Y*, and the last case is

Weibull distribution for X and lognormal distribution for Y. By following the aggregate loss distribution generating procedure, we obtain the aggregate loss distribution of X and Y. Meanwhile, we keep the intermediate results of the procedure and obtain marginal aggregate loss distribution of X and Y respectively. (In practice, one may obtain the aggregate loss distribution of X or Y by directly employing its marginal frequency model and severity model estimated from the original data. However, our study shows the diversification effect is similar.) The capital charges are taken as a certain percentile of the aggregate loss distributions. Table 1 presents the comparison results of capital charges under perfect correlation and dependence structure of the BQNBD. The perfect correlation refers to all severe operational risk losses occurring simultaneously and systematically in the same time period (e.g., one year). The capital charge under this circumstance is the summation of certain (e.g., 99.9^{th}) percentiles for each loss type. Diversification effect (DE) is defined as a proportion of the capital charge reduction due to loss dependence as opposed to the capital charge under perfect correlation assumption.

TABLE 1
VaRs Assuming Perfect Correlation and Diversification Effects

PDF	$f_x = \frac{1}{x\sqrt{2\pi\lambda}} e^{-\frac{(\ln(x)-\theta)^2}{2\lambda^2}}, x > 0$ $\theta = 16, \lambda = 2.5$			$f_x = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, x > 0$ $\lambda = 0.01$		
	$f_y = \frac{1}{y\sqrt{2\pi\lambda}} e^{-\frac{(\ln(y)-\theta)^2}{2\lambda^2}}, y > 0$ $\theta = 18, \lambda = 2.3$			$f_y = \frac{1}{\lambda} e^{-\frac{y}{\lambda}}, x > 0$ $\lambda = 0.02$		
	VaR95%	VaR99%	VaR99.9%	VaR95%	VaR99%	VaR99.9%
X	8.03E+10	1.73E+11	5.45E+11	22143.8	24371.3	26982.0
Y	7.57E+10	1.65E+11	5.38E+11	10375.7	11428.0	12683.4
Perfect correlation	1.56E+11	3.38E+11	1.08E+12	32519.5	35799.3	39665.4
Dependence	1.4E+11	2.75E+11	8.06E+11	31759.6	34673.2	38082.5
Diversification effect	10.2%	18.5%	25.6%	2.3%	3.1%	4.0%
PDF	$f_x = \exp(-(\frac{x}{\lambda})^{\alpha})\frac{\alpha}{\lambda}(\frac{x}{\lambda})^{\alpha-1}, x > 0$ $\alpha = 0.5, \lambda = 1000$			$f_x = \exp(-(\frac{x}{\lambda})^{\alpha})\frac{\alpha}{\lambda}(\frac{x}{\lambda})^{\alpha-1}, x > 0$ $\alpha = 0.1, \lambda = 1000$		
	$f_y = \exp(-(\frac{y}{\lambda})^{\alpha})\frac{\alpha}{\lambda}(\frac{y}{\lambda})^{\alpha-1}, y > 0$ $\alpha = 0.5, \lambda = 1200$			$f_y = \frac{1}{y\sqrt{2\pi\lambda}} e^{-\frac{(\ln(y)-\theta)^2}{2\lambda^2}}, y > 0$ $\theta = 16, \lambda = 3$		
	VaR95%	VaR99%	VaR99.9%	VaR95%	VaR99%	VaR99.9%
X	482019.4	550800.9	635786.4	1.39E+12	8.2E+12	7E+13
Y	544263.5	624131.3	722122.1	3.48E+11	1.01E+12	4.45E+12
Perfect correlation	1026283	1174932	1357909	1.74E+12	9.21E+12	7.45E+13
Dependence	965647.7	1076676	1211723	1.69E+12	8.66E+12	7.05E+13
Diversification effect	5.9%	8.4%	10.8%	2.5%	6.0%	5.3%

Some observations can be drawn from the table above:

1. The severity distribution of *X* and *Y* could be of the same form or a different form, indicating the flexibility of our dependence model.

2. The capital charge reduction rate varies with the choice of severity distributions. Heavier tail severity distributions often lead to a higher diversification effect.

3. The diversification effect becomes higher as risk measurement moves to the tail of the loss distributions. The diversification effect could be as high as more than 25 percent at 99.9^{th} percentile measurement.

Conclusion

The paper details the probabilistic structure of the bivariate quasi-negative binomial distribution. The operational risk diversification effect is illustrated through frequency dependency modeled by the BQNBD under a framework of Monte Carlo simulation. It is shown that the diversification effect in a bivariate case could be as high as more than 25 percent under the framework of the BQNBD dependence model. The method of studying diversification effect is statistically sound but very flexible, easy to extend to multivariate cases and easy to implement in practice.

The BQNBD can be extended to include covariates in the data. Future study will consider regression models based on QNBD and BQNBD for count data.

References

Chernobai, Anna S., Sevtlozar T. Rachev, and Frank J. Fabozzi. 2007. *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis.* New Jersey: John Wiley & Sons Inc.

Consul, P.C. 1989. *Generalized Poisson Distributions: Applications and Properties*. New York: Marcel Dekker Inc.

Genest, C., and J. Neslehova. 2007. "A Primer on Copulas for Count Data." *ASTIN Bulletin* 37: 475-515.

Johnson, N.L., S. Kotz, and N. Balakrishnan. 1997. *Discrete Multivariate Distributions*. New York: John Wiley and Sons Inc.

Li, S. 2007. "The Generalized Lagrangian Distribution: Properties and Application." diss., Central Michigan University.

Li, S., F. Famoye, and C. Lee. 2006. "On Some Extensions of the Lagrangian Probability Distributions." *Far East Journal of Theoretical Statistics* 18: 25-41.

———. 2008. "On Certain Mixture Distributions Based on Lagrangian Probability Models." *Journal of Probability and Statistical Science* 6: 91-100.

Li, S., D. Black, C. Lee, F. Famoye, and S. Li. 2010. "Dependence Models Arising From the Lagrangian Probability Distributions." *Communications in Statistics-Theory and Methods* 39: 1729-42.

Li, S., F. Yang, F. Famoye, C. Lee, and D. Black. 2011. "Quasi-negative Binomial Distribution: Properties and Applications." *Computational Statistics & Data Analysis* (February 18).

Lindskog, F., and A.J. McNeil. 2003. "Common Poisson Shock Models: Applications to Insurance and Credit Risk Modelling." *ASTIN Bulletin* 33, no. 2: 209-38.

McNeil, A.J., R. Frey, and P. Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools.* New York: Princeton University Press.

Nelsen, Roger B. 2006. *An Introduction to Copulas*, 2nd ed. Springer.

Powojowski, M. R., D. Reynolds, and H.J.H. Tuenter. 2002. "Dependent Events and Operational Risk." *Algo Research Quarterly* 5, no. 2: 65-73.

U.S. Office of the Federal Register, National Archives and Records Administration. 2007. "Risk-Based Capital Standards: Advanced Capital Adequacy Framework—Basel II; Final Rule." *Federal Register* 72, no. 235: 69408.

Wang, S. S. 1998. "Aggregation of Correlated Risk Portfolios: Models and Algorithms." accessed July 2008. www.casact.org/pubs/proceed/proceed98/980848.pdf.