An alternative frequency dependence model and its applications[*]

Shubiao Li[1], Dennis Black[1], Fang Yang[2], Felix Famoye[3]and Carl Lee[3]
[1]Comerica Loan Center, Comerica Bank, Auburn Hills, Michigan, USA
[2]JPMorgan Chase & Co, New York, USA
[3]Department of Mathematics, Central Michigan University, Mt. Pleasant, Michigan, USA

Key Words and Phrases: bivariate quasi-negative binomial distribution; marginal distribution; zero-inflation; tail property; dependence; operational risk;

Abstract
In this paper, a multivariate quasi-negative binomial distribution (MQNBD) is proposed to model frequency dependence among different risk types. The property of the marginal distribution of the MQNBD distribution is studied. It is shown 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. Application results show that the marginal distribution is extremely suitable for operational risk or insurance claims where the data is highly skewed, has heavy tails or excessive numbers of zeros; the operational risk diversification effect is illustrated through frequency dependency modeled by the bivariate quasi-negative binomial distribution under framework of Monte Carlo simulation.

1. Introduction

In banking industry, three types of dependence --- loss (severity) dependence, frequency dependence and aggregate loss dependence--can be observed in operational risk loss data (Anna, Chernobai, et al.,2007). Under the advanced measurement approaches (AMA) of Basel II guidelines, operational risk capital charge calculation 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” *(Risk-Based Capital Standards, 2007, pp 69408)*

---

[*] The ideas in this paper represent those of the authors and not of any institution. Contact email: sli@comerica.com

There are many studies prevalent in literature 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 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 detail review on Copulas and dependence can be found in McNeil et al. (2005) and Nelsen (2006).

Frequency dependence occurs when different type of risks shares 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 due to the fact that 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, Chapter 37)]. Lindskog (2001) 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 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 model 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, the marginal distribution, the quasi-negative binomial distribution (QNBD), is studied in section 3. A zero-inflated quasi-negative binomial distribution (ZIQNBD) is defined in section 4. The application of the QNBD and ZIQNBD is presented in section 5. The risk diversification effect is illustrated through frequency dependency modeled by the BQNBD under framework of Monte Carlo simulation in section 6.

2. Bivariate quasi-negative binomial distribution and its probabilistic structure

Li et al. (2006, 2007) defined the class of generalized Lagrangian distributions, in which an extra parameter (the Lagrangian expansion point) was brought into its probability mass function. Let $f(z)$ and $g(z)$ be analytic functions,

$D^{x-1}\Big[\ g(z)\ ^{x}\ f'(z)\Big]_{z=0} \geq 0$ 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

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

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 probability mass function (pmf) of the generalized Lagrangian distribution is

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

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 $0 < \lambda < 1$, $\theta > 0$ and $t > 0$.

Let $X_i(t)$ ($i=1,2,...,m$) 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 some 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 $f_i(z)=e^{\theta_i z}$, $g_i(z)=e^{\lambda_i z}$, $i=1,2$ and suppose 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 (BQNBD). More details in generating multivariate distributions including multivariate quasi-negative binomial distribution (MQNBD), one can refer to Li et al. (2010).

Bivariate random variables (*X, Y)* are said to have a BQNBD if its *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 $\varepsilon_i \geq 0, \delta_i \geq 0, \alpha > 0, i=1,2$.

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

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

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

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

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

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

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

Which 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

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

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. Some properties of the marginal distribution of the BQNBD

The validity of equations (3)-(9) is proved by Li et al. (2010). The probabilistic structure of the BQNBD is also clear. Before we move to frequency dependence study, we first investigate the properties of the marginal distribution (QNBD) and demonstrate that it has ability to modeling operational frequency or insurance claims of a single type of risk, where the data often is highly skewed, has heavy tails or excessive numbers of zeros.

For the sake of simplicity, we suppress scripts in equation (4) and (5) and formally give the definition of the QNBD as below:

A discrete random variable *X* is said to have a QNBD if its *pmf* (Li et al. 2008) is given by

$$P(X = x) = \begin{cases} \dfrac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)} \dfrac{1}{1+cx} \left(\dfrac{1+cx}{1+b+cx} \right)^x \left(\dfrac{b}{1+b+cx} \right)^\alpha, & x = 0,1,2,... \\ 0, & \text{for } x > m \text{ if } c < 0, \end{cases} \tag{10}$$

where $\alpha > 0$, $b > 0$, $m \geq -1/c$, and *m* is the largest positive integer for which 1 + *mc* ≥ 0. The probability of success for QNBD in (10) is $(1+cx)/(1+b+cx)$, which depends on the random variable *X* and parameters *b* and *c*. For the QNBD, the probability of zero class is $P(X = 0) = [b/(1+b)]^\alpha$ and the probability of at least one event occurring is $1-[b/(1+b)]^\alpha$. The QNBD in (10) reduces to the negative binomial distribution (NBD) when the parameter $c = 0$. This additional parameter c of QNBD can be used to describe the pattern of non-zero count.

3.1 The upper tail property of QNBD

The first property of QNBD investigated is the upper tail behavior of the distribution. Two functions $h(x)$ and $\omega(x)$ are said to have the same order, denoted by $h(x) \approx \omega(x)$, if the limit of $h(x)/\omega(x)$ as x approaches infinity is a non-zero constant k. That is, $\lim_{x\to\infty} h(x)/\omega(x) = k \neq 0$, where k is a constant (Walter, 1976). If the function $h(x)$ is pmf of a random variable, then the upper tail behavior of $h(x)$ can be studied through $\omega(x)$. The following theorem shows that the upper tail of QNBD when $c > 0$ has the same order as the function $\omega(x) = x^{-2}$.

Theorem 1: Let $h(x)$ be the pmf of QNBD in (3) and let $\omega(x) = x^{-2}$, then $h(x)$ and $\omega(x)$ have the same asymptotic order when $c > 0$.

Proof: By using the asymptotic (Stirling's) approximation $x! \approx \sqrt{2\pi}x^{x+1/2}e^{-x}$ (Consul and Famoye 2006), we have $\Gamma(x+\alpha) \approx \sqrt{2\pi}(x+\alpha-1)^{x+\alpha-1/2}e^{-(x+\alpha-1)}$ when x is large. Therefore, by equation (3), the pmf of the QNBD is

$$h(x) = P(X = x) = \frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)}\frac{1}{1+cx}\left(\frac{1+cx}{1+b+cx}\right)^{x}\left(\frac{b}{1+b+cx}\right)^{\alpha}$$

$$\approx \frac{(x+\alpha-1)^{x+\alpha-1/2}e^{-(x+\alpha-1)}}{(1+cx)x^{x+1/2}\Gamma(\alpha)e^{-x}}\left(\frac{1+cx}{1+b+cx}\right)^{x}\left(\frac{b}{1+b+cx}\right)^{\alpha}$$

$$\approx \left(\frac{x+\alpha-1}{x}\right)^{x}\frac{e^{-(\alpha-1)}}{(x+\alpha-1)x^{1/2}(1+cx)\Gamma(\alpha)}\left(\frac{1+cx}{1+b+cx}\right)^{x}\left(\frac{b(x+\alpha-1)}{1+b+cx}\right)^{\alpha}$$

$\square \dfrac{1}{(x+\alpha-1)x^{1/2}(1+cx)\Gamma(\alpha)} e^{-b/c}\left(\dfrac{b}{c}\right)^{\alpha} \square \dfrac{1}{x^{2}} = \omega(x)$, which completes the proof. □

Corollary 1: The mean, the variance and higher moments of the QNBD do not exist when $c > 0$.

Proof: By definition, the mean of the QNBD is $E(X) = \sum_{x=0}^{\infty} xP(X = x)$, which is the sum of a series. But by Theorem 1, $P(X = x)$ can be approximated by $P(X = x) \square x^{-2}$. Therefore, $xP(X = x) \square x^{-1}$, thus, the conclusion follows by the fact that $\sum_{x=1}^{\infty} x^{-1} = \infty$. □

The properties of QNBD in corollary 1 are similar to those of a Cauchy distribution. Cauchy distribution plays an important role in probability and statistical theory. It provides a counter-example to some generally accepted results and concepts in statistics. For example, "Poisson showed that a proof given by Laplace for the large-sample justification for Legendre's principle of least squares through the central limit theorem breaks down as the second term in Laplace's expansion of the characteristic function would not be negligible for large sample size *n*" (Johnson et al. 1994, pp. 298). Similarly, the QNBD when $c > 0$ provides a counter-example for the existence of moments for discrete distributions. This may be particularly useful when data are highly over-dispersed with very long or heavy tail.

3.2 The limiting distribution of QNBD

The next theorem shows that the limiting distribution of QNBD is the GPD as α goes to infinity under certain conditions.

Theorem 2: Suppose the ratio α / b exists when both parameters b and α approach infinity, and let the limiting ratio α / b be θ. Then, the limiting distribution of QNBD as both α and b approach infinity is the restricted GPD with parameters θ and c defined in Consul and Famoye (2006, pp. 182).

Proof: Let $\alpha/(1+b+cx)=\varphi(x)$, then $\lim\limits_{\alpha,b\to\infty}\varphi(x)=\lim\limits_{\alpha,b\to\infty}[\alpha/(1+b+cx)]=\theta$. Thus,

$$\lim_{\alpha,b\to\infty}P(X=x)=\lim_{\alpha,b\to\infty}\frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)}(1+cx)^{x-1}\left(\frac{1}{1+b+cx}\right)^{x}\left(\frac{b}{1+b+cx}\right)^{\alpha}$$

$$=\lim_{\alpha,b\to\infty}\frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)\alpha^{x}}(1+cx)^{x-1}[\varphi(x)]^{x}\left(1-\frac{\varphi(x)(1+cx)}{\alpha}\right)^{\alpha} = \theta^{x}(1+cx)^{x-1}e^{-\theta(1+cx)}/x!,$$ which is the restricted GPD given in Consul and Famoye (2006, pp. 182). The last equation uses the facts that $\lim\limits_{\alpha\to\infty}\left(1-\frac{\varphi(x)(1+cx)}{\alpha}\right)^{\alpha}=e^{-\theta(1+cx)}$, $\lim\limits_{\alpha\to\infty}\frac{(x+\alpha-1)!}{(\alpha-1)!\alpha^{x}}=1$, and $\lim\limits_{\alpha\to\infty}\varphi(x)^{x}=\theta^{x}$. □

Note that when $c = 0$, one can apply Theorem 2 to show the limiting distribution of the negative binomial distribution is the Poisson distribution (Johnson et al. 2005, pp 237).

The parameter c in (10) plays an important role in determining the shape of QNBD. Theorem 1 shows that QNBD has a fat tail when $c > 0$, implying QNBD could fit highly over-dispersed data. The limiting distribution provided in Theorem 2 holds for all possible values of c, which suggests that QNBD can also be applied to both under- and over-dispersed data. This is similar to the GPD. When the parameter c is negative, the QNBD becomes a truncated distribution. The probability in (10) is truncated at a positive integer m for which $1 + mc \geq 0$. Thus, the support for (10) is from 0 to integer part of [–

1/*c*]. In this case, the sum of all probabilities may not be unity. It is impossible to quantify the truncation error analytically for QNBD. Consul and Shoukri (1985) showed that the truncation error of GPD is less than 0.5% when the number of non-zero probability classes $m \geq 4$ by means of simulation. Since GPD is a limiting distribution when the limit of the ratio α / b exists, it suggests that the truncation error of QNBD would become smaller as m increases. Further study including computational evaluation will be needed for the error analysis for QNBD.

4. Zero-Inflated QNBD

In operational risk and insurance industry, the frequency data is often characterized with an excessive number of zeros and long or heavy tail properties. Common distributions used to fit data with long or heavy tail distributions are either NBD or GPD. However, for the situation with excessive number of zeros, these distributions may fail to fit the zeros properly. The situation of excessive zeros often arises from the results of clustering (Johnson et al. 2005). For instance, in insurance industry, excess zeros may arise when claims near the deductible are not reported to the insurer, as collected claim payments could be less than the increase in future premiums. In practice, one often assumes that data with excessive zeros arises from two or more data generating processes, one for zero counts and the other for non-zero counts. This assumption motivates the development of zero-inflated probability models (Johnson et al. 2005).

Suppose the original distribution with $pmf\;\; f(x) = P(X = x),\; x = 0,1,2,\ldots$, then a zero-inflated model is a mixture model of the original distribution and a degenerate distribution with all probability concentrated at the origin. Therefore, the probability mass function of a zero-inflated model is given as

$$P(X = x) = \begin{cases} \omega + (1-\omega) f(0), & \text{for } x = 0 \\ (1-\omega) f(x), & \text{for } x = 1, 2, \ldots, \end{cases} \tag{11}$$

Based on (11), one can easily define the zero-inflated version of any discrete distribution. If the random variable *X* is a NBD, then (11) is called a zero-inflated negative binomial distribution (ZINBD). If *X* is the QNBD, then (11) gives the zero-inflated QNBD distribution (ZIQNBD). Similar notation can be used for the GPD and the Poisson distributions. The result in (11) allows a negative weight when the data generating process generates less number of zeros than those from the original probability model (Johnson et al. 2005, pp. 351).

5. Application of QNBD and ZIQNBD

The QNBD has three parameters and allows the probability of success to vary depending not only on parameters, but also on the value of the random variable. The NBD is a special case of QNBD and GPD is a limiting distribution of QNBD. To evaluate the performance of QNBD and its zero-inflated version, ZIQNBD, we apply and compare the four distributions (Poisson (PD), GPD, NBD and QNBD) and their zero-inflated versions (ZIPD, ZIGPD, ZINBD and ZIQNBD) to model two data sets from insurance industry. The NLMIXED procedure of the SAS software is used to estimate the model parameters. In finding the maximum likelihood estimates for the parameters, one can use the default values of 1.0 in PROC NLMIXED in SAS as initials. However, quite often these initials do not work. A better initial is to assume that the data is negative binomial distributed. With this assumption, one can use the NBD moment estimates as initials for the parameters α and *b*, and 0 as the initial for parameter *c*.

The performance of the model fitting are compared by using chi-squared goodness-of-fit statistic, maximized log-likelihood statistic (MLL), Akaike Information Criteria (AIC), and Bayesian Schwartz Criteria (BIC). All these statistics are reported by SAS along with the standard errors of the parameter estimates. Among the eight models considered for each data, we only present the four best models with smallest chi-square and AIC statistics. BIC and MLL statistics are also computed during our study. We found that all the results are consistent. For the sake of brevity, we skip reporting some of these other statistics in the tables. The standard errors of the parameter estimates are listed in parentheses. We deliberately choose not to combine the adjacent classes when the expected frequency for the chi-square is below 5. This will not have a serious effect on our comparison since the same method is used for all models considered in the analysis.

5.1 Motor insurance data set

Table 1 presents automobile insurance claim count data from Yip and Yau (2005). Yip and Yau retrieved the dataset from the SAS Enterprise Miner database and used 2812 complete records.

The data has a sample mean of 0.82 claims and a sample variance of 1.36. Clearly over-dispersion is present. It is also clear there is a spike for the count of zero, which represents claim-free policyholders. This is a common phenomenon in insurance applications. Due to experience rating schemes implemented by insurers that rewards claim-free and penalizes claim violation activities, policyholders have a tendency not to report small claims to avoid being penalized by insurer's experience rating scheme. As a consequence, there is excessive number of zero claims.

Our results show that zero-inflated models outperform the corresponding non-zero-inflated models. This is not surprising due to excessive number of zero claims. Of all the eight models (PD, GPD, NBD, QNBD and their zero-inflated models), ZIQNBD performs the best with the smallest chi-square value of 0.18 (*p*-value = .6714). The asymptotic Wald-test shows that the zero-inflated parameter estimate $\hat{\omega}$ is statistically significant, which suggests the data is zero-inflated. If we do not consider zero-inflated versions, then QNBD performs the best, but the fit is far from adequate (a chi-square value of 87.95 with *p*-value .0000).

Table 1: Automobile insurance claim counts

		Expected Frequencies			
No. of claim	Observed	QNBD	ZIQNBD	ZIPD	ZIGPD
0	1706	1681.81	1706.00	1706.00	1706.00
1	351	483.46	349.77	422.96	338.35
2	408	314.26	411.98	357.38	434.00
3	268	217.83	263.55	201.31	256.68
4	74	109.98	75.92	85.05	69.45
5	5	4.65	4.78	39.30	7.52
Total	2812	2812.00	2812.00	2812.00	2812.00
		$\hat{\alpha} = 0.2700$ (0.0136) $\hat{b} = 0.1750$ (0.0135) $\hat{c} = –0.1900$ (0.0024)	$\hat{\alpha} = 4.5722$ (2.3201) $\hat{b} = 1.6495$ (0.8865) $\hat{c} = –0.1605$ (0.0119) $\hat{\omega} = 0.5556$ (0.0151)	$\hat{\theta} = 1.6899$ (0.0449) $\hat{\omega} = 0.5177$ (0.0123)	$\hat{\theta} = 2.5022$ (0.0992) $\hat{\lambda} = –0.3152$ (0.0325) $\hat{\omega} = 0.5716$ (0.0107)
χ^2		87.95	0.18	72.88	3.67
p-value		0.0000	0.6714	0.0000	0.1596
AIC		6704.8	6610.9	6699.2	6612.8
MLL		–3349.40	–3302.62	–3347.60	–3303.40

An interesting observation here is the estimated truncation point of ZIQNBD, $[-1/c]$ is very close to 6, which is slightly greater than 5, the maximum number of observed claim counts in the dataset. In a particular insurer's system, the policy usually sets an upper bound on the number of claims. Insurers are allowed to cancel insurance policies with high-risk customers who have claim counts exceeding a threshold. Even if insurers do not cancel the policies, high-risk customers generally face such a high premium from experience rating that they do not want to stay with current insurer. ZIQNBD seems to handle the truncated count dataset well.

5.2 Motor vehicle records

Table 2 displays the number of violation points on the motor vehicle records from Flynn and Francis (2009). Again, this is a skewed distribution with a spike at zero. Our results show that zero-inflated models outperform non-zero inflated versions of the models and the four best models are QNBD, ZIQNBD, ZINBD and ZIGPD. The chi-square statistics with their corresponding *p*-values suggest that only ZIQNBD is an appropriate model for the dataset. The ZIQNBD has particularly done very well in fitting the tail part of the data.

Table 2: Number of violation points on motor vehicle records

No. of points	Observed	Expected Frequencies			
		QNBD	ZIQNBD	ZINBD	ZIGPD
0	4659	4556.71	4659.00	4659.00	4659.00
1	1467	1798.58	1451.49	1363.19	1346.66
2	1199	1172.91	1225.77	1314.79	1326.98
3	966	848.98	969.82	1041.87	1053.09
4	727	632.45	725.34	735.96	738.88
5	528	468.72	512.03	481.88	479.55
6	341	337.01	338.89	298.93	295.28
7	213	228.90	208.01	178.09	175.18
8	114	141.73	116.59	102.82	101.13
9	53	75.70	58.40	57.88	57.21
10	20	31.69	25.36	31.92	31.86
11	13	8.62	9.15	17.31	17.53
12	1	0.99	2.57	9.25	9.55
13	2	0.01	0.57	10.11	11.10
Total	10303	10303.00	10303.00	10303.00	10303.00
		$\hat{\alpha}$ = 0.4232 (0.0099) $\hat{b}$ = 0.1702 (0.0060) $\hat{c}$ = –0.0698 (0.0012)	$\hat{\alpha}$ = 1.2017 (0.1164) $\hat{b}$ = 0.3818 (0.0386) $\hat{c}$ = –0.0538 (0.0036) $\hat{\omega}$ = 0.3038 (0.0140)	$\hat{\alpha}$ = 3.3032 (0.2179) $\hat{b}$ = 1.2308 (0.0741) $\hat{\omega}$ = 0.3628 (0.0074)	$\hat{\theta}$ =2.0311 (0.0418) $\hat{\lambda}$ =0.2511 (0.0099) $\hat{\omega}$ = 0.3695 (0.0069)
χ^2		517.70	9.26	61.93	73.28
p-value		0.0000	0.4136	0.0000	0.0000
AIC		36024.0	35892.0	35955.0	35967.0
MLL		–18008.95	–17942.23	–17974.48	–17980.57

In Tables 1 and 2, the ZIQNBD is the best model for analyzing these count data sets. We now compare the ZIQNBD with QNBD and NBD which are nested within ZIQNBD. The ZIQNBD reduces to the QNBD when $\omega = 0$ and it reduces to ZINBD when $c = 0$. The ZIQNBD reduces to the NBD when both parameters ω and c are equal to zero. To compare two nested models, we use the test statistic

$$\lambda = -2\log(L_s / L_c)\,, \tag{12}$$

where L_c is the maximized likelihood for the "full" or "complete" (or larger) model and L_s is the maximized likelihood of the nested (or smaller) model. The statistic in (12) has an asymptotic chi-square distribution with degrees of freedom *k*, where *k* is the difference between the numbers of estimated parameters for the two models. The value of *k* is 1 when ZIQNBD is compared with either ZINBD or QNBD and it is 2 when ZIQNBD is compared with NBD. In Table 3, we compare the ZIQNBD to the QNBD and NBD models.

Table 3: Maximized log-likelihood statistic (MLL) and *p*-values for comparing ZIQNBD with ZINBD, QNBD and NBD

	Data in Table 1		Data in Table 2	
Model	MLL	*p*-value	MLL	*p*-value
ZIQNBD	–3302.62		–17942.23	
ZINBD	*		–17974.48	0.00
QNBD	–3349.40	0.00	–18008.95	0.00
NBD	–3500.98	0.00	–18272.63	0.00

* ZINBD did not converge for the data

The results in Table 3 show that the ZIQNBD outperforms both the QNBD and NBD for the two data sets. For the Automobile insurance claim counts (Table 1), the ZINBD did not converge. In Tables 1 and 2, both parameters *c* and ω are significantly different from 0. Thus, the ZIQNBD model should be applied to fit the two data sets.

6. Frequency dependence and diversification effect of operational risk

Section 3 to 5 demonstrate that the marginal distribution QNBD of the BQNBD provides a better model in fitting count data characterized by skewed, fat tail and excessive number of zeros. This section moves to dependence model and diversification effect study. 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, $\bar{x}$ = 14.48, s_x = 7.88 and $\bar{y}$ = 13.52, s_y = 7.80. The Pearson correlation coefficient between the two risk types is 0.45 .Clearly, both X and Y exhibits 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 parenthesis 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). More details on parameter estimation and other properties of the BQNBD, one can refer to Li at el. (2010).

Value-at-risk (*VaR*) is a standard risk measurement in 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. Suppose 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 to 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 by 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 an 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 distribution, lognormal distribution, uniform distribution, weibull distribution, gamma distribution, Poisson distribution, binomial and negative binomial distribution are embedded in most commonly used statistical software such as SAS, SPSS, R etc..

The severity of *X* and *Y* could be empirical distributions or of any analytic form estimated from data. For illustration purpose and for the sake of confidentiality and simplicity, we simply arbitrarily choose the severity distributions of *X* and *Y*, which are presented in table 4. The distributions include lognormal, exponential, weibull distribution for both *X* and *Y* and the last case is weibull for *X* and lognormal 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 certain percentile of the aggregate loss distributions. Table 4 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 occur 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. 99.9th) 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 4: 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 table 4:

(1) The severity distribution of *X* and *Y* could be of the same form or 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 move to the tail of the loss distributions. The diversification effect could be as high as more than 25% at 99.9th percentile measurement.

7. Conclusion

The paper details the probabilistic structure of the bivariate quasi-negative binomial distribution (BQNBD). The property of the marginal distribution (QNBD) of the BQNBD distribution is studied. Zero-inflated model of the QNBD is defined. It is shown 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. Both theoretic implication and application results show 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; the operational risk diversification effect is illustrated through frequency dependency modeled by the BQNBD under framework of Monte Carlo simulation. It is shown that the diversification effect in bivariate case could be as high as more than 25% 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 case (MQNBD) and easy-to-implement in practice.

The QNBD and 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

Anna, S., Chernobai sevtlozar T., Rachev Frank, J. F. (2007) *Operational Risk: A guide to Basel II Capital Requirements, Models, and Analysis.* John Wiley & Sons, Inc., New Jersey.

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

Consul, P.C. and Famoye, F. (2006) *Lagrangian Probability Distributions*. Birkhäuser, Boston.

Consul, P.C. and Shoukri, M.M. (1985) The generalized Poisson distribution when the sample mean is larger than the sample variance. *Communications in Statistics-Simulation and Computation*, 14: 667-681

Flynn, M. and Francis, L.A. (2009). More flexible GLMs zero-inflated models and hybrid models. *Casualty Actuarial Society E-Forum, Winter 2009: 148-224*

Genest, C., Neslehova, J. (2007) A primer on copulas for count data, *ASTIN Bulletin* 37, 475-515.

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

Johnson, N.L., Kemp, A.W. and Kotz, S. (2005) *Univariate Discrete Distributions (3rd Edition),* Wiley-Interscience, New York

Johnson, N.L, Kotz, S. and Balakrishnan, N. (1994) *Continuous Univariate Distributions, Vol. 1,* Wiley-Interscience, New York

Li, S. (2007) The generalized Lagrangian distribution: Properties and application, Dissertation, Central Michigan University

Li, S., Famoye, F. and Lee, C. (2006) On some extensions of the Lagrangian probability distributions. *Far East Journal of Theoretical Statistics,* 18: 25-41

Li, S., Famoye, F. and Lee, C. (2008) On certain mixture distributions based on Lagrangian probability models. *Journal of Probability and Statistical Science*, 6: 91-100.

Li, S., Black, D.,Lee, C., Famoye, F., and Li, S. (2010) Dependence models arising from the Lagrangian probability distributions. *Communications in Statistics-Theory and Methods, 39*: 1729 – 1742.

Lindskog F., McNeil. A.J. (2003) Common Poisson shock models: applications to insurance and credit risk modelling. *ASTIN Bulletin*, 33(2), 209-238.

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

Nelsen, Roger B. (2006), *An Introduction to Copulas*, Second Edition, Springer.

Powojowski, M. R., Reynolds, D., Tuenter, H.J.H. (2002) Dependent events and operational risk. *Algo Research Quarterly,* Vol.5, No. 2, 65-73.

Risk-Based Capital Standards: Advanced Capital Adequacy Framework—Basel II; Final Rule (2007), Vol. 72, No. 235, p. 69408

Walter, R. (1976) *Principles of Mathematical Analysis (3rd Edition)*, McGraw-Hill, New York

Wang, S. S. (1998) Aggregation of correlated risk portfolios: Models and algorithms, *Retrieved in July 2008 from* www.casact.org/pubs/proceed/proceed98/980848.pdf

Yip, K.C.H. and Yau, K.K.W. (2005) On modeling claim frequency data in general insurance with extra zeros, *Insurance: Mathematics and Economics,* 36(2): 153-163

.