Mathematical Statistics
Stockholm University

Aggregating sectors in the infectious defaults model

Ola Hammarlid

Research Report 2003:11

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2003:11**,
http://www.math.su.se/matstat

Aggregating sectors in the infectious defaults model

Ola Hammarlid*

September2003

Sammanfattning

How to model the dependence between defaults in a portfolio subject to credit risk is a question of great importance. The Infectious Default model by Davis and Lo offers a way to model the dependence. Every company defaulting in this model may 'infect' another company to default. An unsolved question however, is how to aggregate independent sectors, since a naive straightforward computation quickly gets cumbersome, even when homogeneous assumptions are made. Here, two algorithms are derived that overcome the computational problem and further make it possible to use different exposures and probabilities of default for each sector. For an 'outbreak' of defaults to occur in a sector, at least one company has to default by itself. This fact is used in the derivations of the two algorithms. The first algorithm is derived from the probability generating function of the the total credit loss in each sector and the fact that the outbreaks are independent Bernoulli random variables. The second algorithm is an approximation and is based on a Poisson number of outbreaks in each sector. This algorithm is less cumbersome and more numerically stable, but still seems to work well in a realistic setting.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: olah@math.su.se

1. Introduction

To capture the aggregated credit loss of a portfolio in a realistic way, the dependence of defaults between the entities has to be considered. For example, in order to measure credit risk in a portfolio a quantile is usually used, Value at Risk, which relies heavily on the dependence. For contracts such as Collateralized Debt Obligations (CDO), the pricing of the super senior tranche depends mostly upon the dependence. A tranche is an interval of the total credit loss, where the investor is risking his capital. The super senior tranche is the last tranche with the best rating.

The dependence between defaults is a well known empirical feature which has been modelled in a number of ways. Moody's Binomial Expansion Technique (BET) model the dependency of defaults by lowering the number of contracts, called a diversity score, and raising the exposure such that the expected credit loss is unchanged, Gluck and Remeza [6] and Cifuentes and O'Conner [2]. Rating and diversity score is a common way of describing the properties of a CDO contract in the market.

There are other models incorporating default dependencies, which are usually not straightforward to use in a pricing situation and can easily get cumbersome. These models are based on copulas, shot noise models, Cox processes, structural models etc, see for example Embrechts *et al.* [5] or Bielecki and Rutkowski [1].

Davis and Lo propose a new way of modelling the dependency between defaults, through 'infection' [3]. The idea is that a defaulting company may infect another company to default in the following way: A portfolio subject to credit risk contains n bonds, loans or other types of debt. In sector k, out of K sectors, the portfolio is invested in n_k bonds and $n = \sum_{k=1}^{K} n_k$. When bond i in sector k defaults, it is indicated by,

$$\mathbf{1}_i(k) = \begin{cases} 1 & \text{if default,} \\ 0 & \text{otherwise.} \end{cases}$$

The number of defaults in sector k are $N_k = \sum_{i=1}^{n_k} \mathbf{1}_i(k)$ and the total number of defaults are $N = \sum_{k=1}^{K} N_k$. Every default in a sector gives a credit loss L_k.

Let $X_i(k)$ and $Y_{ij}(k)$ be independent Bernoulli random variables with

$$P(X_i(k) = 1) = p_k \qquad \text{and} \qquad P(Y_{ij}(k) = 1) = q_k.$$

The Bernoulli variable $X_i(k)$ represents spontaneous default without external influence and the Bernoulli variable $Y_{ij}(k)$ is the infection default, where company j's spontaneous default might cause company i to default. The indicator of default,

$$\mathbf{1}_i(k) = X_i(k) + (1 - X_i(k)) \left(1 - \prod_{i \neq j} (1 - X_j(k) Y_{ij}(k)) \right).$$

The defaults between different sectors are independent. Davis and Lo stated and proved the following theorem:

Theorem 1.1 *The probability*

$$P(N_k = m) = \binom{n_k}{m}\Bigg(p_k^m(1-p_k)^{n_k-m}(1-q_k)^{m(n_k-m)} + \sum_{i=1}^{m-1}\binom{m}{i}p_k^i(1-p_k)^{n_k-i}(1-(1-q_k)^i)^{m-i}(1-q_k)^{i(n_k-m)}\Bigg).$$

The expected value and variance in each sector,

$$\begin{aligned} E\left[N_k\right] &= n_k\left(1-(1-p_k)(1-p_kq_k)^{n_k-1}\right)\\ Var(N_k) &= E[N_k] + n_k(n_k-1)\beta_k - (E[N_k])^2 \end{aligned}$$

where

$$\begin{aligned} \beta_k = p_k^2 &+ 2p_k(1-p_k)\left(1-(1-q_k)(1-p_kq_k)^{n_k-2}\right)\\ &+ (1-p_k)^2\left(1-2(1-p_kq_k)^{n_k-2} + (1-2p_kq_k+p_kq_k^2)^{n_k-2}\right). \end{aligned}$$

The continuous time version of this model by Davis and Lo [4] is not considered here.

In order to calculate $P(N = m_{\text{total}})$, we have to sum over all the configurations in each sector that gives m_{total} defaults. This summation can be "colossal even for moderate" K [3]. To get the distribution of the overall credit loss a homogeneous exposure and other homogeneous assumptions on the default probabilities are suggested to make the computations possible.

To be of any interest to practitioners in the financial market a credit risk model must be able to handle realistic portfolios. This means that the model has to handle a free choice of p_k, q_k and exposure L_k for a large K. Therefore, we will tackle the aggregation problem by deriving two algorithms. The first algorithm computes the total credit loss, where heterogeneous assumptions on exposures and probabilities between sectors are easily handled. Furthermore, the distribution of the total number of defaults may be computed by putting all the exposures equal to one. This algorithm is similar to the De Pril's algorithm, see section 2.

To derive the second algorithm we replace the zero or one outbreak assumption with a Poisson number of outbreaks. This enables the derivation of a Panjer type of algorithm, which is less cumbersome than the first algorithm and more numerically stable, see section 3.

Finally we compare the two algorithms applied to a portfolio of loans, that tries to be realistic in a real world setting. The result indicates that the difference can be neglected, see section 4.

2. An algorithm for the original model

In this section we state and prove an algorithm to compute the distribution of the total credit loss S in the portfolio. By an outbreak, we mean that at least one company in a sector defaults. An outbreak may only occur if at least one company defaults spontaneously. The number of outbreaks is either zero or one with probability $P(N_k > 0) = 1-(1-p_k)^{n_k}$.

When an outbreak occurs in sector k, each single default causes an integer valued credit loss of L_k and the total loss of $S_k = N_k L_k$, where $N_k > 0$. Conditioned on an outbreak in a sector, the distribution of the number of defaults,

$$P(N_k = m | N_k > 0) = P(N_k = m) / \left(1 - (1 - p_k)^{n_k}\right), \quad m \geq 1, \tag{1}$$

and the probability of the total credit loss given an outbreak

$$P(S_k = mL_k) = P(N_k = m | N_k > 0). \tag{2}$$

Theorem 2.1 *The probability* $P(S = s)$ *can be computed by*

$$P(S = 0) = \prod_{k=1}^{K} (1 - p_k)^{n_k} \quad \text{and} \quad P(S = s) = \frac{1}{s} \sum_{k=1}^{K} v_k(s),$$

where for $s \geq 1$*,*

$$v_k(s) = \frac{1 - (1 - p_k)^{n_k}}{(1 - p_k)^{n_k}} \sum_{m=1}^{[s/L_k]} P(N_k = m | N_k > 0) \Big(m L_k P(S = s - mL_k) - v_k(s - mL_k) \Big)$$

and $v_k(s) = 0$ *otherwise. The notation* $[s/L_k]$ *is the integer part of* s/L_k*.*

This algorithm can easily handle heterogeneous exposure and default probabilities between sectors, even for a large portfolio, which is an improvement of the scheme suggested by Davis and Lo [3].

We will use the same technique as Rolski *et al.* in proving Theorem 4.4.1 in [7]. In fact, this theorem could, after some rearrangements be used to prove Theorem 2.1. We chose not to do so since no intuition is gained on how the algorithm comes about ‡.

Proof: In sector k no default occurs with probability $(1 - p_k)^{n_k}$. Naturally, the probability generating function of sector k,

$$g_k(t) = (1 - p_k)^{n_k} + \sum_{m=1}^{n_k} P(N_k = m) t^{mL_k}.$$

Let $\theta_k = 1 - (1 - p_k)^{n_k}$, then

$$g_k(t) = 1 - \theta_k + \theta_k \sum_{m=1}^{n_k} P(N_k = m | N_k > 0) t^{mL_k} = 1 - \theta_k + \theta_k \hat{g}_k(t),$$

where $\hat{g}_k(t) = \sum_{m=1}^{n_k} P(N_k = m | N_k > 0) t^{mL_k}$. The independence between sectors gives the probability generating function of the total credit loss

$$g(t) = \prod_{k=1}^{K} g_k(t) = \prod_{k=1}^{K} \left(1 - \theta_k + \theta_k \hat{g}_k(t)\right).$$

By the elementary property of the probability generating function we have that

$$P(S = s) = \frac{g^{(s)}(0)}{s!} = \frac{1}{s!} \frac{\mathrm{d}^s}{\mathrm{d}t^s} g(t) \Big|_{t=0}.$$

‡ Anders Martin-Löf kindly pointed out how to make a detail in an early version of the proof more simple.

Therefore, $P(S=0)=\prod_{k=1}^{K}(1-p_k)^{n_k}$, and

$$g^{(1)}(t)=g(t)\frac{\mathrm{d}\log(g(t))}{\mathrm{d}t}=\sum_{k=1}^{K}\frac{\theta_k\hat{g}_k^{(1)}(t)g(t)}{1-\theta_k+\theta_k\hat{g}_k(t)}=\sum_{k=1}^{K}V_k(t),$$

where $V_k(t)=\theta_k\hat{g}_k^{(1)}(t)g(t)(1-\theta_k+\theta_k\hat{g}_k(t))^{-1}$. Put

$$v_k(s)=((s-1)!)^{-1}V_k^{(s-1)}(0),\quad v_k(0)=0, \tag{3}$$

then

$$P(S=s)=\frac{g^{(s)}(0)}{s!}=\frac{1}{s!}\sum_{k=1}^{K}V_k^{(s-1)}(0)=\frac{1}{s}\sum_{k=1}^{K}v_k(s).$$

The idea now is to compute a recursive formula for $v_k(s)$. It is possible to write

$$(1-\theta_k)V_k(t)=\theta_k\hat{g}_k^{(1)}(t)g(t)-\theta_k\hat{g}_k(t)V_k(t). \tag{4}$$

We expand $V_k(t)$ in terms of $v_k(s)$

$$V_k(t)=\sum_{s=1}^{\infty}t^{s-1}v_k(s). \tag{5}$$

The probability generating function of the credit loss conditioned on an outbreak and the first order derivative,

$$\begin{aligned}\hat{g}_k(t) &= \sum_{m=1}^{\infty}t^{mL_k}P(N_k=m|N_k>0),\\ \hat{g}_k^{(1)}(t) &= \sum_{m=1}^{\infty}mL_kt^{mL_k-1}P(N_k=m|N_k>0).\end{aligned}$$

Use these last two expressions and equation (5) to rewrite equation (4),

$$\begin{aligned}(1-\theta_k)\sum_{s=1}^{\infty}t^{s-1}v_k(s) &= \theta_k\sum_{m=1}^{\infty}\sum_{l=1}^{\infty}mL_kt^{mL_k-1+l}P(N_k=m|N_k>0)P(S=l)\\ &\quad -\theta_k\sum_{m=1}^{\infty}\sum_{l=1}^{\infty}t^{mL_k+l-1}P(N_k=m|N_k>0)v_k(l)\\ &= \{s=l+mL_k\}\\ &= \theta_k\sum_{s=1}^{\infty}t^{s-1}\Big(\sum_{m=1}^{[s/L_k]}mL_kP(N_k=m|N_k>0)P(S=s-mL_k)\\ &\quad -P(N_k=m|N_k>0)v_k(s-mL_k)\Big),\end{aligned}$$

where $\{s=l+mL_k\}$ indicates a change of variable and $[s/L_k]$ is the integer part of s/L_k. By this we se that

$$v_k(s)=\frac{\theta_k}{1-\theta_k}\sum_{m=1}^{[s/L_k]}P(N_k=m|N_k>0)\Big(mL_kP(S=s-mL_k)-v_k(s-mL_k)\Big).$$

□

The number of defaults can easily be computed by $L_k=1$ in theorem 2.1.

Corollary 2.2 *The probability $P(N = m)$ can be computed by*

$$P(N=0) = \prod_{k=1}^{K}(1-p_k)^{n_k} \quad \text{and} \quad P(N=h) = \frac{1}{h}\sum_{k=1}^{K} v_k(h),$$

for $k = 1, \ldots, K$ where

$$v_k(h) = \frac{1-(1-p_k)^{n_k}}{(1-p_k)^{n_k}} \sum_{j=1}^{h} P(N_k = j | N_k > 0)\Big(jP(N = h-j) - v_k(h-j)\Big)$$

and $v_k(h) = 0$ otherwise.

3. An algorithm based on Poisson number of outbreaks

In this section we make a small change of assumption compared to the original model. The change is that the number of outbreaks is assumed be a Poisson random variable Λ_k with intensity

$$\mu_k = 1 - (1-p_k)^{n_k}\,.$$

Hence the expected number of outbreaks and the expected credit loss remain unchanged. An alternative choice of intensity could be $\mu_k = -n_k \log(1-p_k)$. Then the distribution would be an upper limit for the distribution of the original model.

This change to a Poisson number of outbreaks can be justified. A default does not have to mean bankruptcy, or a defaulted credit may be replaced by an equivalent credit before the next infectious outbreak. A mathematical argument for this change is that $g(t) = \prod_{k=1}^{K}(1-\theta_k+\theta_k\hat{g}_k(t)) \approx \exp\left(\sum_{k=1}^{K}\theta_k(\hat{g}_k(t)-1)\right)$, which is the probability generating function of a compound Poisson random variable.

On outbreak $l = 0, \ldots, \Lambda_k$ in sector k, the loss $S_{kl} = N_{kl}L_k$, where N_{kl} is the number of defaults. These numbers of defaults, N_{kl}, are independent over sectors and outbreaks. The distribution of the credit loss on outbreak l is still given by the number of defaults conditioned on an outbreak, equation (2). The total loss in a sector $S_k = \sum_{l=0}^{\Lambda_k} S_{kl}$, and therefore the total credit loss $S = \sum_{k=1}^{K}\sum_{l=0}^{\Lambda_k} S_{kl}$.

Theorem 3.1 *Assume the number of outbreaks Λ_k is Poisson distributed with intensity μ_k. Then, $P(S=0) = \exp\left(-\sum_{k=1}^{K}\mu_k\right)$ and*

$$P(S=s) = \frac{1}{s}\sum_{k=1}^{K}\sum_{m_k=1}^{[s/L_k]} \mu_k m_k L_k P(N_k = m_k L_k | N_k > 0) P(S = s - m_k L_k).$$

This is a Panjer type of algorithm. The Panjer algorithm is used in insurance and by CreditRisk+ model by Credit Suisse First Boston, but in a different setting, without dependencies through infection.

Proof: The event that any company default in the portfolio is a Poisson random variable with intensity $\sum_k \mu_k$. The probability of no default is therefore,

$$P(S=0) = \exp\left(-\sum_{k=1}^{K} \mu_k\right).$$

The moment generating function of the loss on an outbreak

$$\tilde{M}_k(\gamma) = \sum_{m=1}^{n_k} \mathrm{e}^{m\gamma L_k} P(N_k = m | N_k > 0).$$

Remember that the moment generating function of a Poisson random variable, $M_{\Lambda_k}(\gamma) = \exp(\mu_k(\gamma - 1))$. Therefore, when we condition on Λ_k,

$$M_k(\gamma) = E\left[E\left[\exp\left(\gamma L_k N_{kl}\right) | \, N_{kl} > 0\right]^{\Lambda_k}\right] = \exp\left[\mu_k(\tilde{M}_k(\gamma) - 1)\right].$$

The assumed independence between sectors gives the moment generating function of the total credit loss,

$$M_S(\gamma) = E\left[\mathrm{e}^{\gamma \sum_{k=1}^{K} S_k}\right] = \exp\left(\sum_{k=1}^{K} \mu_k(\tilde{M}(\gamma) - 1)\right).$$

The first order derivative of $M_S'(\gamma) = M_S(\gamma) \sum_{k=1}^{K} \mu_k \tilde{M}_k'(\gamma)$, where

$$M_S'(\gamma) = \sum_{s=0}^{\infty} \mathrm{e}^{\gamma s} s P(S = s). \tag{6}$$

and also

$$\begin{aligned} M_S'(\gamma) &= \sum_{j=0}^{\infty} \sum_{k=1}^{K} \sum_{m_k=1}^{\infty} \mathrm{e}^{\gamma(j + m_k L_k)} \mu_k m_k L_k P(S_{kl} = m_k L_k) P(S = j) \\ &= \{s = j + m_k L_k\} \\ &= \sum_{s=0}^{\infty} \mathrm{e}^{\gamma s} \sum_{k=1}^{K} \sum_{m_k=1}^{[s/L_k]} \mu_k m_k L_k P(N_k = m_k | N_k > 0) P(S = s - m_k L_k). \end{aligned}$$

The notation $\{s = j + m_k L_k\}$ stands for the change of variables. If we compare this last expression with (6) we see that

$$sP(S = s) = \sum_{k=1}^{K} \sum_{m_k=0}^{s} \mu_k m_k L_k P(N_k = m_k | N_k > 0) P(S = s - m_k L_k), \quad s > 1.$$

□

If we put $L_k = 1$ in theorem 3.1, we get the distribution of the number of total defaults in the portfolio.

Corollary 3.2 *Assume the number of outbreaks in a sector Λ_k is Poisson distributed with intensity μ_k. Then, $P(N = 0) = \exp\left(-\sum_{k=1}^{K} \mu_k\right)$ and*

$$P(N = h) = \frac{1}{h} \sum_{k=1}^{K} \sum_{m_k=1}^{h} \mu_k m_k P(N_k = m_k | N_k > 0) P(N = h - m_k).$$

Table 1. Reference portfolio

Sector	n_k	p_k	q_k	**Adjusted** p_k	L_k
Aerospace & Defense	3	1%	5%	0.9%	5
Automobile	3	2%	5%	1.8%	3
Banking	10	0.1%	10%	0.1%	10
Broadcasting	3	2%	5%	1.8%	2
Buildings & Real Estate	12	1%	5%	0.6%	9
Electronics	6	1%	5%	0.8%	2
Entertainment	4	2%	5%	1.7%	5
Finance	15	0.5%	10%	0.4%	8
Food & Tobacco	2	1%	5%	1.0%	4
Health care	3	1%	5%	0.9%	7
Insurance	6	1%	10%	0.7%	5
Mining, Metals	4	2%	5%	1.7%	3
Oil & Gas	2	3%	5%	2.9%	3
Printing & Publishing	4	1%	5%	0.9%	2
Telecommunications	15	4%	5%	2.6%	12
Utilities	11	3%	5%	2.0%	6

4. A comparison between the two algorithms and discussion

In this section we will compare the two algorithms applied on a credit portfolio. The quantities of the reference portfolio, given in table 1, try to be realistic for a loan portfolio. The investment horizon in the case of regulatory capital is one year, and for CDO usually five years. Usually, investment graded bonds should have a default probability of less than one percent each year. This may, however, vary greatly over time. The reference portfolio could therefore be on both one and five years, where the one year portfolio would have more speculative quantities.

From a regulatory capital perspective, the probabilities $P(S \geq s)$ are of interest. These probabilities are easily derived in both of the algorithms by

$$P(S \geq 0) = 1 \quad \text{and} \quad P(S \geq s) = P(S \geq s-1) - P(S = s).$$

To be able to understand the dependence effect through infections, the expected number of defaults needs to be fixed, see theorem 1.1. Therefore, we adjust the spontaneous probability p of default. The adjusted probabilities in the table are computed such that the expected values of the sectors remain unchanged compared to the case of no infection probability. The adjusted probabilities are found through the Newton-Raphson method.

The mean value credit loss in the reference portfolio is 15. The difference between the two models is presented in graphical form, see figure 1 and 2.

From a regulatory capital perspective, it is the quartiles 95-99.9 % that are of interest, which means $-7 \leq \log(P(S \geq s)) \leq -3$. In this interval the distribution of the Poisson model dominates the true distribution, since we are far to the right of the expected value, where the Poisson arrival of infections makes a difference, see figure 2.

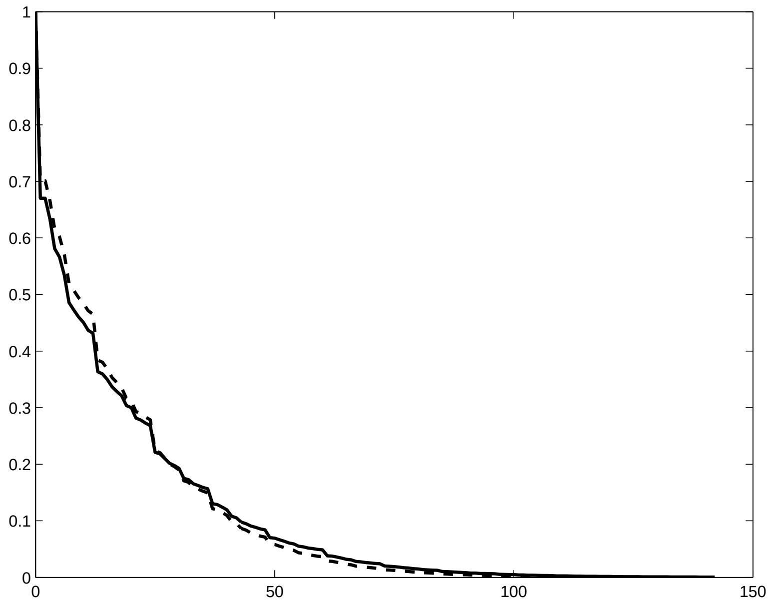

Figure 1. The difference between the Poisson approximation algorithm (solid line) and the exact solution (dashed line) for $P(S \geq s)$.

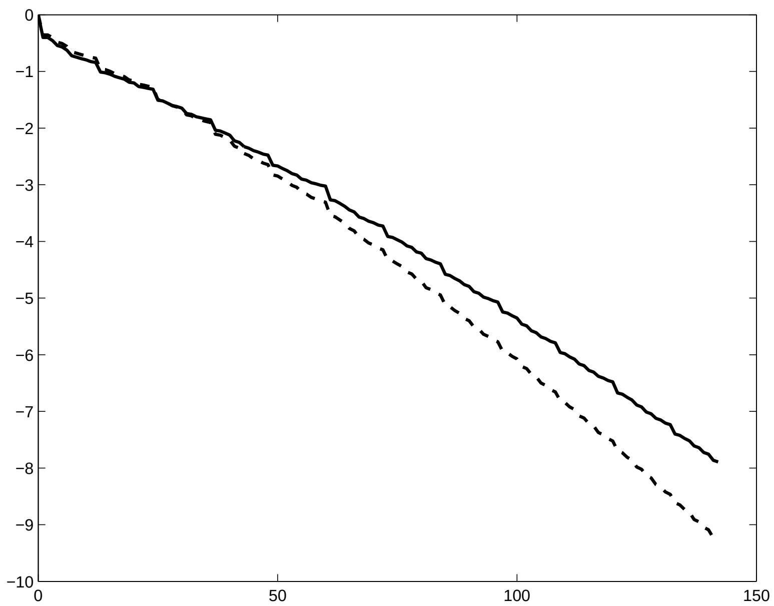

Figure 2. The difference between the Poisson approximation algorithm (solid line) and the exact solution (dashed line) on log scale, $\log(P(S \geq s))$.

A practical problem is that losses can not be too large in absolute numbers, since this may cause numerical problems in the computations. This however, may always be solved by scaling, but with the loss of the exact distribution. For example, if $L_k = 113$ would cause numerical problems, then a scaling to 11 or 1 might solve the problems, but we lose the true distribution. The exact distribution can however easily be 'sandwiched' between an upper and a lower bound after scaling.

The difference between the two algorithms are of less importance since the estimation uncertainty of the probabilities of default and the probabilities of infection in each section are probably of greater magnitude. From a regulatory perspective the Poisson algorithm is an upper bound for small probabilities and could therefore be used as well. However, the De Pril type of algorithm needs to compute and store $K + 1$ vectors compared to the Poisson algorithm that only has to store the probability vector. Also in favor of the Poisson algorithm is the fact that it only adds positive quantities which is numerically more stable.

A default in the infectious default model is something bad, since the defaulting company might infect another company to default. Not all practitioner's agree about this. A default might leave the remaining companies in a better position to avoid default. This is probably best modelled by a decrease in probability to default, but this is another story.

A portfolio in one sector can be very differentiated in exposure. The homogeneous assumption of loss given default L_k, in a sector k, might therefore be to restrictive in practise. This might be solved by how the sectors are defined, such that a sorting by exposure is also made. This however frees the dependence that was assumed to be there in the first place. How to define sectors is a tricky question in practice, especially when the number of companies in the portfolio is small.

Acknowledgments

I thank Anders-Martin-Löf for many useful remarks and suggestions on how to improve an earlier version of the article. I am also grateful to an anonymous referee, pointing out some aspects of the practitioner's point of view.

References

[1] Bielecki T R and Rutkowski M 2002 *Credit Risk: Modeling, Valuation and Hedging* (Springer)

[2] Cifuentes A and O'Conner G 1996 The Binomial Exansion Method Applied to CBO/CLO Analysis, Special Report, Moody's Investment Services

[3] Davis M and Lo V 2001 Infectious Defaults, *Quantitative Finance,* **1**, No 4, 382-387.

[4] Davis M and Lo V 2001 Modelling default correlation in bond portfolios, in *Mastering Risk vol 2: Applications*, ed. Carol Alexander, (Financial Times Prentice Hall) 141-151

[5] Embrechts P, McNeil A and Straumann D 1998 Correlation and Dependency in risk Manegement: Properties and Pitfalls, *Working Paper*

[6] Gluck J and Remeza H 2000 Moody's Approach to Rating Multisector CDOs, Special Report, Moody's Investment Services

[7] Rolski T, *et al* 1999 Stochastic Processes for Insurance and Finance. (Wiley series in probability and statistics)