Mathematical Statistics
Stockholm University

Infectious Defaults on a Graph

Ola Hammarlid

Research Report 2004:9

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 **2004:9**,
http://www.math.su.se/matstat

Infectious Defaults on a Graph

Ola Hammarlid*

September 2004

Sammanfattning

In the Infectious Default Model every default has a chance of 'infecting' another company to default in the same sector. All parameters are assumed homogenous within a sector, which is too restrictive in practice. The remedy suggested is to relax the assumption of just one default per company. This makes possible a model where a default may only 'infect' another company if it belongs to a relation set of the defaulted. These connections between companies generate a graph, where hidden factors are easily added. Furthermore, every company has its own set of parameters. Two algorithms are derived that compute the distribution of the total credit loss. This relaxed model is shown to be a good approximation to the model were only one default per company is allowed, for a realistic portfolio.

KEY WORDS: Infectious default, dependent defaults, collateralized debt obligation, regulatory capital, credit loss, credit portfolio, graph

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

1 Introduction

To describe the aggregated credit loss of a portfolio realistically, dependence in defaults between the entities has to be considered. For example, to measure credit risk in a portfolio a quantile is usually used, Value at Risk, which depends heavily on the dependence. For contracts like Collateralized Debt Obligations (CDO), the pricing of the super senior tranche depends mostly on the dependence.

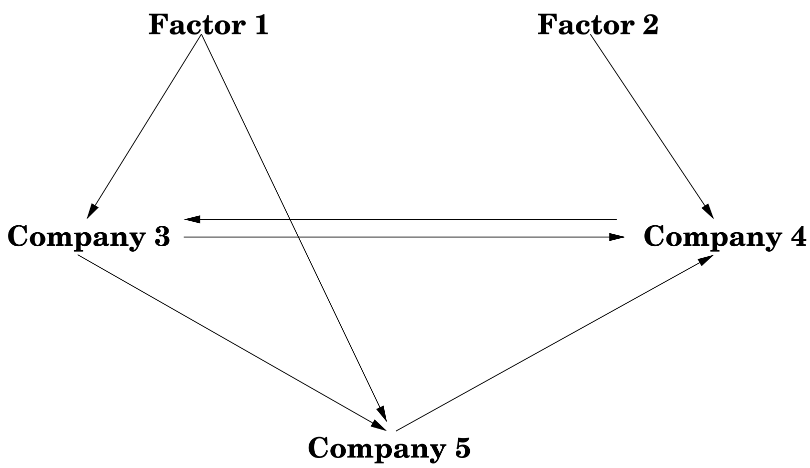
Moody's Binomial Expansion Technique (BET) models dependence by clustering defaults and raising the exposure, such that the expected value is fixed, Gluck and Remeza [6] and Cifuentes and O'Conner [2]. BET is frequently used in the financial market to describe the properties of a CDO. More sophisticated models involves for example copulas, 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]. In this Infectious Default Model (IDM) a defaulting company may 'infect' another company to default. The case of a company defaulting more than once were ruled out. IDM has several practical drawbacks due to the homogeneous assumptions of exposure, probability of default, probability of infection in each sector. On top of that, the question of how to define a sector, is a difficult problem to solve by itself.

We will overcome these drawbacks, by relaxing the assumption that default may only occur once for a company. This is not at all unrealistical. The definition of default contains a lot more than just bankruptcy. In the extreme, a bond issued by a company may be defaulting while another bond issued by the same company is not. However, the definition of default is not the purpose of this article. The difference of the distribution of the credit loss between a 'relaxed' model and a model were default can only take place once is small. Therefore, the method presented here could be used as an approximation, to the one-default restricted model.

A portfolio subject to credit risk contains n bonds, loans or other types of debt. There might be factors among these entities. A factor does not have any loss given default, but might infect a company to default, which results in a credit loss. Examples of factors are, a country, an industry sector or companies that is not in the portfolio. The companies and factors are numbered by $i = 1, \ldots, n$. We will from now on also call factors companies.

A company i might default by itself and then risk to infect (cause) another company j to default. It is only companies that belong to a certain set, A_i, that can be infected, that is, if $j \in A_i$. To get a clear picture of the idea see Figure 1.

The losses given spontaneous default are written L_i and losses given infected default are called K_{ij}. Both these types losses given default are assumed integer valued. Let X_i represent the number of spontaneous defaults of company



Figur 1: Each arrow describes the direction of how a default might infect. For example the default of Company 3 may cause company 4 and 5 to default. Therefore, the relation set of Company 3 is $A_3 = \{4, 5\}$. The factors are viewed as companies that do not have any loss given default, but may infect a company to default.

i and write Y_{ijl} for the number of infectious defaults of company j, caused by company i's default number l. Both X_i and Y_{ijl} will be modelled by Bernoulli or Poisson random variables. In total, combining all possible combinations of distribution assumptions, give us four different models.

A risk manager, as well as an investor in a CDO, are interested in the distribution of the total credit loss S. The main objective is to derive algorithms that compute the distribution of S. The algorithms derived are similar to the De Pril's algorithm and Panjer recursion, known in insurance mathematics, see for example Rolski *et al.* [8].

In Section 2 two algorithms are derived that in combination compute the distribution of the total credit loss, when both the number of spontaneous and infection defaults are Bernoulli distributed. Section 3 handles correspondingly the case of a Poisson number of spontaneous and infection defaults. The last section, Section 4, discusses issues of mixing the distribution assumptions. Furthermore, the relaxed model presented here is compared to the original assumption of just one default per company. The comparison is carried through by simulation.

2 Bernoulli Arrivals of Defaults

In this section we state and prove two algorithms that compute the distribution of the total credit loss S in the portfolio, when the spontaneous and the infectious defaults are Bernoulli random variables. The spontaneous defaults X_i and the infectious defaults Y_{ij} are independent Bernoulli random variables with probability p_i and q_{ij}. The third index l on Y_{ijl} is dropped in this section since X_i only can take the values 0 and 1.

First we study the loss inflicted by company i,

$$S_i = X_i \left(L_i + \sum_{j \in A_i} Y_{ij} K_{ij} \right) = X_i (L_i + Z_i), \tag{1}$$

where $Z_i = \sum_{j \in A_i} Y_{ij} K_{ij}$. Remember that if company i were a factor then $L_i = 0$.

Lemma 2.1 *The probability $P(Z_i = z)$ can be computed by*

$$P(Z_i = 0) = \prod_{j \in A_i} (1 - q_{ij}) \quad \text{and} \quad P(Z_i = z) = \frac{1}{z} \sum_{j \in A_i} v_{ij}(z),$$

where for $z \geq 1$,

$$v_{ij}(z) = \frac{q_{ij}}{1 - q_{ij}} \Big(K_{ij} P(Z_i = z - K_{ij}) - v_{ij}(z - K_{ij}) \Big)$$

and $v_{ij}(z) = 0$ otherwise.

The proof is skipped since it is almost identical to the proof of the following theorem 2.2.

In the following theorem we assume that the distribution of Z_i is already known by Lemma 2.1. In fact the distribution of Z_i does not have to be generated by a Bernoulli number of infections. In fact it could be any distribution of Z_i as long as it is integer valued and positive. For example, it could a Poisson number of infections, which is studied in Section 3.

Theorem 2.2 *The probability $P(S = s)$ can be computed by*

$$P(S = 0) = \prod_{i=1}^{n} (1 - p_i) \quad \text{and} \quad P(S = s) = \frac{1}{s} \sum_{i=1}^{n} v_i(s),$$

where for $s \geq 1$,

$$v_i(s) = \frac{p_i}{(1-p_i)} \sum_{z=0}^{s-L_i} P(Z_i = z)\Big((z+L_i)P(S = s-z-L_i) - v_i(s-z-L_i)\Big)$$

and $v_i(s) = 0$ otherwise. The probability $P(Z_i = z)$ is computed in Lemma 2.1.

Note that by the simple substitution of variables $L_i = 0$, $P(Z_i = K_{ij}) = 1$ and $p_i = q_{ij}$ in the theorem we have in fact proved Lemma 2.1.

To prove Lemma 2.1 we will use the same technique as Hammarlid [7], when aggregating sections in IDM.

Proof: The probability generating function of the loss inflicted by the spontaneous default of company i

$$g_i(t) = 1 - p_i + p_i \sum_{z=0}^{\infty} P(Z_i = z) t^{L_i + z} = 1 - p_i + p_i g_{Z_i + L_i}(t),$$

where $g_{Z_i+L_i}(t) = E\left[t^{L_i+Z_i}\right]$.

The independence between spontaneous defaults give the probability generating function of the total credit loss

$$g(t) = \prod_{i=1}^{n} g_i(t) = \prod_{i=1}^{n} \left(1 - p_i + p_i g_{Z_i+L_i}(t)\right).$$

The elementary property of the probability generating function gives us the probability of the total credit loss, that is

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

Therefore, we have $P(S = 0) = \prod_{i=1}^{n}(1 - p_i)$ and

$$g^{(1)}(t) = g(t)\frac{\mathrm{d}\log(g(t))}{\mathrm{d}t} = \sum_{i=1}^{n} \frac{p_i g^{(1)}_{Z_i+L_i}(t) g(t)}{1 - p_i + p_i g_{Z_i+L_i}(t)} = \sum_{i=1}^{n} V_i(t),$$

where $V_i(t) = p_i g^{(1)}_{Z_i+L_i}(t) g(t)(1 - p_i + p_i g_{Z_i+L_i}(t))^{-1}$. Put

$$v_i(s) = ((s-1)!)^{-1} V_i^{(s-1)}(0), \quad v_i(0) = 0 \tag{2}$$

then

$$P(S = s) = \frac{g^{(s)}(0)}{s!} = \frac{1}{s!}\sum_{i=1}^{n} V_i^{(s-1)}(0) = \frac{1}{s}\sum_{i=1}^{n} v_i(s).$$

It is now only left to determine $v_i(s)$. This is done recursively. First let us write,

$$(1-p_i)V_i(t) = p_i\left(g^{(1)}_{Z_i+L_i}(t)g(t) - g_{Z_i+L_i}(t)V_i(t)\right). \tag{3}$$

Then we expand $V_i(t)$ in terms of $v_i(s)$

$$V_i(t) = \sum_{s=0}^{\infty} t^{l-1}v_i(s), \tag{4}$$

since $v(s) = 0$ if $s \leq 0$. Let us recall that,

$$\begin{aligned}
g(t) &= \sum_{l=0}^{\infty} t^l P(S=l), \\
g_{Z_i+L_i}(t) &= \sum_{z=0}^{\infty} t^{L_i+z} P(Z_i = z), \\
g^{(1)}_{Z_i+L_i}(t) &= \sum_{z=0}^{\infty} (L_i+z)t^{L_i+z-1} P(Z_i = z).
\end{aligned}$$

We use these three expressions and equation (4) to rewrite equation (3),

$$\begin{aligned}
&(1-p_i)\sum_{s=0}^{\infty} t^{s-1}v_i(s) = \\
&= p_i \sum_{z=0}^{\infty}\sum_{l=0}^{\infty} t^{L_i+z-1+l} P(Z_i=z)((L_i+z)P(S=l) - v_i(l)) \\
&= \{\text{Change of variables, } s = l+z+L_i\} \\
&= p_i \sum_{s=0}^{\infty} t^{s-1}\Big(\sum_{z=0}^{s-L_i} P(Z_i=z)\left((L_i+z)P(S=s-z-L_i) - v_i(s-z-L_i)\right)\Big),
\end{aligned}$$

where we have used that $P(S=s) = 0$ when $s \leq 0$ and $P(Z_i = z) = 0$ for $z < 0$. By this we se that

$$v_i(s) = \frac{p_i}{1-p_i}\sum_{z=0}^{s-L_i} P(Z_i=z)\Big((L_i+z)P(S=s-z-L_i) - v_i(s-z-L_i)\Big).$$

The distribution of the number of defaults caused by a single company, N_i, can easily be computed by letting $K_{ij} = 1$ in Lemma 2.1. Then to compute the total number of defaults, N, let $L_i = 1$, except for factors, in Theorem 2.2 and exchange $P(Z_i = z)$ by the distribution of $P(N_i = m)$.

3 Poisson default arrivals

In this section we make a change of assumption. Let X_i be independent Poisson random variable with intensity λ_i, representing the number of spontaneous defaults of company i. Write Y_{ijl} for the infection on company j to default, caused by company i's default number l. The infected default are independent Poisson random variables with intensities λ_{ij}.

If we let $\lambda_i = p_i$ and $\lambda_{ij} = q_{ij}$ then the expected number of the total credit loss remain unchanged compared to the Bernoulli case. An alternative choice of intensity could be $\lambda_i = -\log(1-p_i)$ and $\lambda_{ij} = -\log(1-q_{ij})$. Then the Poisson distribution would be an upper limit for the distribution of the Bernoulli model. A mathematical argument for approximating the Bernoulli model is that $g(t) = \prod_{i=1}^{n}(1 - p_i + p_i g_{Z+L_i}(t)) \approx \exp\left(\sum_{i=1}^{n} p_i(g_{Z+L_i}(t) - 1)\right)$, which is the probability generating function of a compound Poisson random variable.

The distribution of the total loss inflicted by company i,

$$S_i = X_i L_i + \sum_{l=0}^{X_i} \sum_{j \in A_i} Y_{ijl} K_{ij} = \sum_{l=0}^{X_i} \left(L_i + \sum_{j \in A_i} Y_{ijl} K_{ij} \right) = \sum_{l=0}^{X_i} L_i + Z_l(i), \quad (5)$$

where $Z_l(i) = \sum_{j \in A_i} Y_{ijl} K_{ij}$. The equivalent of this equation, in the Bernoulli case is Equation (1).

Lemma 3.1 *Assume the number of infections Y_{ijl} are Poisson distributed with intensity λ_{ij}. Then we have $P(Z_l(i) = 0) = \exp\left(-\sum_{j \in A_i} \lambda_{ij}\right)$ and*

$$P(Z_l(i) = z) = \frac{1}{z} \sum_{j \in A_i} \lambda_{ij} \sum_{v=1}^{z} v P(V_i = v) P(Z_l(i) = z - v).$$

The proof is skipped since the proof can be based on the following theorem 3.2 by some simple substitutions of variables.

Define the random variable V_i and its moment generating function by

$$P(V = K_{ij}) = \frac{\lambda_{ij}}{\sum_{j \in A_j} \lambda_{ij}} \quad \text{and} \quad M_V(\theta) = E\left[e^{\theta V_i}\right]. \quad (6)$$

The loss of Z_l is a compound Poisson random variable with moment generating function

$$M_Z(\theta) = \exp\left(\sum_{j \in A_j} \lambda_{ij} \left(e^{\theta K_{ij}} - 1\right) \right) = \exp\left(\sum_{j \in A_j} \lambda_{ij} \left(M_V(\theta) - 1\right) \right). \quad (7)$$

The total credit loss in the portfolio is $S = \sum_{i=1}^{n} S_i$ where S_i are defined in (5). The spontaneous defaults are independent and distributed, thus

$$M_{S_i}(\theta) = E\left[\exp\left(\sum_{l=0}^{X_i} \theta(L_i + Z_l(i))\right)\right] = \exp\left(\lambda_i\left(e^{\theta L_i} M_Z(\theta) - 1\right)\right). \quad (8)$$

In the following theorem note that the distribution of Z_i is arbitrary. Therefore it is possible to exchange the Poisson assumption of Y_{ijl} and let it be a Bernoulli random variable instead.

Theorem 3.2 *Assume the number of spontaneous defaults is Poisson distributed with intensity* $\sum_{i=1}^{n} \lambda_i$*. Then we have* $P(S_i = 0) = \exp\left(-\sum_{i=1}^{n} \lambda_i\right)$ *and*

$$P(S = s) = \frac{1}{s} \sum_{i=1}^{n} \lambda_i \sum_{z_i=0}^{s} (L_i + z_i) P(Z_i = z_i) P(S = s - z_i - L_i), \quad s > 1.$$

Proof: The probability no default is equal to the probability of no spontaneous defaults, $P(S = 0) = \exp\left(-\sum_{i=1}^{n} \lambda_i\right)$. The first order derivative

$$M_S'(\theta) = \sum_{s=0}^{\infty} e^{\theta s} s P(S = s). \quad (9)$$

But according to Equation (8) we have that

$$\begin{aligned} M_S'(\theta) &= M_S(\theta) \sum_{i=1}^{n} \lambda_i e^{\theta L_i} (L_i M_Z(\theta) + M_Z'(\theta)) \\ &= \sum_{h=0}^{\infty} \sum_{i=1}^{n} \lambda_i \sum_{z_i=0}^{\infty} e^{\theta(h+z_i+L_i)} (L_i + z_i) P(Z(i) = z_i) P(S = h) \\ &= \{\text{Change variables, } s = L_i + z_i + h\} \\ &= \sum_{s=0}^{\infty} e^{\theta s} \sum_{i=1}^{n} \lambda_i \sum_{z_i=0}^{s} (L_i + z_i) P(Z(i) = z_i) P(S = s - z_i - L_i). \end{aligned}$$

The last equality is because when $s - z_i - L_i < 0$ then $P(S = s - z_i - L_i) = 0$. If we compare this last expression with (9) we see that

$$sP(S = s) = \sum_{i=1}^{n} \lambda_i \sum_{z_i=0}^{s} (L_i + z_i) P(Z(i) = z_i) P(S = s - z_i - L_i), \quad s > 1.$$

As before, the distribution of the number of defaults caused by a single company's default, N_i, can easily be computed by letting $K_{ij} = 1$ in Lemma 3.1. Then to compute the total number of defaults, N, let $L_i = 1$, except for factors, in Theorem 3.2 and exchanging $P(Z_i = z)$ by the distribution of $P(N_i = m)$.

4 Comparing the different models

The work in this article relies on the relaxation of the assumption of at most one-default per company. What can we say about the change of distribution when we make this relaxation? 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

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

However, to compare the difference between the models define the norm

$$\| \cdot \| = \sup_A |P(X \in A) - P(Y \in A)|,$$

where X and Y are any random variables. In the case of the relaxed model compared to the one-default per company model, we have that the contributing mass to the difference is when, infection of a default takes place but the infected company already is in default. The probability of such an event,

$$P(\text{No double default for } i \text{ and } j) = p_i p_j \left(q_{ij} \mathbf{1}_{\{j \in A_i\}} + q_{ji} \mathbf{1}_{\{i \in A_j\}} \right), i \neq j,$$

where $\mathbf{1}_{\{.\}}$ denotes an indicator function. Let $\tilde{S}$ be the total credit loss in the model with no more than one default per company. By the above and $P(A \cup B) \leq P(A) + P(B)$ we have that,

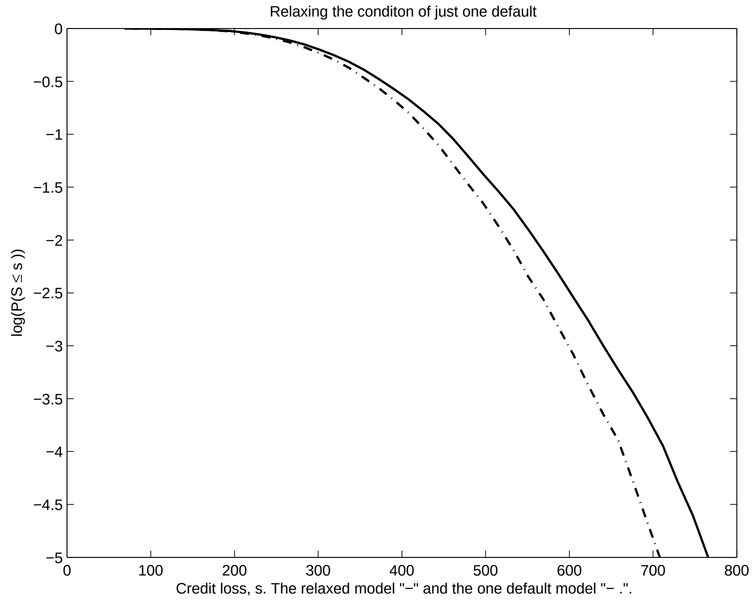
$$\sup_A |P(S \in A) - P(\tilde{S} \in A)| \leq \sum_{i=1}^{n} \sum_{j \neq i} p_i p_j \left(q_{ij} \mathbf{1}_{\{j \in A_i\}} + q_{ji} \mathbf{1}_{\{i \in A_j\}} \right).$$

Assume that the largest relation set is of size n_A and that n_A does not grow with the size of the portfolio. Furthermore, write $\rho = \max_{i \leq n} p_i$ and $\psi = \max_{i,j} q_{ij}$. Then

$$\sup_A |P(S \in A) - P(\tilde{S} \in A)| \leq 2n \cdot n_A \rho^2 \psi.$$

For example in a loan portfolio of investment graded loans the probability of default over one year is probably less that $p = 1\%$. The probability of infection is assumed to be less than $q = 5\%$ and the relation set less than 10, then $\sup_A |P(S \in A) - P(\tilde{S} \in A)| \leq 10^{-4} n$.

We will illustrate, by simulation, that the difference is small even in a more extreme case. Let the reference portfolio contain 150 loans and one factor. For each company i the probability p_i of default is uniformly drawn from the interval $[0, 0.2]$, and q_{ij} is uniformly drawn from the interval $[0, 0.4]$. The losses given default L_i and K_{ij} are uniformly drawn on the sets $1, \ldots, 30$



Figur 2: The dotted line is the credit loss with no more than one default per company and the other line is when more than one default is allowed. The comparison is done by simulation.

and $1, \ldots 20$. The size of the relation set is expected to be 15 and the nodes are uniformly picked over the portfolio. Therefore the approximation of the distance is of no use. By simulation, (10^5 trials) the practical difference is displayed in Figure 2

The difference between the models is of less importance since the estimation uncertainty of the probabilities of default and the probabilities of infection are probably of greater magnitude, at least for investment graded debt. From a regulatory perspective the relaxed model is an upper bound and could therefore be used, even if the one default model is assumed.

Which distribution assumption on the number of defaults should be made? The algorithm of the Bernoulli number of defaults needs to compute and store $n+1$ 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. The Poisson model is in most cases a good approximation to the Bernoulli model, when ever the default probabilities are small.

Another 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.

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.

5 Acknowledgement

I would like to thank Anders Martin-Löf for some useful remarks.

Referenser

[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] Hammarlid O. (2003) Aggregating sectors in the infectious defaults model. *Quantitative Finance,* Vol. **4**, issue 1, pages 64 –69. (http://stacks.iop.org/1469-7688/4/64)

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