Modelling Default Correlation in Bond Portfolios

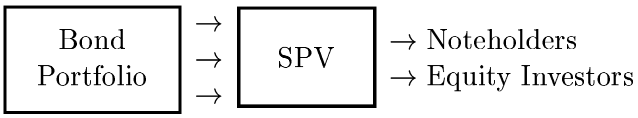
Mark Davis[*]
Visiting Professor, Financial and Actuarial Mathematics Group
Vienna University of Technology
Violet Lo[†]
Assistant Manager, Tokyo-Mitsubishi International plc, London

Abstract

The performance of a CBO structure is critically dependent on the correlation of defaults in a medium-sized portfolio of bonds. We introduce two classes of purely probabilistic models to handle these interaction effects. In the first, we suppose that default of an issuer in a particular industry sector may trigger off defaults of other issuers in the same sector by an 'infection' mechanism. As infection increases we find that the default distribution has increased variance and heavier tails, quantifying the concentration risk. The second model is a continuous-time stochastic process called an enhanced-risk model, that exhibits very similar features. In this case the incremental default probability is increased for all other issuers for a certain random period following occurrence of a default. The paper concludes with some comments on parameter estimation and applications of these models in CBO analysis.

Figure 1 shows the structure of a typical CBO (Collateralized Bond Obligation) transaction. The central box labelled SPV denotes a 'special purpose vehicle', i.e. a company which is set up for this one transaction and dissolved when the transaction terminates. On the closing date of the transaction the SPV accepts principal payments from investors and purchases a portfolio of perhaps 60 high yield bonds. All subsequent payments made to investors are derived from income received from the bond portfolio.

Figure 1: CBO Structure



The investors are in two categories: Noteholders (contributing perhaps 85% of the total principal) and Equity Investors. The former receive a specified coupon (=

---

[*] `m.h.a.davis@btinternet.com`

[†] `violet.lo@t-mi.com`

interest rate) on their investment and are repaid their principal at maturity; they are first in priority of payment. Their coupon is considerably less than the average coupon of the bond portfolio. The Equity Investors are paid residual receipts, with no guaranteed coupon or guarantee of principal repayment. If the SPV actually receives all coupon and principal payments due from the bond portfolio, then the Noteholders will receive everything due to them and the Equity Investors will receive a very high return — perhaps in excess of 20%. If any of the collateral bonds defaults then the Equity Investors will suffer a loss of return, but the Noteholders are protected — up to a point — by the priority rule. (This is what justifies paying a lower coupon to them.) The Equity Investors are in effect making a leveraged investment in the high-yield portfolio.

Clearly, the performance of a CBO depends entirely on the default performance of the underlying high yield portfolio. The rating agencies (S&P, Moody's *et al.*) give credit ratings to companies or individual debt issues, and publish comprehensive statistics of default performance in the various rating categories. Reassuringly, it is indeed true that the historically experienced default rates are a monotone function of the rating category. The yield on lower-rated bonds is higher, to compensate investors for increased default risk. The idea of a CBO is to concentrate this risk, shifting the equity investors to the high-risk end of the risk/return curve. At the same time, the overall risk is reduced by diversification, in that the SPV invests in a medium-size portfolio of bonds rather than just one high-yield issue.

The most difficult question to answer in analysing CBOs is: how effective is this diversification? Obviously, this depends on 'correlation'. If the 60 bonds default independently, then it is an easy matter to predict the default performance of the portfolio from the rating agencies' default statistics. However, there is clearly the potential for 'concentration risk': default of a whole subset of the bond portfolio might be triggered by some common event which affects all the issuers in that subset. The rating agencies deal with this question in various ways. We will describe one of them, Moody's Binomial Expansion Technique, below. This provides in many cases an effective, or at least systematic, solution to the problem. However it does not model the actual default events in the portfolio, since a smaller, 'equivalent' portfolio is considered. It is also a static model in that it only concerns the amount of default in a fixed time interval, with no consideration of the times at which defaults take place. To overcome these drawbacks we introduce two classes of probabilistic models. The first is a static model involving 'default infection' - default of one bond may trigger off default of other related bonds. The other model takes a somewhat similar approach, but in a continuous time stochastic process context. Here, as soon as one bond defaults we enter a 'high risk' period, lasting a random length of time, in which default intensities for the other bonds are increased. The tendency therefore is for occasional periods of high credit spreads and 'bursts' of default - a much more realistic model than one assuming constant credit spreads.

As will be seen, the common feature of these models is heavy-tailed default distributions. As correlation (or 'infection') increases, more weight is pushed out with the tails of the distribution, increasing the risk to the supposedly secure noteholders. The main purpose of the analysis is to quantify this extra risk.

The most obvious feature of this whole area is lack of data. Default events are infrequent and data collected over many years can hardly be supposed to be a sample from a stationary process. Thus reliable statistical estimates of distributional parameters are practically impossible to obtain, particularly the key correlation estimates. In these circumstances there is no point at all in introducing complicated models with lots of additional parameters. Our philosophy is to introduce the simplest models with the smallest number of extra parameters (never more than one or two) that will capture 'concentration risk' in a credible way. Even if statistical estimates are hard to come by, plausible ranges of the parameters can be arrived at on intuitive grounds. Analysis using plausible values gives bounds on asset values and is certainly better than ignoring concentration risk altogether.

1 Moody's Diversity Scores and BET

Consider a collateral portfolio consisting of, say, $n = 60$ bonds. Assume for simplicity that the bond issuers all have the same credit rating and that the notional amounts of the bonds are equal. Moody's and other rating agencies have assembled statistics of defaults in the various rating classes over the last 20 years from which we can infer a distribution function for the default time. We can thus estimate the default probability p for any of the bonds in the portfolio over a given period, say 10 years. If we assume that all the bonds are independent then the distribution function for the number K defaulting is the binominal distribution

$$P[K = k] = C_k^n p^k (1-p)^{n-k}, \tag{1.1}$$

where C_k^n is the binomial coefficient, $C_k^n = n!/(k!(n-k)!)$. The expected number of defaults and the variance are np, $np(1-p)$ respectively.

Moody's Binomial Expansion Technique (BET) – see (Moody's Investment Services, 1997) – is based on the idea that issuers in the same industry sector are related, while issuers in different industry sectors can be treated as independent [1]. Moody's define 32 industry sectors and the following Diversity Score table

Table 1: Moody's Diversity Scores

Number of Firms in the Same Industry	Diversity Score
1	1.0
2	1.5
3	2.0
4	2.3
5	2.6
6	3.0
7	3.2
8	3.5
9	3.7
10	4.0
10 or more	Evaluated case by case

---

[1] This ignores, of course, the possibility of 'general economic factors' affecting all industries.

For example, suppose our portfolio of 60 bonds in distributed as follows

Table 2: Portfolio distribution

No of issuers in sector	1	2	3	4	5
No of incidences	2	7	6	4	2
Diversity	2	10.5	18	9.2	5.2

Thus there are two cases in which issuers are the only representatives of their industry sector, seven cases in which pairs of issuers are in the same sector, etc. The diversities, taken from Table 1 are given in the third row; the total diversity score is 45.

In Moody's Binominal Expansion Technique (BET) we consider the original portfolio of 60 bonds to be equivalent to a portfolio of 45 *independent* bonds with the same default probability but with notional value 60/45 times the original notional. Thus the expected loss, expressed in dollar terms, is the same ($= 60p$). The distribution of the number defaulting in the equivalent portfolio is

$$P[k \text{ defaults}] = C_k^d p^k (1-p)^{d-k}$$

where d is the diversity score. The main purpose of the BET is to evaluate the losses suffered by the Noteholders. Because of the payment priority rule, the Noteholders experience no losses at all up to some level of default, and progressively increasing losses after that. To give a simple evaluation, we calculate the expected loss for an option-like loss function $L(x) = 100 \max(x-K, 0)$ as shown in fig 1[2], where x denotes the fraction of the portfolio notional that defaults.

For the original portfolio with diversity score 60 and individual default probability 0.1, the expected number of defaults is $m = 6$ and the standard deviation is $r = 2.32$. To close approximation therefore, $m + 3\sigma = 13$ and we take the threshold level K as $13/60 = 0.217$. We now plot, in Figure 3 the expected loss and the loss probability[3] as a function of diversity score.

As can be seen, both expected loss and loss probability increase as the diversity is reduced, even though the *expected* proportion of default is constant. For our specimen portfolio with diversity 45, the loss probability is 1.2%, more than double the loss probability for diversity 60 (0.57%).

2 Infectious Defaults

This idea was introduced in our earlier paper (Davis and Lo, 1999). Rather than replacing the original portfolio by a smaller 'equivalent' portfolio we wanted to give a probabilistic model for interaction effects in the original portfolio. To describe this, let n be the portfolio size and Z_i be random variables such that $Z_i = 1$ if bond i defaults and $Z_i = 0$ otherwise. The number defaulting is thus $N = Z_1 + Z_2 + \ldots + Z_n$. The value of Z_i is determined as follows. For $i = 1, \ldots, n$ and $j = 1, \ldots, n$ with

---

[2] The scaling factor 100 is for graphical convenience. The units are arbitrary.

[3] i.e. the probability that losses exceed 0.217

Figure 2: Loss Function

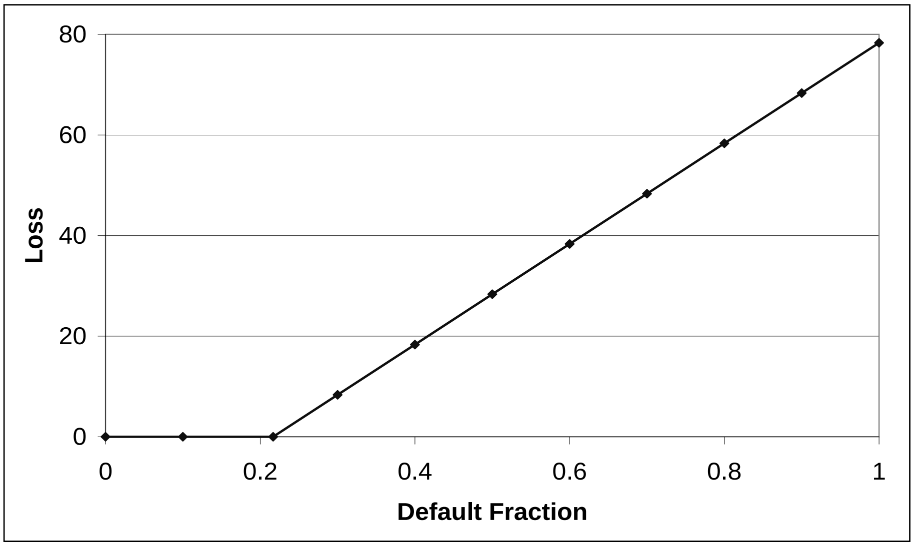
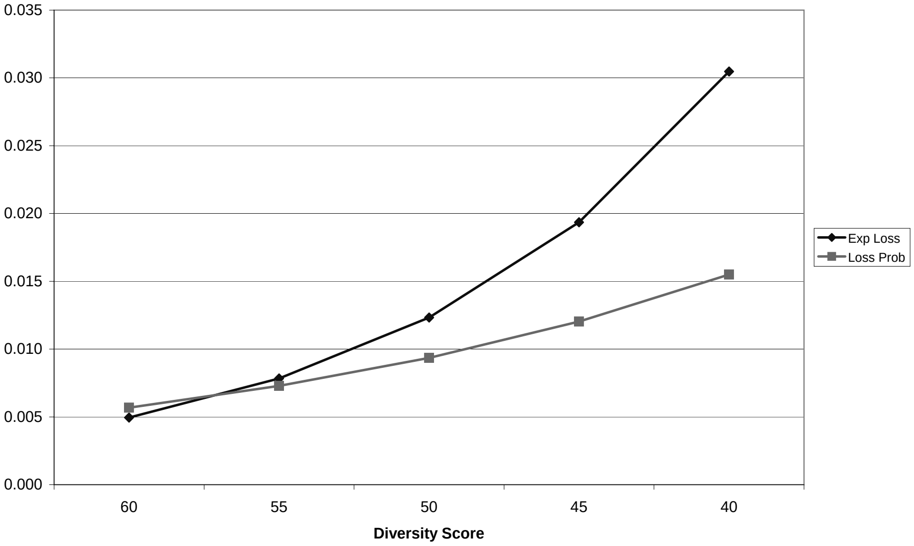


Figure 3: Losses as function of diversity score



$j \neq i$ let X_i, Y_{ij} be independent Bernoulli random variables with $P[X_i = 1] = p$, $P[Y_{ij} = 1] = q$. The constant q is the infection parameter. Then

$$Z_i = X_i + (1 - X_i)\left(1 - \prod_{j \neq i}(1 - X_j Y_{ji})\right). \tag{2.1}$$

The idea here is that bond i may default 'directly' ($X_i = 1$) or may be 'infected' by default of bond j. Indeed, the second term in (2.1) is equal to 1 when $X_i = 0$ and for some j both X_j and Y_{ji} are equal to 1. It is shown in (Davis and Lo, 1999) that the distribution of N is

$$P[N = k] = c_k^n \alpha_{nk}^{pq} \tag{2.2}$$

where

$$\begin{aligned} \alpha_{nk}^{pq} &= p^k (1-p)^{n-k} (1-q)^{k(n-k)} \\ &+ \sum_{i=1}^{k-1} c_i^k p^i (1-p)^{n-i} (1-(1-q)^i)^{k-1} (1-q)^{i(n-k)}. \end{aligned}$$

The expected value is

$$EN = n\left(1 - (1-p)(1-pq)^{n-1}\right). \tag{2.3}$$

When $q = 0$, (2.2) and (2.3) reduce to the Binomial distribution with parameter p. When $q > 0$, EN is increased because we have provided more ways in which each bond can default. However, we want to consider, as in the Diversity Score analysis, a situation in which the probability of default for each individual bond is fixed at p. To achieve this, we need to *rescale* p, i.e. find the number $\hat{p}(q)$ such that $(1 - \hat{p}(q))\,(1 - \hat{p}(q)q)^{n-1} = 1 - p$. We have $\hat{p}(0) = p$ and $\hat{p}(q) < p$ for $q > 0$. Then $EN = np$ when the infection model parameters are $\hat{p}(q)$, q, but the *variance* is increased, as we shall see below.

The next step is to bring in the industry sector distribution. As in the diversity score analysis, we assume that bonds in m different industry sectors default independently. Within each industry sector i we assume that an infection model holds with parameters $(p_i = \hat{p}(q_i), q_i)$. The parameters required are p, the marginal default probability, and q_i, the infection parameter for sector i. Generally it is reasonable to take $q_i = q$ for some fixed q, reducing the model parameters to two (p, q).

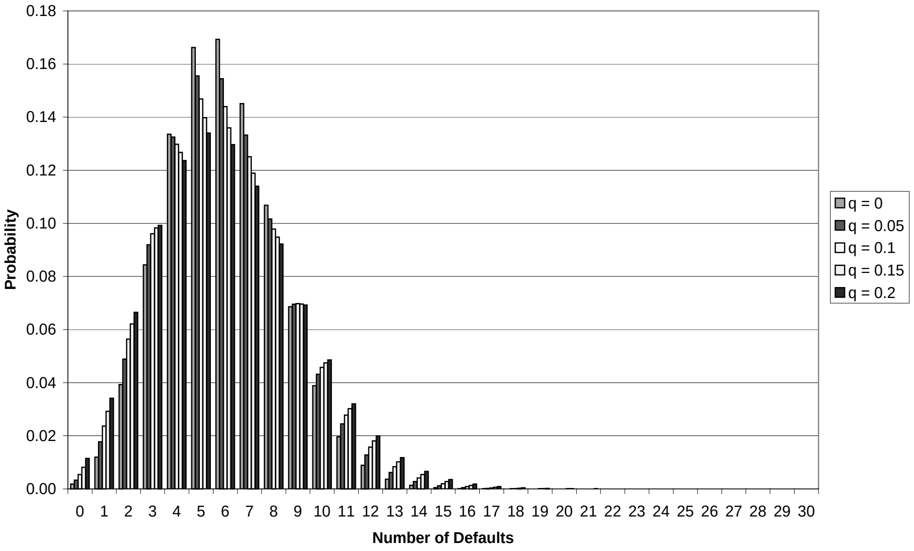
The probability of k_i defaults in sector i, $i = 1, \ldots, m$ is

$$\prod_{i=1}^{m} F(k_i, n_i, p_i, q_i)$$

where n_i is the number of bonds in sector i and $F(k, n, p, q)$ is the distribution function (2.2). The probability that the total number of defaults is k is therefore

$$\sum_{a \in \mathcal{A}_k^m} \prod_{i=1}^{m} F(k_i, n_i, p_i, q_i) \tag{2.4}$$

Figure 4: Infectious Default Distribution



where m is the number of industry sectors and $\mathcal{A}_k^m$ is the set of arrangements $a = \{k_1, \ldots, k_m\}$ of k defaults in the m industry sectors.

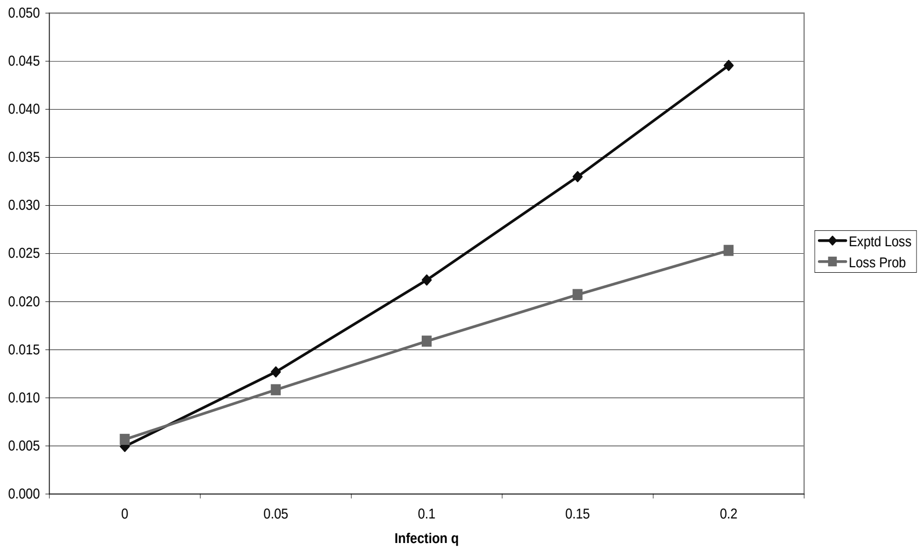
Calculating (2.4) is not completely straightforward, even for moderate n, since the number of elements in $\mathcal{A}_k^m$ can be extremely large. Effective computational methods based on an idea we call 'superboxes' have been developed and are described in (Davis and Lo, 1999). Figure 4 shows the default distributions for values of the infection parameter q ranging from 0 to 0.4, for the 60-bond portfolio with industry distribution as in Table 2. The individual default probability is $p = 0.1$, so all distributions have mean 6. The case $q = 0$ is the Binomial $(60, p)$ distribution. As can be seen, the infection process increases the variance, pushing more weight out into the tails of the distribution. Figure 5 shows the expected loss probability using the same parameters as in section 2 (i.e. the threshold K is $13/60$ corresponding to default of 13 bonds). Comparing Figure 3 and 5 we see that infection has a very similar qualitative effect to reduction in diversity. In terms of expected loss, a diversity score of 45 corresponds to an infection parameter $q = 0.08$ for this portfolio.

3 A Dynamic Model: Enhanced Risk

The infectious defaults model gives a convenient family of distributions for analysing defaults in a given time interval. However this is not the whole story. If we wish to simulate the performance of a CBO structure, for example, then the question of *when* defaults occur is as important as *how many* occur. To model this we need a stochastic process model that produces sequences of random default times according to some prescribed mechanism.

Let us start with the simplest case, in which there are n independent bonds each having a default time that is exponentially distributed with parameter λ. The

Figure 5: Infectious Default Loss



probability of default for any one bond in a time interval T is then $p = 1 - e^{-\lambda T}$ and the distribution of the number defaulting is Binomial (n, p). This is the case analysed in the previous sections with diversity equal n or infection parameter $q = 0$. Let N_t denote the number of defaults in the interval $[0, t]$. It is easy to see that the process

$$M_t = N_t - \int_0^t \lambda(n - N_s) ds \tag{3.1}$$

is a martingale: we have

$$P\left[\text{default in } [t, t+dt] | N_t\right] = E[dN_t | N_t] = \lambda(n - N_t) dt \tag{3.2}$$

i.e. the default or 'hazard' rate is proportional to the number of bonds still 'alive'. Denoting $m(t) = EN_t$ we see by taking expectations in (3.1) that

$$m(t) = n\lambda t - \int_0^t \lambda m(s) ds$$

which is easily solved to give

$$m(t) = n\left(1 - e^{-\lambda t}\right),$$

in agreement with the binomial distribution.

The martingale method gives us a convenient simulation methodology: rather than simulating default times for each of the n bonds we can just generate successive jump times of N_t. Denoting these $T_1, T_2, \ldots$ we know from (3.1) or (3.2) that $T_{i+1} - T_i$ is exponentially distributed with parameter $\lambda(n - i)$. Usually there will only be a few T_i's in the interval $[0,T]$ of interest. This kind of approach has been emphasized by (Duffie and Singleton, 1998).

To introduce some interaction effects into this model we would like to capture the idea that incidence of a default puts everybody else 'at risk', at least for a while. This accords with evidence (from, say, the Russian default of 1998) that when a default takes place, credit spreads for other issuers are increased, settling back to 'normal' levels after some time. In mathematical terms the assumptions are as follows. Initially each bond has hazard rate λ, so the total hazard rate is $n\lambda$. When a default happens, the hazard rate for *all remaining bonds* is increased by a factor $a > 1$, so the total hazard rate is $a(n-1)\lambda$. The multiplying factor continues to be a for an exponentially distributed time with parameter μ, after which it reverts to 1 until the next default happens. Thus 'normal' periods alternate with 'enhanced risk' periods triggered off by an actual default.

The process just described in a piecewise-deterministic Markov process (PDP) in the terminology of (Davis, 1993). The state space is the $2(n+1)$-point set $E = \{(i,j) : i = 0,1; j = 0,1,\dots,n\}$ and x_t is the value of the process at time t. Index $i = 0,1$ corresponds to normal and enhanced risk respectively, while j is the number of bonds still alive. Thus the starting point is $(0,n)$. The generator $\mathcal{A}$ for this process, acting on a test function $f : E \to R$ is

$$\begin{aligned}
\mathcal{A}f(0,j) &= j\lambda\left(f(1,j-1) - f(0,j)\right), \quad j > 0 \\
\mathcal{A}f(0,0) &= 0 \\
\mathcal{A}f(1,j) &= aj\lambda f(1,j-1) + \mu f(0,j) - (aj\lambda + \mu)f(1,j), \quad j > 0 \\
\mathcal{A}f(1,0) &= \mu\left(f(0,0) - f(1,0)\right).
\end{aligned}$$

In state $(0,j)$ the hazard rate is $j\lambda$ and the process moves to $(1,j-1)$ (one extra default, enhanced hazard). In state $(1,j)$ the hazard rate is $h = aj\lambda + \mu$ and the process moves either to $(1,j-1)$, with probability $aj\lambda/h$, corresponding to another default or to $(0,j)$, with probability μ/h, corresponding to the end of an enhanced risk period.

We can compute the distribution of the number of defaults in a fixed time $[0,T]$ by solving the backward equation

$$\begin{aligned}
\frac{\partial}{\partial t}v(t,x) + \mathcal{A}v(t,x) &= 0, \quad (t,x) \in [0,T] \times E && (3.3) \\
v(T,x) &= l(x), \quad x \in E. && (3.4)
\end{aligned}$$

As shown in (Davis, 1993), the solution $v(t,x)$ is in probabilistic terms

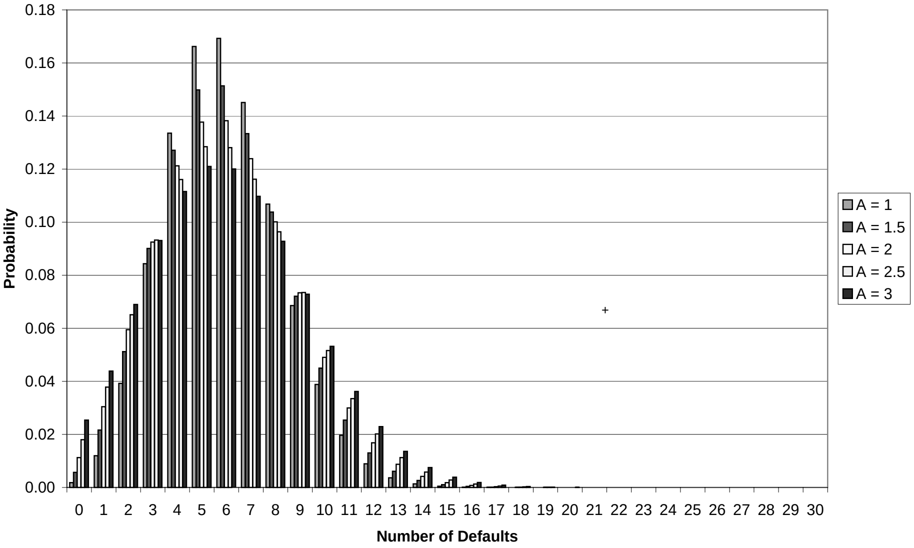
$$v(t,x) = E_{t,x}l(x_T).$$

If we take $l = l_k$ where

$$\begin{aligned}
l_k(0,n-k) &= l_k(1,n-k) = 1 \\
l_k(i,j) &= 0, \quad j \neq n-k
\end{aligned}$$

and denote by v_k the corresponding solution of (3.3), (3.4) then clearly $v_k(0,(0,n))$ is the probability of just k defaults in $[0,T]$ starting with a portfolio of size n.

Figure 6: Distribution for Enhanced Risk Model



Since E is a finite set, (3.3), (3.4) is just a linear ordinary differential equation of dimension $2(n+1)$. It can be solved by computing the matrix exponential or by direct application of Runga-Kutta integration.

As in the infection model, the expected number of defaults increases as we increase the 'enhancement factor' a. To get the correct comparison we must rescale λ so that the individual default probabilities are equal to some prescribed value. Here we take, as before, a portfolio of 60 bonds and a $T = 10$-year time horizon. If $a = 1$ the individual default probability is $p = 1 - e^{-\lambda T}$. For comparison with previous results we take $p = 0.1$ which implies $\lambda = 0.0105$. The distribution is then Binomial (n, p) as noted earlier. For $a > 1$ we choose $\lambda = \lambda_a$ such that the mean is equal to np and calculate the distribution as described above. As regards μ, we somewhat arbitrarily take $\mu = 0.5$ corresponding to an average 2-year period before 'normal' conditions resume. Figure 6 shows the distribution functions for a ranging from 1 to 3. As can be seen, the results are qualitatively very similar to the infection model, with weight being pushed out into the tail as a increases. The expected losses and loss probabilities, analogous to those computed earlier, are shown in Figure 7.

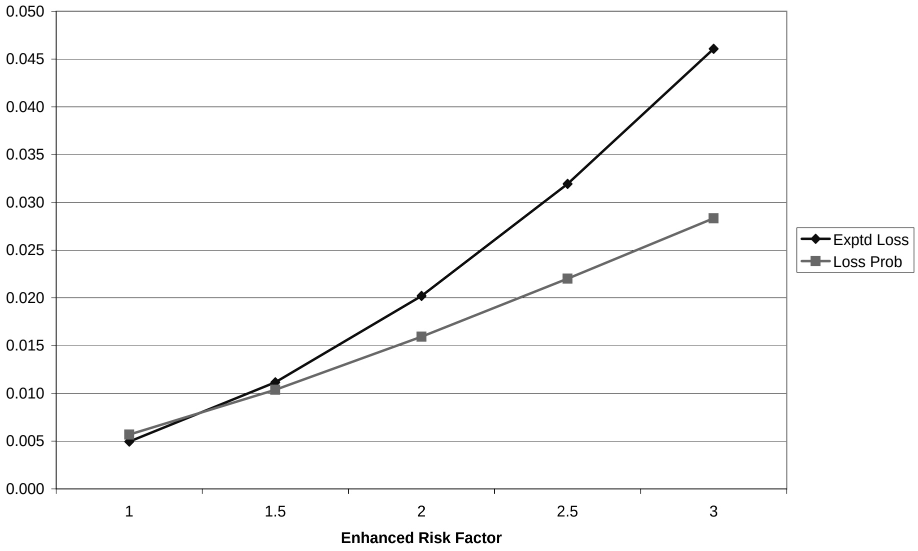
In terms of expected loss, $a = 2$ is roughly equivalent to $q = 0.1$ in the infection model.

Of course, other similar models are easily envisaged. For example, we could apply the enhanced risk model within industry sectors, assuming that the industry sectors are independent. It depends on the application; the model we have given concentrates on the systematic risk affecting all issuers.

4 Parameter Estimation and Applications

These topics will be treated in detail in a later paper.

Figure 7: Enhanced Risk Loss



Estimation is plagued by lack of data. However, both the infection and enhanced risk models (leaving aside μ, the value of which does not seem critical) have only two parameters and hence are fully determined once two statistics of the distribution - for example, the mean and variance - are given. Our suggestion is to use Moody's default statistics to estimate the variance of the number of defaults over the desired time horizon. This information together with Moody's own empirical distributions then fixes the model parameters.

For a CBO structure, there are two questions of interest that can be addressed using our models: the credit rating of the Notes (or senior tranches for a multi-tranche CBO) and the expected returns and return volatility of the equity tranche. For the former, infectious defaults provides an alternative to the Diversity Score analysis that is in some ways attractive. For the latter, a dynamic model is essential, and the enhanced risk model combines realistic patterns of default with ease of simulation.

References

Davis, M. and Lo, V., *Infectious Defaults*, submitted to Credit Metrics Monitor, Risk Metrics Group, 1999.

Moody's Investment Services, *The Binomial Expansion Technique*, 1997.

Davis, M.H.A. *Markov Models and Optimization*, Chapman and Hall, London 1993.

Duffie, D. and Singleton, K. *Simulating Correlated defaults*, working paper, Graduate School of Business, Stanford University 1998.