Derivative Securities, Fall 2010
Mathematics in Finance Program
Courant Institute of Mathematical Sciences, NYU
Jonathan Goodman
`http://www.math.nyu.edu/faculty/goodman/teaching/DerivSec10/resources.html`

Week 9

1 Bond rating model

We gave a simple model of credit risk in which there was a single parameter, λ, the default intensity. But real bonds rarely just default. The companies that assign credit ratings to bonds, the *ratings agencies*, will change their rating of a bond if they feel that the probability of default has changed. It is natural to give a mathematical model of this as a finite state space *Markov chain*.

One of the effects of the more complex Markov chain model is that the probability of default by time t is a more complicated function of t than a simple exponential. This is one way to explain the clear fact that credit spreads are not constant in time. On the contrary, the credit spread tends to increase with time. One could model this by saying that the default intensity is an increasing function of time, and then try to find the term structure of default intensity. But another way is to use the Markov chain model with fixed parameters.

Each rating agency has a small list of bond ratings it uses. For example, Standard and Poors ratings go from AAA, the highest, AA to A and eventually to D, for default. An abstract Markov chain has at any time a state $X(t)$. In a finite state space chain the possible states are $1, 2, \ldots, n$, so $X(t) \in \{1, 2, \ldots, n\}$ for any t. At certain random times, the chain jumps from one state to another. The simple exponential model discussed last week has $n = 2$ states, $X(t) = 1$ meaning the bond is not in default at time t, and $X(t) = 2$ meaning that it is in default. In general, there would be as many states as there are credit ratings.

A Markov chain is characterized by its matrix of *transition rates*. For each pair of states, there is a rate, λ_{ij} of having a transition from i to j. This means that, for $i \neq j$,

$$\lambda_{ij}\, dt \;=\; \Pr\left(X(t+dt) = j \mid X(t) = i\right) \; . \tag{1}$$

What we called λ in our simple jump to default model now is λ_{12}. This means that

$$\lambda\, dt \;=\; \Pr\left(\text{default at } t + dt \mid \text{not default at } t\right) \; .$$

That was the definition we gave last week. The Markov chain is *stationary* if the transition rates do not depend on t.

Markov chains are characterized by the *Markov property*. We have already used this property repeatedly without stating it explicitly. It is the property that the present is all the information about the past that is relevant for characterizing the future. Technically, the Markov property is the fact that the probability distribution of the path $X([t, T])$ conditioned on knowing $X(t)$ is the same as the distribution of the path $X([t,T])$ conditioned on knowing $X([0,t])$. In terms of bonds, this would say, for example, that an AA bond recently downgraded from AAA is no more or less risky than another AA bond that has been rated AA for a long time. All the information about the credit-worthiness of the bond is reflected in its current rating. The ratings agency has taken into account its rating history as much as possible.

The transition rates can be assembled into an $n \times n$ matrix, L, which is the transition rate matrix. This L also is called the *generator* of the Markov chain. It plays the same role, hence the same letter, as the generator of a diffusion process in the backward equation for the diffusion process. The formula (1) does not define the diagonal entries of L. Those are defined by the requirement that the *row sums* of L are equal to zero. This is

$$\sum_{j=1}^{n} \lambda_{ij} \; = \; 0 \; , \tag{2}$$

for all i. This forces the diagonal rates to be

$$\lambda_{ii} \; = \; -\sum_{j \neq i} \lambda_{ij} \; . \tag{3}$$

The formula (1) implies that $\lambda_{ij} \geq 0$ if $j \neq i$, and (3) then implies that $\lambda_{ii} \leq 0$. The formula (3) has a simple probabilistic interpretation, which is that $-\lambda_{ii}$ is the total rate to jump away from state i if you are in state i. Indeed, summing (1) over $j \neq i$ gives that

$$\sum_{j \neq i} \lambda_{ij} dt \; = \; -\lambda_{ii} dt \; = \; \Pr\left(X(t+dt) \neq i \mid X(t) = i \right) \; .$$

In principle it is not necessary to define the λ_{ii}, but using (3) and the matrix L makes it easy to formulate the dynamical equations related to Markov chains.

The first dynamical equation is the *forward equation.* This equation is satisfied by the *occupation probabilities*

$$u_i(t) \; = \; \Pr\left(X(t) = i \right) \; .$$

The derivative $\dot{u}_i = \frac{d}{dt} u_i$ is determined by the rates of transitions into and out of state i. In a small time interval dt, define the in transition probability as du_{i+} and the out probability as du_{i-}. To observe a $j \to i$ transition, X must be in state j at time t (i.e. $X(t) = j$) and make the transition. The probability of this is

$$\Pr(X(t) = j) \cdot \Pr(j \to i \text{ in time } (t, t+dt)) \; = \; u_j(t) \cdot \lambda_{ji} \, dt \; .$$

Summing over these gives

$$du_{i+} \; = \; \sum_{j \neq i} u_j(t) \lambda_{ji} \, dt \; .$$

The probability of observing an $i \to j$ transition, by the same reasoning, is

$$\Pr(X(t) = i) \cdot \Pr(i \to j \text{ in time } (t, t + dt)) \;=\; u_i(t) \cdot \lambda_{ij}\, dt \;.$$

When we sum this over $j \neq i$, the $u_i(t)$ is a common factor on the right (recall (3)):

$$du_{i-} \;=\; u_i(t) \sum_{j \neq i} \lambda_{ij}\, dt \;=\; -u_i(t)\lambda_{ii}\, dt \;.$$

Altogether, we have (recall that leaving state i makes u_i smaller)

$$du_i \;=\; du_{i+} - du_{i-} \;=\; \sum_{j=1}^{n} u_j(t)\lambda_{ji}\, dt \;. \tag{4}$$

These equations can be put in matrix vector form as follows. Let $u(t)$ be the n component row vector $u = (u_1, \ldots, u_n)$. The equations (4) are equivalent to the matrix equation

$$\dot{u} \;=\; uL \;. \tag{5}$$

On the right side is a $1 \times n$ matrix, the row vector u, multiplying the $n \times n$ matrix L, producing the $1 \times n$ vector $\dot{u}$. You will quickly get used to putting the vector on the "wrong" side of the matrix, when it is a row vector.

The differential equation system (5) comes with *initial conditions*, which are the values $u_i(0)$ assembled into the initial vector $u(0) = (u_1(0), \ldots, u_n(0))$. If the initial probabilities are known, then (5) determines the probabilities $u(t)$ for all $t > 0$. For instance, you probably know the bond rating at the present time, so $u_i(0) = 1$ for that rating and $u_i(0) = 0$ for all other ratings. The differential equations (5) form a dynamical system that tells you how the occupation probabilities change over time.

Let us put the simple exponential model in the general framework. The rate for $1 \to 2$ (i.e. not default to default) transitions is λ. The rate for $2 \to 1$ transitions is zero. This makes $\lambda_{11} = -1$ and $\lambda_{21} = \lambda_{22} = 0$. Altogether,

$$L \;=\; \begin{pmatrix} -\lambda & \lambda \\ 0 & 0 \end{pmatrix} \;.$$

The differential equations(5) are

$$\dot{u}_1 \;=\; -\lambda u_1 \;,$$

and

$$\dot{u}_2 \;=\; \lambda u_1 \;.$$

The first has solution $u_1(t) = e^{-\lambda t} u_1(0)$. If we start in the non-default state (nobody buys the Brooklyn bridge), then $u_1(0) = 1$ and $u_2(0) = 0$. Therefore $u_1(t) = e^{-\lambda t}$ and $u_2(t) = 1 - e^{-\lambda t}$. These are the answers we got last week.

In general, it is hard to find explicit solutions to the equations (5). But there are many easy and reliable ways to get the answer numerically. Matlab, for example, is good at computing the exponential of a matrix. $u(t) = u(0)e^{tL}$ is the solution to (5). One also can find the solution by solving the ordinary differential equations numerically, say using Runge Kutta or a linear multistep method. Some of these are discussed in the Numerical Methods class. It probably is not a good idea to solve (5) using eigenvalues and eigenvectors. The matrix L is not symmetric. The eigenvalues do not have to be real, and often will not be. The eigenvectors do not have to be orthogonal and may be a seriously bad basis.

A state, k, is called an *absorbing state* if there are no transitions out of state k, which is to say that $\lambda_{kj} = 0$ for all $j \neq k$. It is common in the general theory of Markov chains to assume there is no absorbing state. But most bond rating systems have an absorbing state, default. Once a bond has defaulted, there is no zombie afterlife. It stays defaulted for ever.

For bond ratings, the company Risk Metrics (and other companies) sells a transition matrix, L, that is estimated from historical data. It would be hard to estimate a risk neutral L from bond price (credit spread) data because there are so many entries of L to estimate. It would make sense to assume that $\lambda_{ij} = 0$ if $|i - j| > 1$, which means that upgrades or downgrades go only one rating at a time. In principle it always should be like this, but there are some downgrades that go several ratings at once. Ignoring these rare events might be worth it, if the result were a useful model. A matrix with $\lambda_{ij} = 0$ if $|i - j| > 1$ is called *tridiagonal*.

In addition to the forward equation (5), there is a backward equation associated to the Markov chain. Imagine a bond[1] "option" with payout $V(i)$ if $X(T) = i$. Define the value function for such a payout by

$$f_i(t) \;=\; E\left[V(X(T)) \mid X(t) = i\right] \; . \tag{6}$$

This also may be expressed as

$$f_{X(t)} \;=\; E\left[V(X(T)) \mid \mathcal{F}_t\right] \; .$$

Of course, the information available at time t includes the state $X(t)$. The value function clearly satisfies final conditions $f_i(T) = V(i)$.

As with the occupation probabilities, it is convenient to organize the values of the value function into a vector, f. The difference is that f is a column vector with $f^t(t) = (f_1(t), \ldots, f_n(t))$. The backward equation is

$$\dot{f} \;=\; Lf \; . \tag{7}$$

There is a derivation of this equation similar to the derivation of (5), but instead I give a derivation that is based on the *duality* relation between the backward and forward equations. This derivation starts from the formula for the unconditional expected payout

$$E[V(X(T))] \;=\; \sum_{i=1}^{n} u_i(t) f_i(t) \;=\; u(t)f(t) \; . \tag{8}$$

[1]This might better be called a bond *contract* because the actions at time T are determined by the contract. It is common to call final time payouts other than vanilla puts and calls contracts rather than options.

Before deriving this, note that the expression on the right is the product of the $1 \times n$ matrix (i.e. row vector) u with the $n \times 1$ matrix (i.e. column vector) f, with the result being the 1×1 matrix, a number. To derive (8), just note that X must be in some state at time t. The "law of total probabilty" (a cousin of the "tower property") states that

$$E[V(X(T))] = \sum_i \Pr(X(t) = i) \cdot E\left[V(X(T)) \mid X(t) = i\right] .$$

The right side of this is exactly the sum in the middle of (8).

Now, the probabilities $u_i(t)$ and the value function values $f_i(t)$ are differentiable functions of t. We know this for u because we wrote the differential equation (5). This is despite the fact that $X(t)$ has jumps. Therefore, the ordinary rules of calculus (i.e. no Ito term) apply. Also, note that the left side of (8) does not depend on t. This means that the right side is constant as a function of t and it's derivative with respect to t is zero. We may differentiate (8) with respect to t using the product rule

$$0 = \frac{d}{dt}\Big(u(t)f(t)\Big) = \dot{u}(t)f(t) + u(t)\dot{f}(t) .$$

From the u dynamics (5) we get (it is important here that matrix multiplication, including multiplication with row and column vectors, is associative)

$$0 = uLf + u\dot{f} .$$

Therefore

$$u\,(Lf - f) = 0 ,$$

for all probability vectors u. It is easy to see (think about this for a minute) that the vector space spanned by all probability vectors is the vector $\mathbb{R}^n$ vector space of column vectors. Therefore the vector $Lf - \dot{f}$ must be zero. That is the derivation of (7).

There are options, contracts more precisely, on bond defaults. The simplest version, which does not seem to exist in the marketplace, would be an option that pays at time T if $X(T)$ is in default. That would correspond to taking $V(n) = 1$ and $V(j) = 0$ for $j \neq n$, assuming n is the default state.

A more realistic credit dependent option is the *credit default swap*, or *CDS*, on a coupon paying bond. Here is an idealized continuous time model of this. An investor buys a bond that pays a continuous and constant coupon of Cdt in each time interval dt until the bond matures at time T_m. At time T_m the company pays the investor the *notional* (or *principal*, or *notional principal*),[2] L. The credit default swap is an agreement with a third party to insure the bond payments. The third party agrees to pay L to the investor upon a *credit event*, which we take just to be a default on the coupon or principal payment. In return, the investor pays the third party protection provider a fixed payment,

[2]There is more on bonds later. For now, recall that if the price of the bond at time $t = 0$ is less than L, it is a *discount bond*. If the price is equal to L, it is sold at *par*.

called the *CDS spread* of sdt in each time interval dt. We formulate this as follows. The investor pays sdt and receives Cdt in time interval dt at each t with $t < T_m$ and $t < T$, where T is the default time. If $T > T_m$, nothing happens. IF $T < T_m$, then at time T, the protection provider pays the investor L. After that, all payments stop.

A real CSD is like this, except that payments are discrete, quarterly for example. Also, there may be a more complicated definition of credit event. For example, a company may have many bond obligations, most of which have nothing to do with the investor who holds the CDS. A default on any one of them may be considered a credit event that triggers the CDS payment.

1.1 Discrete observations

Suppose you fix a time t and ask for the *transition probabilities*

$$P_{ik}(t) \;=\; \Pr(X(t_0+t)=k \mid X(t_0)=i)\;.$$

These do not depend on the starting time t_0, only on the elapsed time t. It is convenient to take $t_0 = 0$. You can find the transition probabilities by solving the forward equation (5) with initial data $u_i(0) = 1$, $u_j(0) = 0$ for $j \neq i$. Then $u_k(t) = P_{ik}(t) = \Pr(X(t) = k)$. In other words, the row vector u is row i of the matrix P. For that reason, the $P_{ij}(t)$ satisfy the forward equation

$$\dot{P}_{ik} \;=\; \sum_j P_{ij}\lambda_{jk}\;.$$

In matrix form, this is

$$\dot{P} \;=\; PL\;. \tag{9}$$

The initial conditions we chose for u imply that the initial condition for P is $P(0) = I$. These properties make $P(t)$ the *fundamental solution* of the ODE (ordinary differential equation) system (5).

It is common to write $P(t) = e^{tL}$ because $P(t)$ is given by the Taylor series for the exponential $e^x = \sum_{n=0}^{\infty}\frac{1}{n!}x^n$. Indeed, you can check that the formula

$$P(t) \;=\; \sum_{n=0}^{\infty}\frac{1}{n!}t^nL^n \tag{10}$$

satisfies (9) just by differentiating the right side with respect to t

$$\dot{P} \;=\; \sum_{n=1}^{\infty}\frac{1}{n!}nt^{n-1}L^n \;=\; \left(\sum_{n=1}^{\infty}\frac{1}{(n-1)!}nt^{n-1}L^{n-1}\right)L \;=\; P(t)L\;.$$

This fundamental solution also may be called the *transition matrix*. We can use it to describe what happens when the continuous time Markov chain is observed at discrete times $t_k = k\delta t$. For example, we may observe a continuous time bond process once each quarter when payments are due. The transition matrix for a particular δt is the transition matrix for a discrete time Markov chain. It has the property, for example, that $P(2\delta t) = P(\delta t)^2$. You can see this either using probabilistic reasoning or by manipulating the formula (10).

1.2 Diffusion processes

The generator L of a Markov chain plays a role similar to that played by the generator of a diffusion process, which we called L also. The backward equation for a diffusion process takes the form

$$\partial_t f \;+\; Lf \;=\; 0\;,$$

where

$$Lf(x,t) \;=\; a(x)\partial_x f \;+\; \tfrac{1}{2}b^2(x)\partial_x^2 f\;.$$

The diffusion is $dX = a(X)dt + b(X)dW$.

The diffusion process also has a forward equation, which is derived from the backward equation much as we derived (7) from (5). The forward equation quantity is $u(x,t)$, which is the probability density of X_t. The dynamics of u and f are related by the same kind of duality we saw above in the discrete case. We have the formula

$$E[V(X_T)] \;=\; \int u(x,t)f(x,t)\,dx$$

just as before, and for the same reason. Since the left side does not depend on t, we get zero if we differentiate the right side

$$0 \;=\; \int \left(\partial_t u(x,t)\right) f(x,t)\,dx \;+\; \int u(x,t)\left(\partial_t f(x,t)\right)\,dx\;.$$

Now substitute the backward equation in the last term on the right:

$$0 \;=\; \int \left(\partial_t u(x,t)\right) f(x,t)\,dx - \int u(x,t)\left(a(x)\partial_x f(x,t)\right)\,dx - \int u(x,t)\left(\tfrac{1}{2}b^2(x)\partial_x^2 f(x,t)\right)\,dx\;.$$

You can integrate by parts, assuming that $u(x,t) \to 0$ so fast as $x \to \pm\infty$ that the boundary terms are zero. The result is

$$0 \;=\; \int \left\{\partial_t u(x,t) \;+\; \partial_x\left(a(x)u((x,t)\right) \;-\; \tfrac{1}{2}\partial_x^2\left(b^2(x)u((x,t)\right)\right\} f(x,t)\,dx\;.$$

We reason from this as before. The only way the integral can vanish for any function $f(x)$ is for the integrand to vanish everywhere. That means that

$$\partial_t u(x,t) \;=\; -\,\partial_x\left(a(x)u((x,t)\right) \;+\; \tfrac{1}{2}\partial_x^2\left(b^2(x)u((x,t)\right)\;. \qquad (11)$$

This is the forward equation for the diffusion process.

The operator on the right side of (11) is the *adjoint* of the backward equation operator, L. It is written

$$L^* u(x) \;=\; -\,\partial_x\left(a(x)u((x)\right) \;+\; \tfrac{1}{2}\partial_x^2\left(b^2(x)u((x)\right)\;.$$

This is the adjoint of L in the sense that $\langle u, Lf\rangle = \langle L^* u, f\rangle$, where $\langle u, f\rangle = \int u(x)f(x)dx$ is the L^2 inner product[3]. The forward equation then takes the symbolic form $\partial_t u = L^* u$.

[3] These comments are for those who have studied this kind of thing. If you have not, just make sure you understand the relation between (11) and the backward equation.

The discrete state space Markov chain forward and backward equations have the same relationship. This is clearer if we write the forward equation for the column vector u^t:

$$\dot{u}^t = L^t u^t .$$

In other words, the matrix that enters into the forward equation is the transpose of the one in the backward equation. “Adjoint” is the term for transpose used in fancier situations.

The forward equation (11) is suitable for propagating the probability density u forward in time. If we are given $u(x,0)$, we can calculate $u(x,t)$ for $t > 0$. Any probability density can be propagated forward in this way. On the other hand, it almost always impossible to propagate probability densities backwards. If someone gives a probability density density $u(x,t)$ and asks what density of X_0, $u(x,0)$ gives rise to it, the answer almost always is that there is no such density. Mathematically, we say that the forward equation is *well posed* for going forward in time but *ill posed* for going backward in time. The backward equation reverses this, it goes backward happily but refuses to go forward.

2 Correlation and Gaussian copulas

Securitization and bundling are important developments in the finance industry, for better or worse. Securitization combines a large number of similar assets into a big pool so that pieces of the pool can be sold as shares to individual investors or institutions. The most important area of securitization is bundles of home mortgages called *collateralized mortgage obligations*, or *CMOs*. Securitization creates a large number of essentially identical shares of large pools of mortgages. This allows investors to buy and sell mortgage assets without studying the details of individual mortgages.[4] Creating a liquid market for mortgage assets made these assets more attractive to investors and therefore more valuable. Making mortgages more valuable to investors makes them less expensive for borrowers.

Diversification is another benefit of securitization. An investor can buy a small fraction of a large number of mortgages rather than buying all of a small number. In this way, the investor will be subject to the average default rate rather than the individual defaults of a few specific people. Of course, the amount of diversification depends on the correlation between defaults. If defaults are independent, the central limit theorem says that the standard deviation of a pool of N mortgages is smaller than the standard deviation of any one of them by a factor $1/\sqrt{N}$ (fine print: after normalizing by the size of the asset, assuming all assets are identical). On the other hand, consider a one company town. If the company closes, a large number of mortgages will default.

Uncertainties in diversification, and default correlation, are particularly important for securities called *tranches* of CMOs. Tranche simply means “slice”

[4]Imaging the investor trying to buy one mortgage having to interview several homeowners to decide which is more creditworthy. Who would have the time to do that?

in French, as in a slice of bread. The tranches of a CMO give some investors higher priority than others in receiving their shares of the mortgage payments. A typical tranche structure might be a *senior tranche* consisting of the first 70% of the payments, a *mezzanine* tranche consisting of the next 20% and an *equity* tranche consisting of the rest. If 60% of the payments are received, the senior tranche holders will receive all the payments and the others will get nothing. If 80% come in, the senior tranche holders will get their 70% and the remaining 10% will go to holders of the mezzanine tranche. If 95% come in, then the senior and mezzanine tranche holders will be paid in full and the equity tranche holders will get paid half (5% out of 10%).

Each tranche is a *fixed income security*, which means that it is a security whose payout structure is defined by a contract. Any fixed income security has its credit spread, which is the difference between the return on the security and the return on a comparable[5] risk free security. Ratings agencies assign credit ratings to tranches of CMOs as they do for other fixed income securities. One of the causes of the financial meltdown was the fact that ratings agencies vastly underestimated the correlation between defaults. A significant number of AAA rated tranches of subprime mortgage CMOs were not fully funded, which put serious pressures on the financial institutions holding them. You are not supposed to have to hedge AAA income streams.

Securitization and tranching have been applied to corporate bonds, leading to *collateralized bond oblications*, or *CBOs*. A general term is *debt* obligations, leading to *CDOs*. At the fit of pre-meltdown enthusiasm, people even created a *CDO squared*, which was a tranched asset based on a bundle of CDOs.

The CDO tranche structures amplify the importance of correlation. Correlation can make the difference between $80\% \pm 2\%$ and $80\% \pm 10\%$. In the former case, a middle aged person could buy into the mezzanine tranche above confident that it will be worth something. In the latter case, the investor has a serious possibility of receiving nothing. In the former case, the senior tranche is unlikely to be touched, while this is a real possibility in the latter.

It is not clear how to model correlation mathematically. The general correlation problem can be stated as follows. Suppose random variables X_i have the same probability density function $f(x)$. You want to construct a joint probability density $g(x_1, \cdots, x_n)$ so that the marginals are all equal to $f(x)$. Saying that the marginal of X_j is $f(x)$ is (the integration leaves out dx_j)

$$f(x) \;=\; \int \cdots \int g(x_1, \ldots, x_n)\, dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_n \; .$$

Of course, one could take $g(x_1, \ldots, x_n) = \prod f(x_i)$, but that just makes the X_i independent. It is not clear how to make the X_i not independent in general. Try to do it in the case of exponential random variables $f(x) = e^{-x}$, say.

[5] How we make the comparison depends on the contractual structure of the fixed income security – when and for how long and what size payments will be. The *duration* of such a security is a weighted average of the payment dates. We will discuss duration when in the coming section on interest rate modeling.

One kind of random variable we do know how to correlate is the normal. Suppose random variables X_i are jointly normal with covariance matrix Σ. Then the variance of X_i is $\Sigma_{ii} = \sigma_i^2$. The correlation between X_i and X_j is

$$\rho_{ij} \;=\; \frac{\text{cov}(X_i, X_j)}{\sqrt{\text{var}(X_i)\text{var}(X_j)}} \;=\; \frac{\Sigma_{ij}}{\sigma_i \sigma_j} \; .$$

Of course, $\rho_{ii} = 1$ by definition. The ρ_{ij} may be arranged into a correlation matrix, ρ. We will see next week that any ρ that is positive definite in fact is the correlation matrix of jointly normal random variables. That means that even after you have specified the distributions of the individual X_i as normal, you can correlate them in any way allowed by linear algebra.

The copula method uses this to created correlated random variables from other distributions. I discuss the case of $\{0, 1\}$ random variables that represent bond defaults. Next week I will discuss more general cases, such as correlated default times. The method described here is a simple special case of the general method we will discuss next week.

Our goal is to create a joint distribution of random variables $X_1, \ldots, X_n$ so that each X_i is equal either to zero (not default) or 1 (default). Suppose that $p = \Pr(X_i = 1)$ is the same for each i. Let $Z_0, \ldots, Z_n$ be independent standard normals. Z_0 will represent a *market risk* factor, while the other Z_i are *idiosyncratic* risk factors that effect X_i only. The strength of the market risk factor is a. Define

$$Y_i \;=\; aZ_0 + \sqrt{1 - a^2} Z_i \; . \tag{12}$$

This formula is chosen so that $\text{var}(Y_i) = 1$ for all i and $\rho(Y_i, Y_j) = E[Y_i Y_j] = a^2$ for all $i \neq j$. Choose a cutoff y_c so that $\Pr(Y_i < y_c) = p$. This is the same as saying $N(y_c) = p$. Set $X_i = 1$ if $Y_i < c$ and $X_i = 0$ otherwise. The defaults are all the ones with $Y_i < c$. If $\rho = a^2$ is close to 1, then the formula (12) makes the Y_i approximately equal to each other, because $\sqrt{1 - a^2}$ is close to zero. This makes the X_i highly correlated.