



UNIVERSITY OF BERGAMO

DEPARTMENT OF MANAGEMENT, ECONOMICS
AND QUANTITATIVE METHODS

Working paper n. 1/2016

ricerca_meq@unibg.it

Series Quantitative Methods

Portfolio loss modeling: an
infectious framework

Gianluca Farina, Rosella Giacometti



Portfolio loss modeling: an infectious framework

Gianluca Farina and Rosella Giacometti
University of Bergamo, Via dei Caniana 2, Bergamo, Italy

December 1, 2015

Abstract

This paper introduces a new model that takes inspiration from the approach of Davis and Lo (2001) but differentiates from it by relaxing the homogeneous assumption and by introducing few restrictions on the shape of the infection mechanism that allow for a good level of tractability. In particular, the contagion mechanism proposed is the result of two independent components: an infection attempt generated by defaulting firms and a failed defense from healthy ones. Within this framework it is possible to prove several theoretical results regarding marginal default probabilities and joint default/survival events. Finally, we provide an efficient recursive algorithm for the portfolio loss distribution similar, in spirit, to the one commonly used for CID (conditionally independent) models.

1 Introduction

One of the main problems in credit modeling is default clustering: it has been observed, especially during recessions, that defaults are not uniformly spaced in time but rather tend to concentrate over small periods of time. Even the most used approaches to introduce dependence among defaults (via exposure to common factors, by correlating the default intensity processes and by direct application of copula methods) struggle to replicate this observed pattern. One valid alternative is to add asymmetric dependency structure to the modeling framework; these effects, usually referred to as contagion or infection, not only have the ability to increase the probability of observing extreme losses in the portfolio but can also account for default clustering¹.

From a pure technical point of view, adding these type of effects introduces a looping mechanism that makes calibration more problematic: the probability of default of each name can impact and is impacted by the probability of default of the others. Several attempts have been proposed in order to resolve the looping issue when adding such effects.

Jarrow and Yu [2001] suggest to separate the firms into two groups: primary names (i.e. belonging to the first group) can only default idiosyncratically while names in the second group can also default because of infection starting from the first set of names. Their work generalizes reduced-form models by making default intensities depend on counterpart default. The primary/secondary separation has been applied by other authors too, for example Rösch and Winterfeldt [2008]: they start from a one-factor model where the number of defaults of primary names can affect the default probabilities of secondary names².

Another paper starting from classical factor models but adding contagion mechanism is the one by Neu and Kühn [2004]. Firms can have either mutually supportive or competitive relationships between them; a default will hence decrease the default probability of competitors but increase the same quantity for firms that had positive inflows from the name in distress (for example suppliers/clients).

Egloff et al. [2007] instead add micro-structural dependencies via a directed weighted graph and show how even well diversified portfolios carry significant credit risk when such inter-dependencies are accounted for. The open problem in their approach (as well as in other network based models) is the calibration of the weights that form the neural network.

Yu [2007] works with an intensity based model where the default intensities are driven also by past defaults history, in addition to exogenous factors. A special case of this model is the copula approach presented by Schönbucher and Schubert [2001].

Two other valuable readings are the works by Giesecke and Weber [2004] and by

¹See the work from Azizpour et al. [2008] for evidence about default clustering and the role of contagion.

²Numerical results are shown for homogeneous portfolios only, although we suspect that a mixture of numerical integration and convolution techniques (for example, fast Fourier transform) could be effective.

Frey and Backhaus [2010]; the first adds contagion processes to more standard common factor approaches while the second applies a Markov-chain model with default contagion to the problem of dynamically hedge CDO products.

Our work belongs to a branch of the literature started by Davis and Lo [2001]. In their model, built using independent Bernoulli variables, firms can default in two ways, either idiosyncratically or via infection from other defaulting names. Unfortunately, in their most generic (and therefore elegant) specification of the model, one has to rely on Montecarlo techniques in order to obtain the portfolio loss distribution. This is not very efficient given that the variable space required is of quadratic order with respect to the size of the portfolio considered (for a portfolio of size n , one has n^2 binary variables to deal with). The solution proposed by the authors leads to closed form results for the loss distribution at the cost of quite strict assumptions³.

Other authors have used Davis and Lo's original approach as starting point of more sophisticated models. Sakata et al. [2007] extended Davis and Lo's original model by assuming that idiosyncratic defaults might in fact be avoided with the help of non-defaulted names. In their model there is hence not only a (negative) contagion that causes more defaults but also a positive one, called recovery spillage, that prevents entities from defaulting. The main results they obtain are similar to the ones presented in Davis and Lo [2001] in terms of both complexity and assumptions required (homogeneous portfolio).

Another very interesting work is the one by Cousin et al. [2013] that extends the original approach in many ways. First, the authors use a multiple time step model where defaults happening at the previous time interval can still cause contagion later on. In addition, they consider the case where more than one infection is needed to cause a default by contagion and, from a theoretical point of view, they relax several of the original paper assumptions. Unfortunately, the results for the most generic specification of their model involve quite heavy calculations that are computationally very demanding. In fact, the numerical applications shown are based on assumptions that are in line with Davis and Lo [2001] with respect to complexity and flexibility⁴.

We can notice how from one side we have very generic extensions of the model that are cumbersome to use and from the other side extremely fast versions that can only deal with unrealistic instances of the model due to the severe restrictions imposed on the starting assumptions. We suspect that neither one of the two cases is actually usable in real life and that this has been the main reason this type of model has not been applied to a wider set of problems. Ideally, one would like to keep the contagion mechanism from the generic form of the model but benefit from the speed of the restricted versions: the model we are presenting in this work is a compromise that moves toward that ideal solution. From the computational point of view, portfolio losses can be calculated via

³They require that the portfolio considered is completely homogeneous with respect to the probability of idiosyncratic default, the infection rates and even the losses given default (LGD).

⁴In particular, the authors require that the Bernoulli variables used are exchangeable.

a recursive algorithm with reasonable costs. From the modeling side, we can specify different default probabilities as well as different infection rates. We also added an immunization mechanism that healthy names can use to protect themselves against infections. The only concession we had to make with respect to the original model is the following: instead of having several different contagion shocks when a name defaults (one for each name in the portfolio), we now model the event that the defaulting name might infect the entire system. We had to exclude the possibility that one defaulting firm infects some firms but not others in favor of the more draconian scenario in which it either tries to infect all names or none. If from one side this difference makes our model **not** an extension of Davis and Lo's one, from the other it is crucial in obtaining efficient loss distribution algorithms.

The rest of the paper is structured as follows: section 2 introduces our new model and provides explanations and examples regarding the parameters choice. Section 3 presents some theoretical results regarding marginal and joint default distribution. Section 4, instead, describes the algorithm for the portfolio loss distribution (including some hints on implementation strategy). Section 5 concludes while proofs of few theoretical results are provided in the appendix.

2 The new model

Our model follows closely in the steps of Davis and Lo's one (although it is not a direct extension of it). In their paper, given a portfolio of n entities, the authors modeled the default probability of entity i via the binary variable Z_i that can either take value 0 (indicating survival) or 1 (default) and that is constructed according to the following equation

$$Z_i = X_i + (1 - X_i) \cdot \left[1 - \prod_{i \neq j} (1 - X_j \cdot Y_{i,j}) \right] \quad (1)$$

The main assumption behind last equation is the existence of n variables $X_i, i = 1, \dots, n$ and $n \cdot (n - 1)$ variables $Y_{i,j}, i, j = 1, \dots, n, j \neq i$. Variables X are responsible for idiosyncratic default while $Y_{i,j}$ governs the possibility that name i is infected by name j , given j 's idiosyncratic default. The additional assumption taken is that the variables X_i and $Y_{i,j}$ are i.i.d. according to a Bernoulli distribution.

Dependency among the variables Z_i is hence introduced via the infection mechanism triggered by the Y s. This very elegant and extremely flexible formulation has only one main drawback in our opinion: there are no closed-form formulas (or even semi-analytical techniques) that can be applied to determine the total portfolio loss distribution. The reason for this is the asymmetry embedded in the $Y_{i,j}$ variables that allow for name j to selectively "infect" some names but not others.

In order to solve this problem, the model we propose drops the above highly

asymmetric framework in favor of a less flexible one where each name, upon idiosyncratic default, can either infect no other name, or spread an infection attempt to the entire system. The economic interpretation of the infection channel is that the idiosyncratic default of the firm is read as a shock from the market capable of triggering more defaults. In addition, the other names can resist the infection attempt via an immunization mechanism that protects them against every possible infection. From one side we have simplified the original model by dropping the "selective" infection mechanism but, on the other side, we have enriched it by allowing names to develop immunization⁵. The following equations show our new approach:

$$Z_i = X_i + [1 - X_i] \cdot [1 - U_i] \cdot \left\{ 1 - \prod_{i \neq j} [1 - X_j \cdot V_j] \right\} \quad (2)$$

$$P[X_i = 1] = p_i, \quad P[U_i = 1] = u_i, \quad P[V_i = 1] = v_i \quad (3)$$

where we have postulated the existence of $3N$ mutually independent Bernoulli variables $X_1, \dots, X_n, V_1, \dots, V_n, U_1, \dots, U_n$. As in the Davis and Lo paper, we also have the two mechanisms for default seen earlier, i.e. idiosyncratic defaults and contagion. In particular name i infects name j only if two independent conditions are satisfied:

1. **Infection attempt from i :** name i defaults idiosyncratically and attempt to spread the infection to **all** other names. This component is driven by the V variables.
2. **Failed defense from j :** name j fails to defend itself from **every** possible infection. This component is driven by the U variables.

Note the independence assumption among the building blocks of (2) that serves two purposes: firstly, it will be crucial when we will prove theoretical results in later sections as it allows to split joint distribution in an easy way. Secondly, it represents a tractable mechanism of creating dependency: we use a specific combinations of independent variables to generate dependent ones, in exactly the same way it is done, for example, in factor models where independent building blocks (the common and the idiosyncratic factors) are assembled together to create dependency.

The set of parameters $[p_i, u_i, v_i]$ for a given name i defines its behavior in the model. In particular high values of v represent names that, upon default, are extremely infective. This can be used for pivotal names that are regarded as critical for the well being of the entire system. Via v_i it is possible to tune the shock that the idiosyncratic default of entity i has on the rest of the system. Low values of u should be used for names that have weak defenses or,

⁵For the sake of clarity, it is worth mentioning here that we are using the words "infection" and "immunization" in a metaphorical way. Our use has nothing to do with the medical interpretation the two words usually carry.

differently put, are strongly dependent on the health of the rest of the system. The exact economic interpretation of the immunization depends on the nature of the system we are trying to model. The role of p , instead, is to control only the probability of idiosyncratic default, not the final probability of default for the name. It is important to understand that the final probability of default is the result not only of p but also of the rate of infections from the other names as well as its own ability to gain immunization via u .

With respect to the problem of modeling the loss given default (LGD) of each name, the only assumption we had to make is that all losses can be approximated as the product of a discrete number and a minimum loss level D :

$$\text{Approximated } LGD_i = D \cdot d_i, \quad d_i \in \mathbb{N} \quad (4)$$

Note that D represents the minimum loss we can represent inside the model. This strategy to discretize the LGDs is used, for example, in Andersen et al. [2003]. To give an example, assume all the names in the portfolio have LGDs in the set $\{33\%, 40\%, 44\%\}$; if we choose D to be 10%, the above names will be associated with $\{3, 4, 4\}$ units of losses, respectively. The approximated LGDs will hence be $\{30\%, 40\%, 40\%\}$, leading to rounding errors of magnitude $\{3\%, 0\%, 4\%\}$. Note how a finer choice of D , for example $D = 5\%$, leads to the approximated LGDs $\{35\%, 40\%, 45\%\}$ that show smaller rounding errors ($\{2\%, 0\%, 1\%\}$), at the cost of dealing with higher units of losses ($\{7, 8, 9\}$). Higher units of losses translates in heavier calculation burdens as the dimension of the vector used for the portfolio loss distribution increases⁶. We opted for a simple way of implementing this discretization strategy for a given value of D obtained by setting

$$d_i = \text{round}(\text{Expected } LGD_i / D) \quad (5)$$

3 Theoretical results

In this section we will provide some useful results: at first, we will explore single name default probability under the model assumptions. We will then move to calculate the probability of joint default/survival for any subset of names. A few necessary tools and additional notational shortcuts will be introduced along the way.

Proposition 3.1 *Let $A \subseteq \{1 \cdots, n\}$; the probability that at least one name in A spreads an infection is given by the quantity I_A defined as*

$$I_A := \left\{ 1 - \prod_{j \in A} [1 - p_j \cdot v_j] \right\} \quad (6)$$

⁶This instance of the common compromise between accuracy and speed makes the choice of D in real life applications a problem more complicated than one suspects. Parcell [2006] offers a very effective strategy to minimize the rounding errors.

Proof. The probability that an infection starts from inside A is given by

$$P\{X_j \cdot V_j = 1 \text{ for at least one } j \in A\} = 1 - P\{X_j \cdot V_j = 0, \forall j \in A\} \quad (7)$$

We can now use the independence assumption on both X and V to get

$$\begin{aligned} P\{X_j \cdot V_j = 0, \forall j \in A\} &= \\ \prod_{j \in A} P\{X_j \cdot V_j = 0\} &= \\ \prod_{j \in A} [1 - P\{X_j \cdot V_j = 1\}] &= \\ \prod_{j \in A} [1 - P\{X_j = 1, V_j = 1\}] &= \\ \prod_{j \in A} [1 - P\{X_j = 1\} \cdot P\{V_j = 1\}] &= \\ \prod_{j \in A} [1 - p_j \cdot v_j] &= \end{aligned} \quad (8)$$

where the penultimate passage is obtained thanks to the independence of X_j and V_j . ■

3.1 Marginal distribution

We will now focus on single name properties. Let \bar{A} be the complement of set A . The following result gives the formula for the probability of default of single names:

Proposition 3.2 *Let $\tilde{p}_i := P\{Z_i = 1\}$. We have:*

$$\tilde{p}_i = p_i + [1 - p_i] \cdot [1 - u_i] \cdot I_{\{\bar{i}\}} \quad (9)$$

Proof. We have that

$$P\{Z_i = 1\} = \frac{P\{Z_i = 1|X_i = 1\} \cdot P\{X_i = 1\} + P\{Z_i = 1|X_i = 0\} \cdot P\{X_i = 0\}}{P\{X_i = 1\} + P\{X_i = 0\}} \quad (10)$$

It is easy to see that

$$\begin{aligned} P\{Z_i = 1|X_i = 1\} &= 1 \\ P\{X_i = 1\} &= p_i \\ P\{X_i = 0\} &= 1 - p_i \end{aligned}$$

so that the only part left to calculate is $P\{Z_i = 1|X_i = 0\}$; using the expression for Z_i in (2) we have

$$P\{Z_i = 1|X_i = 0\} = P\left\{(1 - U_i) \cdot \left[1 - \prod_{i \neq j} (1 - X_j \cdot V_j)\right] = 1\right\} \quad (11)$$

Both variables in the last expression can only take binary values $(0, 1)$ so their product can only be 1 if they both take value 1. This implies that

$$\begin{aligned} &P\left\{(1 - U_i) \cdot \left[1 - \prod_{i \neq j} (1 - X_j \cdot V_j)\right] = 1\right\} \\ &= P\left\{(1 - U_i) = 1, \left[1 - \prod_{i \neq j} (1 - X_j \cdot V_j)\right] = 1\right\} \end{aligned} \quad (12)$$

We can use the independence assumption between the various building blocks to split the right hand side as

$$P\{(1 - U_i) = 1\} \cdot P\left\{\left[1 - \prod_{i \neq j} (1 - X_j \cdot V_j)\right] = 1\right\} \quad (13)$$

end hence, thanks to proposition 3.1

$$P\{Z_i = 1 | X_i = 0\} = (1 - u_i) \cdot I_{\{\bar{i}\}} \quad (14)$$

that concludes the proof. ■

Intuitively, name i can default in two ways: idiosyncratically (with probability p_i) or by contagion if it survives ($1 - p_i$), fails to defend itself ($1 - u_i$) and an external infection is active ($I_{\bar{i}}$).

3.2 Joint default/survival probability

In this subsection we will show results regarding joint survival/default events. The notation might seem complicated at first so we will add plenty of examples. Let introduce ps , us , bd and id as the vectors defined by

$$\begin{aligned} ps_i &= (1 - p_i) \cdot (1 - u_i) && \text{(non immune) partial survival} \\ us_i &= (1 - p_i) \cdot u_i && \text{(full) unconditional survival} \\ bd_i &= p_i \cdot (1 - v_i) && \text{(non infective) benign default} \\ id_i &= p_i \cdot v_i && \text{infective default} \end{aligned}$$

They will be useful for simplifying formulas later on. Let A be a subset of $\{1, \dots, n\}$ and let also $m_A = |A|$. Given an integer $h \leq m_A$ and two vectors x and y , let introduce the following operator

$$\Lambda_A^h(x, y) = \sum x_{i_1} \cdots x_{i_h} \cdot y_{j_1} \cdots y_{j_{m_A-h}}$$

where the summation is intended over all possible choices of indices i_1, \dots, i_h and $j_1, \dots, j_{m_A-h} \in A$. It is the sum of all possible products of factors of m_A elements of the 2 vectors x, y with exactly h elements taken from x and the rest from y .

Similarly, but now for 3 vectors (x, y, z) rather than 2, we can define

$$\Theta_A^{h,k}(x, y, z) = \sum x_{i_1} \cdots x_{i_h} \cdot y_{j_1} \cdots y_{j_k} \cdot z_{r_1} \cdots z_{r_{m_A-h-k}}$$

It is the sum of all possible products of factors of m_A elements of the 3 vectors x, y, z with exactly h elements of x and k of y ⁷.

To clarify further, let's look at a simple example; assume $A = \{1, 2, 3\}$, we have

$$\Lambda_A^1(x, y) = x_1 y_2 y_3 + y_1 x_2 y_3 + y_1 y_2 x_3$$

⁷Note that we have

$$\Lambda_A^h(x, y) = \Theta_A^{h, m_A-h}(x, y, \cdot) \quad (15)$$

so that we can consider Λ_A as a special case of Θ_A .

as the 3 terms on the right hand side represent all the possible ways we can take products of elements of x and y with only one element coming from x . Similarly:

$$\begin{aligned}\Lambda_A^0(x, y) &= y_1 y_2 y_3 \\ \Lambda_A^2(x, y) &= x_1 x_2 y_3 + x_1 y_2 x_3 + y_1 x_2 x_3 \\ \Lambda_A^3(x, y) &= x_1 x_2 x_3\end{aligned}$$

$$\Theta_A^{1,1}(x, y, z) = x_1 y_2 z_3 + x_1 z_2 y_3 + y_1 x_2 z_3 + z_1 x_2 y_3 + y_1 z_2 x_3 + z_1 y_2 x_3$$

where the last equation shows an example applied to Θ instead. Finally, lets use the following shortcut

$$\Pi_A(x) = \prod_{i \in A} x_i$$

The following result allows to calculate the probability of every possible combination of survival/default:

Proposition 3.3 *Let A and B be two non empty subsets of $\{1, \dots, n\}$ with $A \cap B = \emptyset$. Define with C the set of all the names that are neither in A nor in B : $C = \overline{A \cup B}$ and let finally $P(A, B)$ represent the probability of all names in A default while all names in B survive, i.e.*

$$P(A, B) = P(Z_i = 1, Z_j = 0, \forall i \in A, \forall j \in B)$$

Then, we have

$$P(A, B) = P_1 + P_2 + P_3 \tag{16}$$

with

$$\begin{aligned}P_1 &= \Pi_A(bd) \cdot (1 - I_C) \cdot \Pi_B(1 - p) \\ P_2 &= \left[\sum_{h=1}^{m_A} \sum_{k=0}^{m_A-h} \Theta_A^{h,k}(id, ps, bd) \right] \cdot (1 - I_C) \cdot \Pi_B(us) \\ P_3 &= \left[\sum_{h=0}^{m_A} \Lambda_A^h(p, ps) \right] \cdot I_C \cdot \Pi_B(us)\end{aligned} \tag{17}$$

Proof. The three cases we need to identify are 1) infection free world, 2) infection starting inside A but not in C and 3) infection starting in C (and possibly but not necessarily in A). Lets see each of them in detail.

1. This is the simplest case. Names in A can only default idiosyncratically in non infective way with probability $\Pi_A(bd)$. Names in B don't need to get immunization and only need to avoid idiosyncratic default. The total probability of this case is hence

$$P_1 = \Pi_A(bd) \cdot (1 - I_C) \cdot \Pi_B(1 - p)$$

2. We are still requiring no infection coming from C (with an associated probability of $1 - I_C$) but now we need to consider contagion effects coming from names in A . Given that at least one of these names is required to

default in an infective way, the rest can either be infected (partial survival) or default (both with or without spreading of the infection). Names in B now cannot accept anything less than full, unconditional survival. The total tally is

$$P_2 = \left[\sum_{h=1}^{m_A} \sum_{k=0}^{m_A-h} \Theta_A^{h,k}(id, ps, bd) \right] \cdot (1 - I_C) \cdot \Pi_B(us)$$

3. In this last case, the infection starts from C with probability of I_C . Names in B still need full survival ($\Pi_B(us)$) and names in A can either default (in both ways, infective or not) or default by infection. This leads to the last equation

$$P_3 = \left[\sum_{h=0}^{m_A} \Lambda_A^h(p, ps) \right] \cdot I_C \cdot \Pi_B(us)$$

■

The three components reflect different cases; in particular, the first corresponds to an infection-free world where names in A default on their own in non-infective way and names in B survive without needing necessarily to immunize themselves. In the second component the infection is internal to A and names in B need full immunization. The same is true for the last case where the infection is now starting in C and names in A default either idiosyncratically (in an infective or benign way) or by contagion.

When $B = \bar{A}$, equation (16) gives the probability of exclusive defaults of the entire subset A :

$$P(A, \bar{A}) = \Pi_A(bd) \cdot \Pi_{\bar{A}}(1 - p) + \left[\sum_{h=1}^{m_A} \sum_{k=0}^{m_A-h} \Theta_A^{h,k}(id, ps, bd) \right] \cdot \Pi_{\bar{A}}(us) \quad (18)$$

Similarly, when $B = \emptyset$, we obtain the probability that at least every name of A defaults:

$$P(A, \emptyset) = \left\{ \Pi_A(bd) + \left[\sum_{h=1}^{m_A} \sum_{k=0}^{m_A-h} \Theta_A^{h,k}(id, ps, bd) \right] \right\} \cdot (1 - I_{\bar{A}}) + \left[\sum_{h=0}^{m_A} \Lambda_A^h(p, ps) \right] \cdot I_{\bar{A}} \quad (19)$$

3.2.1 Calculating Θ_A and Λ_A

Equations (16)-(18)-(19) are very interesting but they would be of no practical use if we cannot provide an efficient way of calculating Θ_A and Λ_A . Luckily there is a relationship linking Θ_A and Λ_A to Θ_W and Λ_W where W is a strict subset of A :

Proposition 3.4 *Let A be a subset of $\{1, \dots, n\}$ containing at least 2 elements and let $t \in A$; let also $W = A/t$ be the subset of names of A excluding t . The following identities hold:*

$$\Lambda_A^h(x, y) = x_t \cdot \Lambda_W^{h-1}(x, y) + y_t \cdot \Lambda_W^h(x, y) \quad (20)$$

$$\Theta_A^{h,k}(x, y, z) = x_t \cdot \Theta_W^{h-1,k}(x, y, z) + y_t \cdot \Theta_W^{h,k-1}(x, y, z) + z_t \cdot \Theta_W^{h,k}(x, y, z) \quad (21)$$

Proof. We will only show proof of equation (21) as the one for equation (20) is analogous. Let s be one of the terms that form $\Theta_A^{h,k}(x, y, z)$; there are 3 possible cases for s regarding the value in position t : it contains either x_t , or y_t or z_t . In the first case, we can write

$$s = x_t \cdot g_1$$

where g_1 is a combination of $h - 1$ values taken from x (but not in the t position), k values from y (again not in the t position) and therefore is part of $\Theta_W^{h-1,k}(x, y, z)$. We can repeat the same reasoning for all other terms that contain x_t to get the first part of the right hand side: $x_t \cdot \Theta_W^{h-1,k}(x, y, z)$. The second and third cases are similar. In the second case, i.e. s contains y_t , we can write

$$s = y_t \cdot g_2$$

and we know that g_2 will have h values from x , $k - 1$ from y and therefore will belong to $\Theta_W^{h,k-1}(x, y, z)$; repeating the reasoning for all terms of the form $s = y_t \cdot g_2$ we conclude the second part of the right hand side: $y_t \cdot \Theta_W^{h,k-1}(x, y, z)$. Finally, when

$$s = z_t \cdot g_3$$

we have that g_3 is part of the summation of $\Theta_W^{h,k}(x, y, z)$ leading to the last part of the right hand side: $z_t \cdot \Theta_W^{h,k}(x, y, z)$. ■

As an example, let again $A = \{1, 2, 3\}$; we have already seen that

$$\Lambda_A^1(ps, bd) = ps_1bd_2bd_3 + bd_1ps_2bd_3 + bd_1bd_2ps_3.$$

Using the right hand side of (20) instead we would have

$$\Lambda_A^1(ps, bd) = ps_1 \cdot \Lambda_{A/1}^0(ps, bd) + bd_1 \cdot \Lambda_{A/1}^1(ps, bd)$$

and the identity is satisfied as

$$\Lambda_{A/1}^0(ps, bd) = \Lambda_{\{2,3\}}^0(ps, bd) = bd_2bd_3$$

and

$$\Lambda_{A/1}^1(ps, bd) = \Lambda_{\{2,3\}}^1(ps, bd) = ps_2bd_3 + bd_2ps_3.$$

We can recursively use (20) and (21) until we reach (on the right hand side) the case $W = \{i\}$ for which we have⁸:

$$\Lambda_{\{i\}}^h(ps, bd) = 0, \text{ if } h > 1 \text{ or } h < 0$$

$$\Lambda_{\{i\}}^1(ps, bd) = ps_i, \quad \Lambda_{\{i\}}^0(ps, bd) = bd_i$$

$$\Theta_{\{i\}}^{h,k}(id, ps, bd) = 0, \text{ if } h + k > 1 \text{ or } \min(h, k) < 0$$

⁸These results come straight from the definitions of Θ and Λ .

$$\Theta_{\{i\}}^{1,0}(id, ps, bd) = id_i, \quad \Theta_{\{i\}}^{0,1}(id, ps, bd) = ps_i, \quad \Theta_{\{i\}}^{0,0}(id, ps, bd) = bd_i$$

In the homogeneous case we can get simplified expressions for Θ and Λ :

$$\Theta_A^{h,k}(x, y, z) = C_h^{m_A} \cdot C_k^{m_A-h} \cdot x^h \cdot y^k \cdot z^{m_A-h-k}$$

$$\Lambda_A^h(x, y) = C_h^{m_A} \cdot x^h \cdot y^{m_A-h}$$

4 Portfolio loss distribution

In this section we will present a recursive algorithm that can be used in order to calculate the portfolio loss distribution. Let L_n represents the total amount of losses of a portfolio with n names:

$$L_n = \sum_{i=1, \dots, n} d_i \cdot Z_i$$

For a given integer h , we will show an efficient algorithm for calculating $P\{L_n = h\}$ similar, in spirit, to the one presented by Andersen et al. [2003] for conditionally independent models. In such modeling framework, one can obtain conditional default probabilities that are independent from each other by conditioning on the value of a common factor. Once the conditional default probabilities are obtained, one needs only to compute their convolution to get the portfolio loss distribution (under the chosen value of the common factor). Integrating numerically over the common factor leads to the unconditional portfolio loss distribution.

Andersen, Sidenius, and Basu showed that an efficient way of performing the convolution of the independent conditioned default probabilities is to construct the portfolio loss distribution by adding each name one by one via the following recursive relationship

$$P\{L_{n+1} = h\} = P\{L_n = h\} \cdot (1 - p_j) + P\{L_n = h - d_j\} \cdot p_j \quad (22)$$

where p_j is the conditional probability of default of name j and d_j are the units of losses associated with it.

The intuition behind (22) is the following. There are two ways in which we can obtain h units of losses when adding an extra name: either we already reached h losses with the previous names and the new one survived (first component) or the name defaulted adding d_j units of losses to the $h - d_j$ ones already reached before adding j (second component).

In our model, we can exploit the independence of the building blocks to obtain a similar recursive algorithm that constructs the portfolio loss distribution by adding an extra name; this will be the aim of the rest of this section.

Let start by defining L_n^C as the the units of losses due to contagion events and $L_n^I = L_n - L_n^C$ for losses due to idiosyncratic effects; in mathematical terms

$$L_n^C := \sum_{i=1, \dots, n} d_i \cdot (Z_i - X_i)$$

$$L_n^I := \sum_{i=1, \dots, n} d_i \cdot X_i$$

When adding an extra name to the calculations, there are two situations we need to face: either we are in a *contaminated* world, i.e. there has already been an infectious default, or we are in an *infection-free* case. In the latter case we need to consider the scenario of the first infective default that triggers potential losses, i.e. losses that have not been realized yet but have accumulated as the result of previous names surviving without getting full immunization. Let L_n^R represent the number of units of such potential losses:

$$L_n^R := \sum_{i=1, \dots, n} d_i \cdot [1 - X_i] \cdot [1 - U_i]$$

Note that $P\{L_n^R > 0, L_n^C > 0\} = 0$ as every unit of L_n^R becomes part of L_n^C if a contagion is active.

Let define two quantities that will take the role played by $P\{L_n = h\}$ in (22):

$$\begin{aligned} \alpha_n(h, k) &:= P\{L_n^I = h, L_n^C = 0, L_n^R = k\} \\ \beta_n(h, k) &:= P\{L_n^I = h, L_n^C = k, L_n^R = 0\} \end{aligned}$$

So $\alpha_n(h, k)$ represents the probability of realizing h units of losses in an uncontaminated world of n names, in which there are also k units of losses at risk should an infection appear. On the other hand, $\beta_n(h, k)$ represents the probability of realizing $h + k$ units of losses (in a contaminated universe of n names), of which h are due to idiosyncratic defaults and k are due to pure infection. There are at least two reasons why the above quantities are interesting. The first is that they are sufficient to describe the distribution of L_n :

$$\begin{aligned} P\{L_n = h\} &= \\ P\{L_n^I + L_n^C = h\} &= \\ \sum_k [P\{L_n^I = h, L_n^C = 0\} + \sum_k [P\{L_n^I = k, L_n^C = h - k\}]] &= \\ \sum_k [P\{L_n^I = h, L_n^C = 0, L_n^R = k\}] + \sum_k [P\{L_n^I = k, L_n^C = h - k, L_n^R = 0\}] &= \\ \sum_k \alpha_n(h, k) + \sum_k \beta_n(k, h - k) & \end{aligned} \tag{23}$$

The second is that we can calculate them via a recursive relationship⁹ that links $[\alpha_{n+1}(\cdot, \cdot), \beta_{n+1}(\cdot, \cdot)]$ to $[\alpha_n(\cdot, \cdot), \beta_n(\cdot, \cdot)]$:

$$\begin{aligned} \alpha_{n+1}(h, k) &= (1 - p_j) \cdot u_j \cdot \alpha_n(h, k) &+ \\ &= (1 - p_j) \cdot (1 - u_j) \cdot \alpha_n(h, k - d_j) &+ \\ &= p_j \cdot (1 - v_j) \cdot \alpha_n(h - d_j, k) & \\ \beta_{n+1}(h, k) &= (1 - p_j) \cdot u_j \cdot \beta_n(h, k) + p_j \cdot \beta_n(h - d_j, k) &+ \\ &= (1 - p_j) \cdot (1 - u_j) \cdot \beta_n(h, k - d_j) &+ \\ &= p_j \cdot v_j \cdot \alpha_n(h - d_j, k) & \end{aligned} \tag{24}$$

Boundary conditions:

$$\begin{aligned} \alpha_0(0, 0) &= 1 \\ \alpha_0(i, j) &= 0 \quad \forall (i, j) \neq (0, 0) \\ \beta_0(i, j) &= 0 \quad \forall i, j \end{aligned} \tag{25}$$

⁹A derivation of this system of equations can be found in appendix 6.

It is important to note that the final distribution of the portfolio losses does **not** depend on the order chosen when adding names in the above algorithm. This can be seen by applying twice the system 24. Lets prove it only for α as the β case is similar. Suppose we want to add two names, i first and then j , to a set of m names. In order to shorten the notation, lets indicate with $\bar{p} = 1 - p$, $\bar{v} = 1 - v$ and $\bar{u} = 1 - u$. When we add name j , we would apply (24) to a portfolio of $m + 1$ names obtaining

$$\alpha_{m+2}(h, k) = \begin{aligned} & \bar{p}_j \cdot u_j \cdot \alpha_{m+1}(h, k) && + \\ & \bar{p}_j \cdot \bar{u}_j \cdot \alpha_{m+1}(h, k - d_j) && + \\ & p_j \cdot \bar{v}_j \cdot \alpha_{m+1}(h - d_j, k) \end{aligned} \quad (26)$$

Each of the three terms α_{m+1} on the right hand side can be explicitly written by applying (24) again:

$$\alpha_{m+1}(h, k) = \begin{aligned} & \bar{p}_i \cdot u_i \cdot \alpha_m(h, k) && + \\ & \bar{p}_i \cdot \bar{u}_i \cdot \alpha_m(h, k - d_i) && + \\ & p_i \cdot \bar{v}_i \cdot \alpha_m(h - d_i, k) \end{aligned} \quad (27)$$

$$\alpha_{m+1}(h, k - d_j) = \begin{aligned} & \bar{p}_i \cdot u_i \cdot \alpha_m(h, k - d_j) && + \\ & \bar{p}_i \cdot \bar{u}_i \cdot \alpha_m(h, k - d_j - d_i) && + \\ & p_i \cdot \bar{v}_i \cdot \alpha_m(h - d_i, k - d_j) \end{aligned} \quad (28)$$

$$\alpha_{m+1}(h - d_j, k) = \begin{aligned} & \bar{p}_i \cdot u_i \cdot \alpha_m(h - d_j, k) && + \\ & \bar{p}_i \cdot \bar{u}_i \cdot \alpha_m(h - d_j, k - d_i) && + \\ & p_i \cdot \bar{v}_i \cdot \alpha_m(h - d_j - d_i, k) \end{aligned} \quad (29)$$

Substituting (27), (28) and (29) into (26) and rearranging terms, we can write

$$\begin{aligned} \alpha_{m+2}(h, k) = & \bar{p}_i \cdot u_i \cdot \bar{p}_j \cdot u_j \cdot \alpha_m(h, k) && + \\ & \bar{p}_i \cdot \bar{u}_i \cdot \bar{p}_j \cdot \bar{u}_j \cdot \alpha_m(h, k - d_i - d_j) && + \\ & p_i \cdot \bar{v}_i \cdot p_j \cdot \bar{v}_j \cdot \alpha_m(h - d_i - d_j, k) && + \\ & \bar{p}_i \cdot u_i \cdot \bar{p}_j \cdot \bar{u}_j \cdot \alpha_m(h, k - d_j) + \bar{p}_j \cdot u_j \cdot \bar{p}_i \cdot \bar{u}_i \cdot \alpha_m(h, k - d_i) && + \\ & \bar{p}_i \cdot u_i \cdot p_j \cdot \bar{v}_j \cdot \alpha_m(h - d_j, k) + \bar{p}_j \cdot u_j \cdot p_i \cdot \bar{v}_i \cdot \alpha_m(h - d_i, k) && + \\ & \bar{p}_i \cdot \bar{u}_i \cdot p_j \cdot \bar{v}_j \cdot \alpha_m(h - d_j, k - d_i) + \bar{p}_j \cdot \bar{u}_j \cdot p_i \cdot \bar{v}_i \cdot \alpha_m(h - d_i, k - d_j) \end{aligned}$$

Every line in the last equation is symmetric with respect to i and j and then we can invert the order between i and j without changing the final result¹⁰.

4.1 The time dimension

At this point it is worth emphasizing that the approach we propose is intrinsically static as the dynamics of the variables involved are fixed¹¹. We can use

¹⁰Technically, this only proves that the order of the last 2 names added is not important. It is quite easy to extend the reasoning to the entire sequence using induction on m .

¹¹Of course, every day we could calibrate the model and obtain a different parametrization for each day. This gives information regarding the evolution of the parameters rather than proper dynamics.

the model at different points in time in order to obtain the portfolio loss distribution for any given time t but such valuations will be static snapshots bearing no formal relationship with other time steps. Adding the time dimension in the model is technically possible, for example in the spirit of Cousin et al. [2013], but will compromise the tractability (and the possibility of working with heterogeneous portfolios) we strongly want to preserve.

The above said, we can at least hope in some consistency between loss distributions generated at different time horizons. The following result ensures that the probability of observing no losses is a (decreasing) function only of the p_i . The formal proof can be found in appendix 7 but the intuition behind the result is straightforward as the only way we experience no losses is that every name survives idiosyncratically (default by contagion requires at least one idiosyncratic default).

Proposition 4.1

$$P\{L_n = 0\} = \prod_{j=1, \dots, n} (1 - p_j) \tag{30}$$

Assume we use the model at two time points $t_1 < t_2$; using the above result, we can be sure that the probability of observing no losses will decrease in time if the probabilities of idiosyncratic default in the two specifications of the model are increasing, i.e.

$$p_i(t_1) \leq p_i(t_2), \quad \forall i \implies P\{L_n(t_1) = 0\} \geq P\{L_n(t_2) = 0\}$$

where we indicated with $x(t)$ the quantity x measured at time t .

5 Conclusions

In this paper we presented a new model that uses Davis and Lo [2001] as a starting point. Unlike other extensions of such model, the one introduced here can achieve reasonable performances with heterogeneous portfolios as we provided both theoretical and practical results for the efficient computation of the portfolio loss distribution. The price we had to pay in comparison to the original model is a reduction of flexibility.

References

- Leif Andersen, Jakob Sidenius, and Susanta Basu. All your hedges in one basket. *Risk Magazine*, 16(11):67–72, 2003.
- Shariar Azizpour, Kay Giesecke, et al. Self-exciting corporate defaults: contagion vs. frailty. *Working Paper*, 2008.
- Areski Cousin, Diana Dorobantu, and Didier Rullière. An extension of Davis and Lo’s contagion model. *Quantitative Finance*, 13(3):407–420, 2013.
- M Davis and V Lo. Infectious defaults. *Quantitative Finance*, 1(4):382–387, 2001.
- Daniel Egloff, Markus Leippold, and Paolo Vanini. A simple model of credit contagion. *Journal of Banking & Finance*, 31(8):2475–2492, 2007.
- Rüdiger Frey and Jochen Backhaus. Dynamic hedging of synthetic cdo tranches with spread risk and default contagion. *Journal of Economic Dynamics and Control*, 34(4):710–724, 2010.
- Kay Giesecke and Stefan Weber. Cyclical correlations, credit contagion, and portfolio losses. *Journal of Banking & Finance*, 28(12):3009–3036, 2004.
- Robert A Jarrow and Fan Yu. Counterparty risk and the pricing of defaultable securities. *The Journal of Finance*, 56(5):1765–1799, 2001.
- Peter Neu and Reimer Kühn. Credit risk enhancement in a network of interdependent firms. *Physica A: Statistical Mechanics and its Applications*, 342(3):639–655, 2004.
- Ed Parcell. Loss unit interpolation in the collateralised debt obligation pricing model. *Working Paper*, 2006.
- Daniel Rösch and Birker Winterfeldt. Estimating credit contagion in a standard factor model. *Risk*, pages 78–82, 2008.
- Ayaka Sakata, Masato Hisakado, and Shintaro Mori. Infectious default model with recovery and continuous limits. *Journal of the Physical Society of Japan*, 76(5):054801, 2007.
- Philipp J Schönbucher and Dirk Schubert. Copula-dependent default risk in intensity models. *Working Paper*, 2001.
- Fan Yu. Correlated defaults in intensity-based models. *Mathematical Finance*, 17(2):155–173, 2007.

6 Derivation of the system (24)

In order to obtain a set of equations for $\alpha_n(\cdot, \cdot)$, consider that there are 3 ways of reaching $\alpha_{n+1}(h, k)$ starting from $\alpha_n(h, k)$ and adding a new name:

1. **Full survival**

$$(1 - p) \cdot u \cdot \alpha_n(h, k) \quad (31)$$

The name survives $(1 - p)$ and protects itself from future aggressions (u) . No losses are realized neither potential ones added.

2. **Partial survival**

$$(1 - p) \cdot (1 - u) \cdot \alpha_n(h, k - d) \quad (32)$$

The name survives $(1 - p)$ but fails to protect itself against future aggressions $(1 - u)$. Its d units of losses are at risk should an infection spread.

3. **Non-infectious default**

$$p \cdot (1 - v) \cdot \alpha_n(h - d, k) \quad (33)$$

The name defaults directly (p) but it is not trying to start an infection $(1 - v)$.

Similarly, there are 4 ways of reaching $\beta_{n+1}(h, k)$:

1. **Full survival**

$$(1 - p) \cdot u \cdot \beta_n(h, k) \quad (34)$$

The name survives $(1 - p)$ and protects itself against the current and future infections (u) .

2. **Default by contagion**

$$(1 - p) \cdot (1 - u) \cdot \beta_n(h, k - c) \quad (35)$$

The name survives $(1 - p)$ but fails to protect itself against the existing infection $(1 - u)$.

3. **Direct default**

$$p \cdot \beta_n(h - d, k) \quad (36)$$

The name defaults (p) and in this case we don't need to consider separately the cases in which it spreads or not the infection as we are already in an infected world.

4. **First infection**

$$p \cdot v \cdot \alpha_n(h - d, k) \quad (37)$$

The name defaults (p) and spreads the contagion (v) in a previously uncontaminated world causing the k units of potential losses to become real ones.

Putting together the previous equations, we get system (24).

7 Proof of equation (30).

Lets write here again the formula we want to prove:

$$P\{L_n = 0\} = \prod_{j=1, \dots, n} (1 - p_j) \quad (38)$$

Proof. The assert can be easily proved by recursion on n . Lets start with the case $n = 2$. A direct application of equation (23) leads to

$$P\{L_2 = 0\} = \alpha_2(0, 0) + \alpha_2(0, d_1) + \alpha_2(0, d_2) + \alpha_2(0, d_1 + d_2)$$

For each of the four terms on the right hand side we can apply recursively equation 24 (keeping in mind 25) and obtain

$$\begin{aligned} \alpha_2(0, 0) &= (1 - p_1) \cdot (1 - u_1) \cdot (1 - p_2) \cdot (1 - u_2) \\ \alpha_2(0, d_1) &= (1 - p_1) \cdot (u_1) \cdot (1 - p_2) \cdot (1 - u_2) \\ \alpha_2(0, d_2) &= (1 - p_1) \cdot (1 - u_1) \cdot (1 - p_2) \cdot (u_2) \\ \alpha_2(0, d_1 + d_2) &= (1 - p_1) \cdot (u_1) \cdot (1 - p_2) \cdot (u_2) \end{aligned}$$

Summing the four equations above we get

$$P\{L_2 = 0\} = (1 - p_1) \cdot (1 - p_2) \cdot [(1 - u_1)(1 - u_2) + u_1(1 - u_2) + (1 - u_1)u_2 + u_1u_2]$$

from which the assert as

$$(1 - u_1)(1 - u_2) + u_1(1 - u_2) + (1 - u_1)u_2 + u_1u_2 = 1$$

Assume now that (38) holds for $n - 1$ and lets prove it for n . We have

$$P\{L_n = 0\} = P\{L_n = 0 | L_{n-1} = 0\} \cdot P\{L_{n-1} = 0\}$$

The above is true since $P\{L_n = 0 | L_{n-1} > 0\} = 0$. Thanks to the inductive hypothesis we have that $P\{L_{n-1} = 0\} = \prod_{j=1, \dots, n-1} (1 - p_j)$; coming to $P\{L_n = 0 | L_{n-1} = 0\}$, this represents the probability that a portfolio of n names suffers no losses given the already $n - 1$ names experience no defaults. It is then immediate to see that the only way this can be possible is that also the n th name does not default idiosyncratically. Exploiting the independence among the various components, we get

$$P\{L_n = 0 | L_{n-1} = 0\} = (1 - p_n)$$

and hence the assert. ■

Editor

Department of Management, Economics and Quantitative Methods
University of Bergamo
Via dei Caniana n° 2
24127 Bergamo
Italy
Tel. +39 35 2052538
Fax +39 35 2052549

The editorial office has complied with its obligations under Article 1, D.L.L. 31.8.1945, n. 660 and subsequent amendments