

Incorrectly Applying Default Correlation Theory: The Causes of the Subprime Mortgage Crisis of 2008

Sehyeok Park & Rajit Chatterjea

Received October 24, 2024

Accepted April 30, 2025

Electronic access May 31, 2025

This paper argues that the incorrect application of mathematical models, particularly the Gaussian copula, played a critical role in the 2007-2008 financial crisis. The misinterpretation of default correlation theory led to systemic underestimation of credit risk, exacerbating the collapse of mortgage-backed securities and collateralized debt obligations. Using a detailed mathematical analysis, this study examines how financial institutions misapplied risk assessment models and how flaws in risk management within the insurance sector contributed to market instability. Key findings indicate that the reliance on oversimplified correlation assumptions led to a fragile financial system, ultimately triggering widespread economic fallout. The paper highlights the necessity of refining risk models to better capture financial dependencies and prevent future crises.

1 Introduction

As remarked by¹, there is no single factor that is solely responsible for the crisis that unfolded in 2007-2008, although it is acknowledged that the main issue was the transfer of risk of mortgage default between two parties: mortgage lenders on one hand and the banks, hedge funds and insurance companies on the other. This happened through a process called securitization, which is undertaken because institutions usually seek to reduce their costs as well as tax obligations.

1.1 Creation of Collateralized Debt Obligations

Using the terminology in¹, a Collateralized Debt Obligation (CDO) is a product that is purchased or arranged through the process of taking out loans or assets. A common act we may have experienced such as student loans, auto loans, margin stocks are all part of CDOs. The process of creating a CDO is simple. Banks gather various assets such as mortgages, debts, etc and make them into a whole, the level of the CDO is determined by the consumer. Then with the funds the banks receive from the consumers, they regenerate new assets.

There is a wide pool of benefits of loaning and purchasing a CDO. One of the biggest part of CDOs attracting consumers is that it is available for majority of the consumers while receiving the needed assets immediately and also possibly enhancing the consumers credit. A purchase of a CDO would not only affect the consumer, but has an even larger impact on the economic side. The funds the banks receive will be used to create other assets and produce liquidity for the financial market.

However, when purchasing a CDO, the consumer should be cautious and look over the liquidity, structural, rating, credit

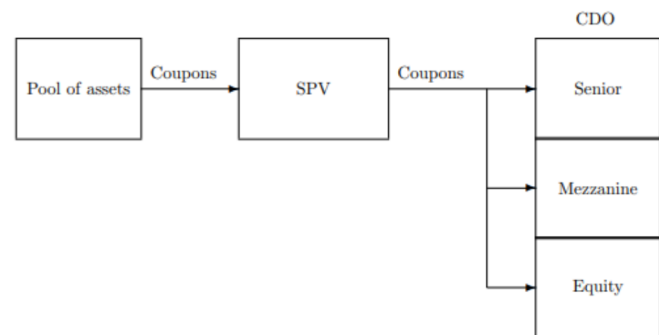


Fig. 1 CDO Structure

risk, and many other factors. Most commonly people lack the understanding of what a CDO actually is, and how much damage it can cost if not properly handled. The risk of a CDO is often overlooked due to its complexity making the investors difficult to understand. CDOs can be seen as a complicated box filled with matching pieces. In many cases investors do not know or fully understand what is inside the box, which could lead to a higher risk than what they have anticipated. Liquidity risk is a risk of when a liquidity could become an illiquidity, meaning the the asset has decreased in value and there is no purchaser wanting to buy the liquidity. Structural risk is the one associated with the matching puzzles. It has various tranches creating cash flows. Senior tranches receive the payments earlier than the junior and the junior receiving payments earlier than the lower, equity tranches; which results in the lower the tranches is the riskier it is. most assets also have a rating risk, the risk of damage a business can impact towards the industry or company, which is proven to be inaccurate most the times.

Class	Rating	Coupon	% Capital Structure
A	Aaa/AAA	Libor + 45bp	70%
B	A2/A	Libor + 145bp	15%
C	Baa2/BBB-	Libor + 245bp	7%
D	Ba3/NR	Libor + 645bp	4%
Equity	Not Rated	Expected Return 25% - 30%	4%

Table 1 CDO Table

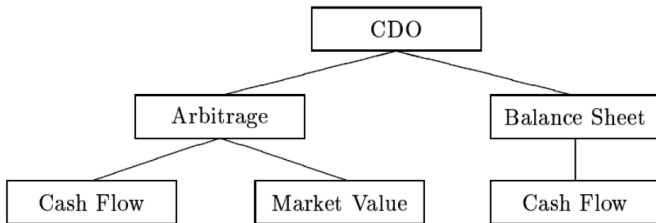


Fig. 2 Assets Process

1.2 Securitization

As there are various causes to this crisis, the most impactful was the failure of securitization, an act of arranging and organizing loans and mortgages to create profit in the capital market. The pool of assets being arranged into a package are consumer loans, debts, mortgages, and any other illiquid assets. This action is not used to make profit but to minimize the loss of the bank. When making the portfolios, the banks consider tranches, different section that is made consisting the types of assets. The portfolios made become accessible to the public, attracting investors for a fixed rate of return: the money generated from the assets within the portfolio.

What role did securitization play in causing the mortgage crisis? During this time banks formed risky portfolios such as CDOs and MBS (mortgage-backed securities) which were misrated leading people to think of it as reliable and safe investments. The misleading information has spread, causing many investors to dive in without the proper knowledge of the products. As housing prices declined and interest rates increased, people started defaulting their mortgages. The banks that have bought these securities fell into bankruptcy, leading to a cause of the mortgage crisis of 2008.

1.3 Hazard Rate Functions

Following,² suppose X is a discrete random variable assuming values in $N = (0, 1, \dots)$ with probability mass function $f(x)$ and survival function $S(x) = P(X \geq x)$. We can think of X as the random lifetime of a device that can fail only at times in N . The hazard rate function of X is defined as:

$$h(x) = \frac{f(x)}{S(x)} \quad (1)$$

at points x for which $S(x) > 0$. The hazard rate is also called the failure rate or intensity function. If X has a finite support $(0, 1, \dots, n)$, $n < \infty$, then $h(n) = 1$.

So given $f(x)$ or $S(x)$ we can determine $h(x)$. Here is how:

$$h(x) = \frac{S(x) - S(x+1)}{S(x)} \iff \frac{S(x+1)}{S(x)} = 1 - h(x) \iff S(x) = \prod_{t=0}^{x-1} (1 - h(t)) \quad (2)$$

Here we assume that $x \geq 1$. If $x = 0$ then $S(x) = 1$. This form of $h(x)$ shows that we can use it to model the life distribution. In this regard we have the following result:

A necessary and sufficient condition that $h : N \rightarrow [0, 1]$ is the hazard rate function of a distribution with support N is that $h(x) \in [0, 1]$ for $x \in N$ and $\sum_{x=0}^{\infty} h(x) = \infty$

In this case the probability mass function $f(x) = h(x) \prod_{t=0}^{x-1} (1 - h(t))$.

Associated with this idea is the concept of a random variable called the time until default, also known as the survival time. This will apply to a security for instance. In order to define the time until default, we need a clearly defined origin of time, a time scale for measuring time and what actually defines a default. The time origin is defined to be the current time, the time scale is the years used for continuous models, or number of periods, and the default is defined by credit rating agencies such as Moody's.

The survival function defined earlier $S(x)$ gives the probability that the security will attain the age x . Suppose we have an existing security A . The time until default, T is a continuous random variable that measures the length of time from today until the point in time when default occurs. Suppose $F(t)$ represent the distribution function of T i.e. $F(t) = P(T \leq t), t \geq 0$. So here $S(t) = 1 - F(t) \equiv Pr(T > t), t \geq 0$, with the assumption that $F(0) = 0 \iff S(0) = 1$. This is reasonable since every security will exist at time $t = 0$. The probability density function is defined as:

$$f(t) = F'(t) = -S'(t) = \lim_{\Delta \rightarrow 0^+} \frac{Pr(t \leq T < t + \Delta)}{\Delta} \quad (3)$$

For a security that has survived x years, the future lifetime for that security is $T - x$, as long as $T > x$. The following notations are used in this regard:

$${}_t q_x = Pr(T - x \leq t | T > x), t \geq 0$$

$${}_t p_x = 1 - {}_t q_x = P(T - x > t | T > x), t \geq 0$$

${}_t q_x$ is understood as the probability that the security A will default within the next t years, given that it survives for x years. For $X = 0 \iff {}_t p_0 = S(t), x \geq 0$.

For $t = 1$, one writes $p_x = Pr(T - x > 1 | T > x)$ and likewise q_x is defined. q_x is termed the marginal default probability. This is the probability of default in the next year, conditional on the survival until the beginning of the year.

Another way to define the Hazard rate function can be the ratio $\frac{f(x)}{1-F(x)}$. The Hazard rate function gives the instantaneous probability of default for a security that is of age x . To write this another way:

$$Pr(x < T \leq x + \Delta x | T > x) = \frac{F(x + \Delta x) - F(x)}{1 - F(x)} \sim \frac{f(x)\Delta x}{1 - F(x)} \quad (4)$$

$$\text{From here, } h(x) = \frac{f(x)}{1-F(x)} = -\frac{S'(x)}{S(x)} \iff S(t) = e^{-\int_0^t h(s)ds}$$

From here,

$${}_t p_x = e^{-\int_0^t h(s+x)ds}, {}_t q_x = 1 - e^{-\int_0^t h(s+x)ds} \quad (5)$$

Now, $F(t) = 1 - S(t) = 1 - e^{-\int_0^t h(s)ds}$ and $f(t) = S(t)h(t)$. If the hazard rate h is constant over a period $[x, x + 1]$. If this is true then $f(t) = he^{-ht}$. This means that the survival time follows an exponential distribution with parameter h . The survival probability over the interval $[x, x + t]$ for $0 < t \leq 1$ is given by:

$${}_t p_x = 1 - {}_t q_x = e^{-\int_0^t h(s)ds} = e^{-ht} = (p_x)^t$$

The crucial idea here is that the mathematical model of a default process is similar to modeling a hazard function. Li (2000) gives several reasons for such an assumption:

1. We get information on the immediate default risk of each entity which is known to be alive at time t
2. Groups comparisons are easier with hazard rate functions
3. Hazard rate functions can be adapted to cases of stochastic default fluctuations
4. Hazard rate functions are similar to short rate models in the context of interest rate derivatives

With this assumption, the joint survival function of two entities A and B, with survival times T_A, T_B is given by $S_{T_A, T_B}(s, t) = Pr(T_A > s, T_B > t)$. The joint distributional function is:

$$F(s, t) = Pr(T_A \leq s, T_B \leq t) = 1 - S_{T_A}(s) - S_{T_B}(t) + S_{T_A, T_B}(s, t) \quad (6)$$

With this background, we can define the default correlation of two entities A and B with respect to their survival times T_A and T_B :

$$\rho_{AB} = \frac{Cov(T_A, T_B)}{\sqrt{Var(T_A)Var(T_B)}} = \frac{E(T_A T_B) - E(T_A)E(T_B)}{\sqrt{Var(T_A)Var(T_B)}} \quad (7)$$

This is also called the survival time correlation. This is more general compared to the discrete default correlation that depends on a single period. The discrete default correlation can be written as follows. Suppose $f(s, t)$ represents the joint distribution of two survival times T_A, T_B , let $E_1 = [T_A < 1], E_2 = [T_B < 1]$. Then the discrete default correlation is defined as $q_1 q_2 = Pr[E_1 E_2] = \int_0^1 \int_0^1 f(s, t) ds dt, q_1 = \int_0^1 f_A(s) ds, q_2 = \int_0^1 f_B(t) dt$.

2 How insurance theory works

Following³ an insurance system is a mechanism for reducing the adverse financial impact of random events that prevent fulfillment of reasonable expectations.

Insurance actually has a mathematical construct. Let's look at an example. Suppose a decision maker has wealth w and faces a loss in the next period. We use a random variable, X , to model this loss. Suppose an insurance contract pays $I(x)$ for the loss of x . Since the contract is feasible, $0 \leq I(x) \leq x$. It is assumed that all feasible contracts with $E[I(X)] = \beta$ can be purchased for the same price P . The decision maker has a utility function $u(w)$ and is risk averse so that $u''(w) < 0$, and has decided on the value P . With the setting, the question is, which insurance contract must be purchased to maximize the expected utility of the decision maker, given the values of β and the premium to be paid P ?

A typical insurance contract only pays out when the loss amount is above a deductible amount d :

$$I_d(x) = \begin{cases} 0 & \text{if } x < d \\ x - d & \text{if } x \geq d \end{cases} \quad (8)$$

This type of insurance is called stop-loss or excess-of-loss insurance. This leads to our first result in this area.

If a decision maker has wealth w , is risk averse so that $u''(w) < 0$, faces a random loss X and will spend P on insurance, and the insurance market offers for a payment of P all feasible insurance contracts of the form $I(x), 0 \leq I(x) \leq x$, with $E[I(X)] = \beta$, with the decision maker having an expected

utility being maximized by purchasing an insurance policy of the form:

$$I_d(x) = \begin{cases} 0 & \text{if } x < d \\ x - d^* & \text{if } x \geq d \end{cases} \quad (9)$$

then the quantity d^* is a solution of the equation $\beta - \int_d^\infty (x - d)f(x)dx = 0$

In the above, the expected claims are the β . So if $f(x)$ represents the pdf and $F(x)$ denotes the cdf of the random loss X , we are assuming that $\beta = \int_d^\infty (x - d)f(x)dx \equiv \beta = \int_d^\infty (1 - F(x))dx$. These are equivalent maybe shown by integrating by parts.

We need a result to prove the idea above. Suppose we have a risk-averse investor i.e. $u''(w) < 0$, and suppose this is true for all w in the interval $[a, b]$. If $u''(w) < 0$ for all w in $[a, b]$, then for w and z in $[a, b]$,

$$u(w) - u(z) \leq (w - z)u'(z)$$

We use the fact that a line tangent to $u(w)$ at the point $(z, u(z))$ has the equation: $y - u(z) = u'(z)(w - z)$ and is above the graph of the function $u(w)$ except at the point $(z, u(z))$. So:

$$u(w) - u(z) \leq u'(z)(w - z)$$

The proof can easily be shown to hold in the case $u'(w) < 0$.

Now, recall that by $I(x)$ we mean the set of all feasible insurance contracts. Using the above result:

$$\begin{aligned} u(w - x + I(x) - P) - u(w - x + I_{d^*}(x) - P) \\ \leq (I(x) - I_{d^*}(x))u'(w - x + I_{d^*}(x) - P) \end{aligned} \quad (10)$$

We also have:

$$(I(x) - I_{d^*}(x))u'(w - x + I_{d^*}(x) - P) \leq (I(x) - I_{d^*}(x))u'(w - d^* - P) \quad (11)$$

To see the above equation, we consider three cases:

- If $I_{d^*}(x) = I(x)$ then both sides of the above is 0, so it holds trivially
- If $I_{d^*}(x) > I(x)$. In this case, $I_{d^*} > 0$ so that $x - I_{d^*}(x) = d^*$ using the fact that the insurance payout only occurs above a certain deductible, so that equality holds in the above with each side equal to $(I(x) - I_{d^*}(x))u'(w - d^* - P)$
- Finally consider $I_{d^*}(x) < I(x)$. Again using the principle of minimum deductible payment, $I_{d^*}(x) - x \geq -d^*$ and $I_{d^*}(x) - x - P \geq -d^* - P$, so that we have: $u'(w - x + I_{d^*}(x) - P) \leq u'(w - d^* - P)$, using the fact that $u''(x)$ is negative and that $u'(x)$ is decreasing.

So we have shown that in all cases:

$$(I(x) - I_{d^*}(x))u'(w - x + I_{d^*}(x) - P) \leq (I(x) - I_{d^*}(x))u'(w - P - d^*) \quad (12)$$

Combining this with $u(w - x + I(x) - P) - u(w - x + I_{d^*}(x) - P) \leq (I(x) - I_{d^*}(x))u'(w - x + I_{d^*}(x) - P)$ we get: $E(u(w - X + I(x) - P) - E(u(w - X + I_{d^*}(x)))u'(w - d^* - P) = (\beta - \beta)u'(w - d^* - P) = 0$ So we have:

$$E(u(w - X_I(X) - P)) \leq E(u(w - x + I_{d^*}(x) - P)) \quad (13)$$

So the way to maximize the expected utility is to select $I_{d^*}(x)$ which is the stop-loss policy.

2.1 Collective Risk Models: An Introduction

Based on the work of⁴, we define a collective risk model assumes that there is a random process which generates claims for an entire portfolio. The idea here is to think of the portfolio in its entirety instead of a smaller subset. Suppose N denotes the number of claims produced by a portfolio in a specific time period. Denote by X_i the amount produced by claim i . Set $S = \sum_{i=1}^N X_i$. This represents the aggregate of claims. Here N is itself a random variable which is related to the frequency of the claim. Collective risk models assume the following:

- The variables X_i are identically distributed random variables
- The random variables N, X_1, X_2, \dots are mutually independent

Since the X_i variables are i.i.d., let $P(x)$ represent the common distribution function of these variables. Suppose X is a random variable with this distribution function. Set $p_k = E[X^k]$. This is the k th moment about the origin. Set $M_X(t) = E[e^{tX}]$ to be the moment generating function of X . Let $M_N(t) = E[e^{tN}]$ be the moment generating function of the number of claims, and let $M_S(t) = E[e^{tS}]$ be the moment generating function of aggregate claims. Denote by $F_S(s)$ to be the distribution function of the aggregate claims. Recall the following formulas for mean and variance:

$$E[W] = E[E[W|V]] \quad \text{Var}[W] = \text{Var}(E[W|V]) + E[\text{Var}(W|V)] \quad (14)$$

Let's see why these are true.

It is straightforward to show that $E[g(X)] = E[E[g(X)|Y]]$. To see this note the following steps:

$$\begin{aligned} E[g(X)|Y = y] &= \sum_{x=0}^{\infty} g(x)f_{X|Y}(x|y) \\ &= \sum_{x=0}^{\infty} g(x) \frac{P(X=x, Y=y)}{P(Y=y)} \end{aligned}$$

$$\begin{aligned} \text{Now, } E[E[g(X)|Y]] &= \sum_{y=0}^{\infty} E[g(X)|Y = y]P(Y = y) \\ &= \sum_{y=0}^{\infty} \sum_{x=0}^{\infty} g(x) \frac{P(X=x, Y=y)}{P(Y=y)} P(Y = y) \\ &= \sum_{x=0}^{\infty} g(x) \sum_{y=0}^{\infty} P(X = x, Y = y) \end{aligned}$$

$$= \sum_{x=0}^{\infty} g(x)P(X = x) = E[g(X)]$$

Setting $g(X) = X$ we get the desired result.

Now to prove the second statement, set $g(X) = X^2$ to get $E[X^2] = E[E[X^2|Y]]$. Since $Var(X|Y) = E(X^2|Y) - (E(X|Y))^2$, we have:

$$E[Var(X|Y)] = E(E(X^2|Y)) - E(E(X|Y))^2 = E(X^2) - E(E(X|Y))^2$$

. From here:

$$Var(E(X|Y)) = E(E(X|Y)^2) - (E(E(X|Y)))^2 = E(E(X|Y)^2) - (E(X))^2$$

The result follows from here.

Now let's use these results.

$$E(S) = E(E(S|N)) = E(p_1N) = p_1E(N)$$

$$Var(S) = E(Var(S|N)) + Var(E(S|N))$$

$$= E(NVar(X)) + Var(p_1N)$$

$$= E(N)Var(X) + p_1^2Var(N)$$

$$Var(S) = E[Var(S|N)] + Var[E(S|N)]$$

$$= E[NVar(X)] + Var(p_1N)$$

$$= E[N]Var(X) + p_1^2Var(N)$$

$$\text{Here } Var(X) = p_2 - p_1^2$$

These statements simply mean that the expected value of aggregate claims is the product of the expected number of claims and the expected individual claim. The variance of aggregate claims has a component which depends on the variability of the number of claims and a component that depends on the variability of an individual claim.

The moment generating function of S is written as:

$$M_S(t) = E[e^{tS}] = E[E[e^{tS}|N]] = E[M_X(t)^N] = M_N[\log M_X(t)] \quad (15)$$

We will need some background in sums of random variables now.

2.2 Sums of random variables

Suppose we have $S = X + Y$, the sum of two random variables. We are concerned with the event $S = X + Y \leq s$, so that the distribution function of S is:

$$F_S(s) = Pr(S \leq s) = Pr(X + Y \leq s)$$

We can write this as:

$$F_S(s) = \sum_{\text{all } y \leq s} Pr(X + Y \leq s | Y = y) Pr(Y = y)$$

$$= \sum_{\text{all } y \leq s} Pr(X \leq s - y | Y = y) Pr(Y = y)$$

$$= \sum_{\text{all } y \leq s} F_X(s - y) f_Y(y)$$

The probability density function is obtained by replacing F by f . We can write these as:

$$F_S(x) = \int_0^s Pr(X \leq s - y | Y = y) f_Y(y) dy$$

$$= \int_0^s F_X(s - y) f_Y(y) dy$$

$$f_S(s) = \int_0^s f_X(s - y) f_Y(y) dy$$

These integrals are called convolutions. This can be used to see the distribution of multiple random variables i.e. $S =$

$X_1 + X_2 + \dots + X_k$. If F_i is the distribution function of X_i , and $F^{(k)}$ is the distribution function of the sum $X_1 + \dots + X_k$, then the recursion $F^{(n)} = F_n * F^{(n-1)}$ exists. Here $*$ represents the convolution operator. So for instance $F^{(2)} = F_2 * F_1$.

2.3 Distribution of aggregate claims

As an example, suppose N , the number of claims, has a geometric distribution given by:

$Pr(N = n) = pq^n, n = 0, 1, 2, \dots$ with $0 < q < 1$ and $p = 1 - q$. In this case, $M_N(t) = E[e^{tN}] = \frac{p}{1 - qe^t}$, so that $M_S(t) = \frac{p}{1 - qM_X(t)}$. In this case, the distribution function of S is given by:

$$F_S(x) = Pr(S \leq x) = \sum_{n=0}^{\infty} Pr(S \leq x | N = n) Pr(N = n) = \sum_{n=0}^{\infty} Pr(X_1 + X_2 + \dots + X_n \leq x) Pr(N = n) = P * P * \dots * P(x) = P^{*n}(x)$$

So we have $F_S(x) = \sum_{n=0}^{\infty} P^{*n}(x) Pr(N = n)$

2.4 Applications of Risk Theory

Typically we are interested in a compound Poisson distribution, which can be written as follows: (N is random variable)

$$N \sim \text{Poisson}(\lambda)$$

$$Y = \sum_{n=1}^N X_n$$

Here each of the X_i variables are i.i.d. which are also independent of N .

Suppose $U(t)$ represents the insurer's surplus, as in, the excess of the initial fund together with the premiums collected over claims that have been paid. Denote by $c(t)$ the premiums collected through time t and by $S(t)$ the aggregate claims paid through time t . Suppose u is the surplus at time 0 then:

$$U(t) = u + c(t) - S(t), t \geq 0 \quad (16)$$

$U(t)$ is usually called the surplus process, $S(t)$ is called the aggregate claims process. Suppose the premium rate is $c > 0$ and is constant. Say $c(t) = ct$, a linear function. If surplus becomes negative, we will say that ruin has occurred. The time of ruin, $T = \min\{t : t \geq 0, U(t) < 0\}$ is considered ∞ if $U(t) \geq 0, \forall t$. Consider the function $\psi(u, t) = Pr(T < t)$, which represents the probability of ruin before time t . Denote by U_n to be the discrete time surplus process: $U_n : n = 0, 1, 2, \dots$

With this setting, suppose u is the initial surplus, and we are looking at n periods, then $U_n = u + nc - S_n$. Here S_n is the aggregate of claims in the first n periods. Suppose W_i represents the sum of claims in period i , and assume that these are all i.i.d. with $E[W_i] = \mu < c$.

So here, $U_n = u + (c - W_1) + (c - W_2) + \dots + (c - W_n)$. Set $\tilde{T} = \min\{n : U_n < 0\}$ to be the time of ruin. Set $\tilde{\psi}(u) = Pr(\tilde{T} < \infty)$ to be the probability of ruin. Define the adjustment coefficient \tilde{R} to be a positive solution of this equation:

$$M_{W-c}(r) = E[e^{r(W-c)}] = 1$$

2.5 Specific properties of the Poisson Distribution

We start with a theorem following², where we look at sums of Poisson random variables.

If S_1, S_2, \dots, S_m be mutually independent random variables such that S_i has a compound Poisson distribution with parameter i and the distribution function of the claim amount is $P_i(x)$, with $i = 1, 2, \dots, m$, then $S = S_1 + S_2 + \dots + S_m$ has a compound Poisson distribution with $\lambda = \sum_{i=1}^m \lambda_i$ and $P(x) = \sum_{i=1}^m \frac{\lambda_i}{\lambda} P_i(x)$

We will use moment generating functions. Let $M_i(t)$ represent the moment generating function of $P_i(x)$. Then the m.g.f. of S_i is $M_{S_i}(t) = e^{\lambda_i[M_i(t)-1]}$. Since S_1, S_2, \dots, S_m were assumed to be independent, the m.g.f. of their sum is:

$$M_S(t) = \prod_{i=1}^m M_{S_i}(t) = e^{\sum_{i=1}^m \lambda_i[M_i(t)-1]} \quad (17)$$

From there we can write $M_S(t) = e^{\lambda[\sum_{i=1}^m \frac{\lambda_i}{\lambda} M_i(t)-1]}$

We will address the following question. Suppose x_1, x_2, \dots, x_m be m different numbers and suppose that N_1, N_2, \dots, N_m are mutually independent random variables. Suppose each N_i has a Poisson distribution with parameter λ_i . We seek the distribution of $x_1N_1 + x_2N_2 + \dots + x_mN_m$. In order to solve this problem, we interpret x_iN_i to have a compound distribution with Poisson parameter λ_i .

Then this sum $x_1N_1 + \dots + x_mN_m$ has a compound Poisson distribution with $\lambda = \sum_{i=1}^m \lambda_i$, the probability function $p(x)$ is defined to be $\frac{\lambda_i}{\lambda} x = x_i, i = 1, 2, \dots, m$, and 0 otherwise. Let x_1, x_2, \dots, x_m denote the discrete values for individual claim amounts, and suppose $\pi_i = p(x_i), i = 1, 2, \dots, m$. In addition, recall that the multinomial probability distribution has the form:

$$Pr(N_1 = n_1, N_2 = n_2, \dots, N_m = n_m) = \frac{n!}{n_1!n_2!n_m!} \pi_1^{n_1} \pi_2^{n_2} \dots \pi_m^{n_m} \quad (18)$$

With the S_i defined in Theorem (2.2) having a compound Poisson distribution with parameter λ and probability function of claim amounts given by the discrete probability function $\pi_i = p(x_i), i = 1, 2, \dots, m$ two properties can be shown:

- N_1, N_2, \dots, N_m are mutually independent
- N_i has a Poisson distribution with parameter $\lambda_i = \lambda \pi_i, i = 1, 2, \dots, m$

We will see why this is true. Suppose there are m claim amounts, the number of claims having a multinomial distribution with parameters $n, \pi_1, \pi_2, \dots, \pi_m$. So now, we are told $N = \sum_{i=1}^m N_i = n$, we do the following calculation:

$$\begin{aligned} E[e^{\sum_{i=1}^m t_i N_i}] &= \\ &= \sum_{n=0}^{\infty} E[e^{\sum_{i=1}^m t_i N_i} | N = n] Pr(N = n) = \\ &= \sum_{n=0}^{\infty} (\pi_1 e^{t_1} + \dots + \pi_m e^{t_m})^n \frac{e^{-\lambda} \lambda^n}{n!} \quad (19) \end{aligned}$$

2.6 The Theory of Copulas

Following the methodology of⁵, we need a few definitions to start off. Denote by \mathbb{R} the interval $(-\infty, \infty)$, by \bar{R} the interval $[-\infty, \infty]$. A rectangle in \bar{R}^2 is the Cartesian product of two intervals: $B = [x_1, x_2] \times [y_1, y_2]$. The vertices of this rectangle are the points $(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)$. A two place real function H is a function whose domain $domH$ is a subset of \bar{R}^2 and whose range $RanH$ is a subset of \mathbb{R}

Let S_1 and S_2 be nonempty subsets of \bar{R} , and let H be a two-place real function such that $DomH = S_1 \times S_2$. A two-place real function H is a function whose domain, $domH$ is a subset of \bar{R}^2 , and whose range $RanH$ is a subset of \mathbb{R} .

Let $B = [x_1, x_2] \times [y_1, y_2]$ be a rectangle all of whose vertices are in $DomH$. Then the H-volume of B is given by: $V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$.

In terms of the first order differences of H on the rectangle B , define: $\Delta_{x_1}^{y_2} H(x, y) = H(x_2, y) - H(x_1, y)$

$$\Delta_{y_1}^{x_2} H(x, y) = H(x, y_2) - H(x, y_1)$$

Then, the H volume of a rectangle B is the *second order difference* of H on B : $V_H(B) = \Delta_{y_1}^{x_2} \Delta_{x_1}^{y_2} H(x, y)$.

A two place real function H is 2-increasing if $V_H(B) \geq 0$ for all rectangles B whose vertices lie in $DomH$.

When H is 2-increasing, the H-volume is called an H-measure of B , or sometimes quasi-monotone.

We will need two results that are of importance in understanding this theory.

Let S_1 and S_2 be non-empty subsets of \bar{R} and let H be a 2-increasing function with domain $S_1 \times S_2$. Let x_1, x_2 be in S_1 with $x_1 \leq x_2$ and let y_1, y_2 be in S_2 with $y_1 \leq y_2$. Then the function $t \mapsto H(t, y_2) - H(t, y_1)$ is non-decreasing on S_1 and the function $t \mapsto H(x_2, t) - H(x_1, t)$ is non-decreasing on S_2 .

Let S_1 and S_2 be non-empty subsets of \bar{R} , and let H be a grounded 2-increasing function with domain $S_1 \times S_2$. Then H is non-decreasing in each argument.

Here a function H from $S_1 \times S_2$ into \mathbb{R} is *grounded* if $H(x, a_2) = 0 = H(a_1, y)$ for all (x, y) in $S_1 \times S_2$.

Suppose S_1 has greatest element b_1 and S_2 has greatest element b_2 . A function H from $S_1 \times S_2$ into \mathbb{R} has margins, and the margins of H are functions F and G given by:

$$\begin{aligned} \text{Dom}F &= S_1, F(x) = H(x, b_2) \forall x \in S_1 \\ \text{Dom}G &= S_2, G(y) = H(b_1, y) \forall y \in S_2 \end{aligned}$$

Suppose the H function is grounded 2-increasing with margins whose domain is $S_1 \times S_2$. Let (x_1, y_1) and (x_2, y_2) be any points $S_1 \times S_2$ then:

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

Now we can define copulas. The standard approach in this direction is to define subcopulas as a class of grounded 2-increasing functions with margins. Then copulas are defined as subcopulas with domain \mathbb{I}^2 . Here $\mathbb{I}^2 = \mathbb{I} \times \mathbb{I}$, where $\mathbb{I} = [0, 1]$.

A two-dimensional subcopula (2-subcopula) is a function C' with the following properties:

- $\text{Dom } C' = S_1 \times S_2$, where S_1, S_2 are subsets of \mathbb{I} containing 0 and 1
- C' is grounded and 2-increasing
- For every u in S_1 and every v in S_2 , $C'(u, 1) = u, C'(1, v) = v$

So for every $(u, v) \in \text{Dom}C'$, $0 \leq C'(u, v) \leq 1$. This means that $\text{Ran}C'$ is a subset of \mathbb{I} .

A two-dimensional copula is a 2-subcopula C whose domain is \mathbb{I}^2 . It is a function C from \mathbb{I}^2 to \mathbb{I} with the following properties:

- For every $u, v \in \mathbb{I}$, $C(u, 0) = 0 = C(0, v)$ and $C(u, 1) = u, C(1, v) = v$
- For every $u_1, u_2, v_1, v_2 \in \mathbb{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$ then $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$

A subcopula and a copula are different, and these differences matter. Let's look at some results in this area. For example, suppose C' is a subcopula. Then for every (u, v) in $\text{Dom}C'$,

$$\max(u + v - 1, 0) \leq C'(u, v) \leq \min(u, v) \quad (20)$$

Let (u, v) be an arbitrary point in $\text{Dom}C'$. Now $C'(u, v) \leq C'(u, 1) = u$ and $C'(u, v) \leq C'(1, v) = v$. Together these yield $C'(u, v) \leq \min(u, v)$. Now since $V_{C'}([u, 1] \times [v, 1]) \geq 0$. These imply $C'(u, v) \geq u + v - 1$. In addition, $C'(u, v) \geq 0$. Together this means $C'(u, v) \geq \max(u + v - 1, 0)$.

Every copula is also a subcopula, but the reverse is not true.

We state and look at a proof of one of the most important theorems in this theory, called Sklar's theorem.

2.7 Sklar's Theorem

Here we are looking at m dimensional copulas. This is a function C from the unit m -cube $[0, 1]^m$ to the unit interval $[0, 1]$ which satisfies the following:

- $C[1, \dots, 1, a_n, 1, \dots, 1] = a_n$ for every $n \leq m$ and for all $a_n \in [0, 1]$

- $C[a_1, \dots, a_m] = 0$ for $a_n = 0$ for any $n \leq m$
- C is m -increasing

Let's understand what these are saying. The first property says that once the realizations of $m - 1$ variables are known with marginal probabilities 1, then the joint probabilities of the m outcomes is the same as the probability of the remaining uncertain outcomes. The second property is saying that the joint probability of all outcomes is 0, if the marginal probability of any outcome is 0. The third property states that the C volume of any m -dimensional interval is non-negative.

We will use the proof given in⁶

First let $[a, b]$ and $[c, d]$ be non-empty intervals of \mathbb{R} and suppose $G: [a, b] \rightarrow [c, d]$ be a non-decreasing mapping with $c = \inf_{x \in [a, b]} G(x)$, and $d = \sup_{x \in [a, b]} G(x)$.

Thus, as G is a mapping, this gives $a = \inf\{x \in \mathbb{R}, G(x) > c\}, b = \sup\{x \in \mathbb{R}, G(x) > d\}$. By lep and uep are meant the lower end point and upper end point and in this case, $a = lep(G)$, $b = uep(G)$. The inverse function $\forall u \in [lep(G), uep(G)], G^{-1}(u) = \inf\{x \in \mathbb{R}, G(x) \geq u\}$. Following⁶, we observe the following result:

Let G be a non-decreasing right continuous function. Then G^{-1} is left continuous and we have: $\forall u \in [c, d], G(G^{-1}(u)) \geq u$ and $\forall x \in [a, b], G^{-1}(G(x)) \leq x$, in addition we also have: $\forall x \in [lep(G), uep(G)], G^{-1}(G(x) + 0) = x$

We also recall the definition of a distribution function on $\mathbb{R}^d, d \geq 1$. A mapping $\mathbb{R}^d \rightarrow \mathbb{R}$ is a distribution function iff :

- F is right continuous
- F assigns to non-negative volumes to any cuboid $[a, b]$ with $a = (a_1, a_2, \dots, a_d) \leq b = (b_1, b_2, \dots, b_d)$ which is equivalent to $a_i \leq b_i$ for all $1 \leq i \leq d$ so that $\Delta F(a, b) = \sum_{\epsilon \in \{0, 1\}^d} (-1)^{s(\epsilon)} F(b + \epsilon * (a - b)) \geq 0$, where $(x, y) * (X, Y) = (x_1 X_1, x_2 X_2, \dots, x_k Y_k)$ and $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_d)$ runs over $\{0, 1\}^d$ and $s(\epsilon) = \epsilon_1 + \dots + \epsilon_d$

In addition, in order to become a cumulative distribution function, we also need:

- $\lim_{\exists i, 1 \leq i \leq k, t_i \rightarrow -\infty} F(t_1, \dots, t_k) = 0$
- $\lim_{\forall i, 1 \leq i \leq k, t_i \rightarrow \infty} F(t_1, \dots, t_k) = 1$

In⁶, a copula on \mathbb{R}^d is a cdf C , with marginal cdf's defined in the following manner, for $1 \leq i \leq d$:

$$C_i(s \in \mathbb{R}) = C(+\infty, \dots, \underbrace{\text{inf ty}}_{i\text{-th argument}}, +\infty, \dots, +\infty) \quad (21)$$

where these are all equal to the $(0, 1)$ uniform cdf, which is defined as

$$x \rightarrow x1_{[0,1]} + 1_{[1,+\infty]} \quad (22)$$

So for $s \in [0, 1]$

$$C_i(s) = C(1, \dots, 1, \underbrace{s}_{i \text{ th argument}}, 1, \dots, 1) = s \quad (23)$$

Now Sklar's theorem reads: For any cdf F on $\mathbb{R}^d, d \geq 1$, there exists a copula C on \mathbb{R}^d such that: $\forall x \in \mathbb{R}^d, F(x) = C(F_1(x), \dots, F_d(x))$

For $s = (s_1, s_2, \dots, s_d) \in [0, 1]^d$, set $C(s) = F(F_1^{-1}(s_1 + 0), F_2^{-1}(s_2 + 0), \dots, F_d^{-1}(s_d + 0))$ C will assign non-negative volumes to cuboids of $[0, 1]^d$ using the definition of $\Delta F(a, b)$ above, with arguments of the form $F_i^{-1}(\circ + 0), 1 \leq i \leq d$.

C is right continuous since F is right continuous, since of the $F_i^{-1}(\circ + 0), 1 \leq i \leq d$.

Using the result earlier for $G^{-1}(G(x) + 0) = x$ with the above result, the theorem of Sklar follows

Following⁶, let's see why the result $G^{-1}(G(x) + 0) = x$ holds. For $G^{-1}(G(x) + h)$ we consider the limit of $h \searrow 0$. For any $h > 0, G^{-1}(G(x) + h)$ is the infimum of the set $y \in [a, b]$ such that $G(y) \geq G(x) + h$. All such y satisfies $y \geq x$, so that $G^{-1}(G(x) + 0) \geq x$.

Next we show that $G^{-1}(G(x) + 0) \leq x$. First, $G(x + h) \searrow G(x)$ as G is right continuous and the right hand limit exists of the non-decreasing function $G^{-1}(\circ)$, so that $G^{-1}(G(x + h)) \searrow G^{-1}(G(x + 0))$.

Now $G^{-1}(G(x + h)) \leq x + h$ since $G^{-1}G(x) \leq x$, this shows that $G^{-1}(G(x) + 0) \leq x$ as $h \searrow 0$.

2.8 Properties of the Copula function

We will illustrate important properties as given in the most important paper on this topic, the paper by⁷. Suppose we have m uniform random variables, U_1, U_2, \dots, U_m . The joint distribution function C is defined as:

$$C(u_1, u_2, \dots, u_m, \rho) = Pr[U_1 \leq u_1, \dots, U_m \leq u_m] \quad (24)$$

Given the univariate marginal distribution functions $F_1(x_1), F_2(x_2), \dots, F_m(x_m)$, the function $C(F_1(x_1), F_2(x_2), \dots, F_m(x_m)) = F(x_1, x_2, \dots, x_m)$, which is a multivariate distribution function with univariate marginal distributions given by $F_1(x_1), F_2(x_2), \dots, F_m(x_m)$. To see this:

Note that the C function can be written thus:

$$\begin{aligned} C &= Pr[F_1^{-1}(U_1) \leq x_1, F_2^{-1}(U_2) \leq x_2, \dots, F_m^{-1}(U_m) \leq x_m] \\ &= Pr[X_1 \leq x_1, X_2 \leq x_2, \dots, X_m \leq x_m] \\ &= F(x_1, x_2, \dots, x_m) \quad (25) \end{aligned}$$

In addition we can show that the marginal distribution of X_i is $F_i(x_i)$ as follows:

$$\begin{aligned} C(F_1(+\infty), F_2(+\infty), \dots, F_i(x_i), \dots, F_m(+\infty), \rho) \\ = Pr[X_1 \leq +\infty, X_2 \leq +\infty, \dots, X_i \leq x_i, \dots, X_m \leq +\infty] \\ = Pr[X_i \leq x_i] \quad (26) \end{aligned}$$

Sklar's theorem given earlier shows that converse, namely that if $F(x_1, x_2, \dots, x_m)$ is a multivariate joint distribution with univariate distribution functions $F_1(x_1), \dots, F_i(x_i), \dots, F_m(x_m)$ then there exists a copula function $C(u_1, u_2, \dots, u_m)$ such that

$F(x_1, x_2, \dots, x_m) = C(F_1(x_1), F_2(x_2), \dots, F_m(x_m))$. If each F_i is continuous then C is unique.

For the purpose of this paper, we will look only at properties of the Bivariate Copula Function $C(u, v, \rho)$ for uniform random variables U, V defined over the area $\{(u, v) | 0 < u \leq 1, 0 < v \leq 1\}$, with ρ a correlation parameter that is not necessarily equal to the Pearson's correlation coefficient.

There are three main properties:

- As U and V are positive, $C(0, v, \rho) = C(u, 0, \rho) = 0$
- Since U and V are bounded above by 1, the marginal distributions are $C(1, v, \rho) = v, C(u, 1, \rho) = u$
- For independent random variables U and $V, C(u, v, \rho) = uv$

2.8.1 Examples of Copula Functions Following⁷, the following copula functions are of interest

- The Frank Copula function is defined as: $C(u, v) = \frac{1}{\alpha} \ln[1 + \frac{(e^{\alpha u} - 1)(e^{\alpha v} - 1)}{e^{\alpha} - 1}]$, $-\infty < \alpha < \infty$
- Bivariate Normal: $C(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v), \rho)$, $-1 \leq \rho \leq 1$, where Φ_2 is the bivariate normal distribution function with correlation coefficient ρ and Φ^{-1} is the inverse of a univariate normal distribution function

It is also possible, following⁷ that given two uniform random variables, u and v , that are independent, to have a Copula function $C(uv) = \min(u, v)$. In this regard, we have the Frchet-Hoeffding Boundary Copulas theorem. We start by looking at the simplest cases. Firstly, an independent copula structure is given by $C(F(x), F(y)) = F(x)F(y)$. A minimum Copula is given by $C(F(x), F(y)) = \min(F(x), F(y))$, and a maximum Copula is given by $C(F(x), F(y)) = \max(F(x), F(y))$.

2.9 Understanding Frchet-Hoeffding Bounds

Here we follow the work of⁸. We start with the following result.

Suppose we have random variables X_1, X_2, \dots, X_d whose dependence structure is given by a Copula C . Let $T_i: \mathbb{R} \rightarrow \mathbb{R}$,

$i = 1, \dots, d$ be strictly increasing functions. Then the dependence structure of the random variables $T_1(X_1), \dots, T_d(X_d)$ is also given by the copula C

In addition, there are the Frechet - Hoeffding bounds. This essentially puts a pyramid inside which every copula has to lie. Such a pyramid gives a lower bound $C(u, v) = \max u + v - 1, 0$ and an upper bound $C(u, v) = \min(u, v)$.

This implies that such functions don't change the dependence structure. It was proven independently by Hoeffding and Frechet that a copula will always lie in between certain bounds. For instance consider two uniform random variables U_1 and U_2 . If $U_1 = U_2$, these are extremely dependent on each other. In this case the copula is given by:

$$C(u_1, u_2) = P(U_1 \leq u_1, U_1 \leq u_2) = \min(u_1, u_2) \quad (27)$$

Such a copula is always attained if $X_2 = T(X_1)$, where T is a monotonic transformation. Random variables of this kind are called comonotonic. There is also the idea of a countermonotonic random variable. For this, we need $1 - u_2 < u_1$. We then have:

$$\begin{aligned} C(u_1, u_2) &= P(U_1 \leq u_1, 1 - U_1 \leq u_2) \\ &= P(U_1 \leq u_1, 1 - u_1 \leq u_2) \\ &= u_1 + u_2 - 1 \quad (28) \end{aligned}$$

In other cases, this is 0. This brings us to the theorem on Frechet- Hoeffding bounds

Consider a copula $C(u) = C(u_1, \dots, u_d)$. Then

$$\max \sum_{i=1}^d u_i + 1 - d, 0 \leq C(u) \leq \min(u_1, u_2, \dots, u_d) \quad (29)$$

We start with the observation that a Copula function $C : [0, 1]^2 \rightarrow [0, 1]$ satisfies three conditions

- $C(u, 0) = C(0, v) = 0, \forall u, v \in [0, 1]$
- $C(u, 1) = u, C(1, v) = v, \forall u, v \in [0, 1]$
- For all $u_1 < u_2$ and $v_1 < v_2, \in [0, 1]$, the following is true: $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \leq 0$

Now using the second property from above, we see that

$$C(u, v) \leq C(u, 1) \leq u \quad (30)$$

and

$$C(u, v) \leq C(1, v) \leq v \quad (31)$$

Together these give the upper bound. Now we take $u_1 = u, v_1 = v, v_2 = 1$ in the third property gives:

$$C(u, v) - u - v + 1 \geq 0 \quad (32)$$

As $C(u, v) \geq 0$, this gives the lower bound of the Frechet-Hoeffding theorem.

2.10 Construction of Credit Curves

2.10.1 Kaplan Meier Estimator Following⁹, let T be the random variable that describes an individual's survival time, and let $t_{(f)}$ a time for an event drawn from T . Here f denotes the ascending order of event times as in for example $t_{(1)} \leq t_{(2)}$. With this terminology, the Kaplan Meier survival estimates, $\hat{S}_{t_{(f)}}$ is given by the following formula:

$$\begin{aligned} \hat{S}_{t_{(f)}} &= \hat{S}_{t_{(f-1)}} \hat{P}(T > t_{(f)} | T \geq t_{(f)}) \\ &= \prod_{i=1}^f \hat{P}(T > t_{(i)} | T \geq t_{(i)}) \quad (33) \end{aligned}$$

Here \hat{P} is the estimated conditional probability of surviving past time $t_{(i)}$, given survival to at least time $t_{(i)}$. The \hat{S} is the Kaplan-Meier survival estimate of the previous time step. In this regard, one could also write:

$$\prod_{i=1}^f \hat{P}(T > t_{(i)} | T \geq t_{(i)}) = \prod_{i=1}^f \left(1 - \frac{D_{(i)}}{n_{(i)}}\right) \quad (34)$$

Here $D_{(i)}$ is the number of events occurring at $t_{(i)}$ and $n_{(i)}$ is the number of individuals that have survived.

2.10.2 Cox Proportional Hazards Model Here we use two components

- Baseline hazard rate as a function of time
- Effect parameters

Let $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ be a set of p explanatory variables, for subject i . The Cox model is then written in terms of the hazard function $h(t, x_i)$ and is defined as:

$$h(t, x_i) = h_0(t) \exp(\beta_1 x_{i1} + \dots + \beta_p x_{ip}) \quad (35)$$

Here β_i is the coefficient related to explanatory variable x_i , and $h_0(t)$ is the baseline hazard function. This model is also called a proportional hazard model. Suppose j and j' be two observations with the corresponding predictors θ_j and $\theta_{j'}$. Then the hazard ratio for these observations is given by:

$$\frac{h_j(t)}{h_j'(t)} = \frac{h_0(t)\exp(\theta_j)}{h_0(t)\exp(\theta)} = \frac{\exp(\theta_j)}{\exp(\theta)} \quad (36)$$

This is independent of time t .

3 Markov Chain Model

A stochastic process, with discrete time parameter, is a sequence of random variables X_1, X_2, \dots . The state of the process is given by X_1 . Here X_n is the state of the process at time n .

A Markov Chain is a stochastic process where the next state of the process only depends on the current state, and is not influenced by any of the previous states. This is called the Markovian property. The stochastic process $\{X_n, n = 1, 2, 3, \dots\}$ with state space I is said to be a discrete time Markov Chain if for each $n = 1, 2, \dots$, the following is true:

$$P(X_{n+1} = i_{n+1} | X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n), i_1, i_2, \dots, i_{n+1} \in I \quad (37)$$

A Markov Chain is called time homogeneous if given states $i, j \in I$:

$$P(X_{n+1} = j | X_n = i) = P(X_n = j | X_{n-1} = i) = p_{ij} \forall n \quad (38)$$

This is independent of n which represents the time.

The p_{ij} are called transition probabilities. These satisfy two conditions:

- $p_{ij} \geq 0$
- $\sum_{j \in I} p_{ij} = 1, i \in I$

A transition matrix gives a matrix of transition probabilities. This matrix is of the following form:

$$P = \begin{bmatrix} \dots & \dots & \dots \\ \vdots & \ddots & \vdots \\ \dots & \dots & \dots \end{bmatrix} \quad (39)$$

It must also be true that the following holds

$$\sum_{j=1}^k p_{ij} = 1, i = 1, 2, \dots, k \quad (40)$$

This is a way of saying that the sum of all transitional probabilities from one state i to all other states including itself is 1. To extend this from one step to m steps, we simply raise the

matrix elements to powers. In the following representations, $m = 2, 3, \dots$, where we are looking at the probability to migrate from state i to state j in m steps. This is given by:

$$P^m = \begin{bmatrix} \dots & \dots & \dots \\ \vdots & \ddots & \vdots \\ \dots & \dots & \dots \end{bmatrix} \quad (41)$$

Here p_{ij}^m is the probability of going from state i to state j in m steps.

Note that the transition matrix is estimated. One of the ways to achieve this is the cohort method. Define

$$\Delta t_k = t_k - t_{k-1} \quad (42)$$

Now the transition rate between two states i and j is estimated by:

$$\hat{p}_{ij}(\Delta t_k) = \frac{N_{ij}(\Delta t_k)}{N_i(t_k)} \quad (43)$$

Here $N_{ij}(\Delta t_k)$ denotes the number of entities that migrated from state i to state j during the period Δt_k and $N_i(t_k)$ is the number of entities that started in state i at time t_k . If there were no transitions between state i and j , then $\hat{p}_{ij} = 0$.

3.1 Term Structure of Default Rates

In⁷, three methods are shown

- Historical default information from rating agencies
- Merton option theoretical approach
- Implied approach using market price of default bonds or asset swap spreads

3.1.1 Understanding credit risks via CreditMetrics We will reference¹⁰. CreditMetrics is a framework for quantifying credit risk in portfolios. We are interested in the section on portfolio risk calculations. In most financial risk estimations, there are three main directions

- Estimating particular individual parameters such as expected default frequencies
- Estimating volatility of value which are the unexpected losses
- Estimating volatility of value within the context of a specific portfolio

Of these the most important is the idea of unexpected losses. As seen in¹⁰, there are several difficulties in terms of unexpected losses, and the industry has taken an approach that is dangerous. Firstly, as¹⁰ states, since it is difficult to explicitly

address correlations, a lot of the time it is assumed that the correlations are all zero, or all equal one which corresponds to the cases of perfectly correlated or perfectly positively correlated, but the issue is that these are not realistic. Other times, the practitioners assume that the correlations will be the same as that of some index portfolio. This needs a different type of analysis, because it assumes that a specific portfolio somehow mirrors the market in question, which may be the case but only under the assumption that there are parallels between the correlations and profile of composition.

3.1.2 Asset Value Model We are concerned with joint probabilities in terms of defaults. Following¹¹, denote by JDF to be the joint default frequency of firm 1 and firm 2, which is the actual probability of both firms defaulting together. Let ρ_D represent the default correlation for firms 1 and 2.

Then

$$\rho_D = \frac{JDF - EDF_1 EDF_2}{\sqrt{EDF_1(1 - EDF_1)EDF_2(1 - EDF_2)}} \quad (44)$$

Here EDF represents the probability of default.

Define the following variables. Set $X_i \equiv$ face value of security i . Then let $P_i \equiv$ price of security i , per \$1 of face value,

$V_p \equiv$ portfolio value $\equiv P_1X_1 + P_2X_2 + \dots + P_nX_n$,
 $w_i \equiv$ value proportion of security i in portfolio $\equiv \frac{P_iX_i}{V_p}$,
 $\rho_{ij} \equiv$ loss correlation between security i and j . In addition we have

$$w_1 + w_2 + \dots + w_n = 1 \quad (45)$$

Define $EL_i \equiv$ expected loss for security i , $EL_p \equiv$ portfolio expected value $= w_1EL_1 + \dots + w_nEL_n$. Now if $UL_i \equiv$ unexpected loss for security i . Then

$$UL_p = \sqrt{w_1w_2UL_1UL_1\rho_{11} + w_1w_2UL_1UL_2\rho_{12} + \dots + w_1w_nUL_1UL_n\rho_{1n} + \dots + w_nw_nUL_nUL_n\rho_{nn}} \quad (46)$$

The above equation gives the unexpected loss for the portfolio in question. Default correlation, which was the crux of the problems during the crisis of 2008 measures the strength of the default relationship between two borrowers. If there is no relationship, it means the default is zero. However, if borrowers are correlated, the probability of both defaulting is higher. According to¹¹, the joint probability of default is defined to be the likelihood that both firms market asset values will be below their respective default points in the same time period.

This probability depends on three factors

- Current asset values in terms of the market
- The asset volatilities
- The correlation between the market asset values

Denote by N_2 the bivariate normal distribution function, by N^{-1} the inverse normal distribution function, by ρ_a the correlation between firm 1's asset return and firm 2's asset return.

In this case, the JDF is given by

$$JDF = N_2(N^{-1}(EDF_1), N^{-1}(EDF_2), \rho_A) \quad (47)$$

Ideally, a firm's return can be written as the sum of the composite factor return and firm specific effects. The composite factor returns include country factor returns and industry factor returns. The country factor return has four components: the global economic effect, the regional factor effect, the sector factor effect and the country specific effect. Finally the industry factor return has four components, the global economic effect, the regional factor effect, the sector factor effect and finally the industry specific effect.

The composite (custom) market factor index for firm k can be written as:

$$\phi_k = \sum_{c=1}^{\bar{c}} w_{kc}r_c + \sum_{i=1}^{\bar{i}} w_{ki}r_i \quad (48)$$

In this equation set, $w_{kc} \equiv$ weight of firm k in country c , the $w_{ki} \equiv$ weight of firm k in country i , the r_c the return index for the country c , the r_i the return index for the country i .

Finally the ϕ_k the composite custom market index factor for firm k .

It is also true that

$$\sum_{c=1}^{\bar{c}} w_{kc} = \sum_{i=1}^{\bar{i}} w_{ki} = 1 \quad (49)$$

3.2 The KMV and Merton Models

The KMV corporation refers to Kealhofer, McQuown and Vasicek (KMV), which used to provide quantitative credit analysis, and was thereafter acquired by Moody's in 2002. We start with the example provided by¹¹. First consider a firm that has a single asset that has 1 million shares of Microsoft stock. Assume that it has a single fixed liability, which is a one year discount note with a par amount of 100 million dollars. The firm is otherwise funded by equity. In a year the company will either be able to pay off the note by virtue of the market value of its business, or it will default. The equity of this company is equivalent to 1 million call options on Microsoft stock, each with an exercise price of 100 dollars. The maturity time is 1 year. This entire example shows that the equity of a company can be thought of as a call option on the company's underlying assets. This means, according to¹¹, the value of equity will depend on three factors. These are the market value of the company's assets, the volatility and the payment terms of the liabilities. Merton's original model from 1974 has specific properties:

- The company has equity, a single debt liability and no other obligations

- The liability has continuous fixed coupon flow and infinite maturity
- The company has no other cash payouts like equity dividends

Merton showed that, assuming that the company's assets follow a lognormal process, this model can be solved to show a closed form process for the value of a company's debt. The aim of this model shows the company's debt.

The KMV model is based on probability of default of the company as a whole, rather than the valuation of the debt. The KMV model has the following properties, following¹¹.

- The company could have debt or non debt fixed liabilities, in addition to common equity and preferred stock
- Warrants, convertible debt and convertible preferred stock is allowed
- Short term obligations can be demanded by creditors and long term can be treated as perpetuities
- Any and all classes of liabilities are allowed to make fixed cash payouts
- The default occurs when the market value of a company's assets falls below a fixed point, called the default point. This default point depends on both the nature and the extent of the fixed obligations
- Default occurs on the company as a whole

In¹¹, the distance to default, $DD(h)$ is defined to be the number of standard deviations to the default point by horizon h .

This is calculated as

$$DD(h) = \frac{\ln(A) - \ln(DPT) + (\mu_A - \frac{1}{2}\sigma_A^2)h}{\sigma_A h^{1/2}} \quad (50)$$

Here A is the current market value of the company's assets, DPT is the company's default point, μ_A is the expected market return to the assets per unit of time, σ_A is the volatility of the market value of the company's assets per unit of time. The KMV model has a focus on default risk measurement and not debt valuation. The reason for this is that debt valuation actually has default risk measurement built into it. The other issue is that the using a lognormal model, there will be differences between actual realized default rates and predicted default rates. An example given by¹¹ is that of a firm that is more than 4 standard deviations from its default point has essentially 0 probability of default, although in reality the default probability is around 0.5percent, and this is actually significant in real life. This simply means that on paper, being 4 standard deviations away is equivalent to being better than AAA grade in investment, but

having a default probability of 0.5percent means it is not even investment grade.

There is a reason why the default risk measurements are used in place of debt valuations. The debt valuations already have the default risk measurements contained in them. In other words, if the default risk measurement is accurate, so is the debt valuation. Keep in mind that the distance to default is an ordinal measure, not an absolute measure. In this regard one needs, for instance, a log normal asset value distribution of the Merton approach. The solution to this problem is the KMV EDF (Expected Default Frequency) credit measure. The EDF is the probability of default within a given time period.

3.3 Prediction of Default Rates

When it comes to bond yields, one needs to consider specific ideas. These are:

1. *Spread Volatility* The average yield spread that corresponds to a given agency rating grade will change significantly with time
2. *Considerable variation in the shape of spread curves* This variation is significant when given as a function of term

3.4 Default Rates and Firm Values

Suppose one has a cash flow F , due at a single future date, t . Suppose r is the continuous discount cash rate to t for a default risk free cash flow. The option theoretic formula for the value of the cash flow today, V is given by

$$V = Fe^{-(r_t)}(1 - q_t(LGD)) \quad (51)$$

Here q_t is the so called risk neutral cumulative default probability to t , and LGD is the loss given default term, the expected percentage loss if the borrower defaults). There is a relationship between q_t the risk neutral cumulative default probability and p_t the actual cumulative default probability to t . Under the assumption of lognormality, this is given by:

$$q_t = N(N^{-1}(p_t) + (\frac{\mu_A - r}{\sigma_A}t^{1/2})) \quad (52)$$

Here N and N^{-1} represent the standard cumulative normal distribution and it's inverse function, μ_A represents the instantaneous expected return to the asset, and σ_A represents the volatility of asset returns. The cash flow is valued as if the default probability were q_t , and this is larger than the actual probability p_t . Another approximate relation between p_t and q_t is given by:

$$q_t \approx 2N[N^{-1}(\frac{p_t}{2}) + [\frac{\mu_A - r}{\sigma_A}t^{1/2}]] \quad (53)$$

Assuming a risk premium of $\mu - r$ is determined by the capital asset pricing model, we write

$$\mu - r = \left[\frac{\text{cov}(r_A, r_M)}{\text{var}(r_M)} \right] (\mu - r) \quad (54)$$

We can write this as

$$\frac{\mu - r}{\sigma_A} = \rho \lambda \quad (55)$$

Here r_A is the asset return, r_M is the market return, μ_M is the expected return, σ_A is the standard deviation of asset return, σ_M is the standard deviation of market return, ρ is the correlation of r_A and r_M , and λ is the market Sharpe ratio. We can write these equations together as

$$q_t = 2N \left[N^{-1} \left(\frac{P_t}{2} + \rho \lambda (t^{1/2}) \right) \right] \quad (56)$$

In the case of multiple cash flows, the valuation formula becomes

$$V = \sum C_t e^{-rt} (1 - q_t LCD) \quad (57)$$

Let us recapitulate the model and show an example. The EDF is a forward looking measure of the actual probability of default. The KMV model is based on the structural approach to calculate EDF, with the credit risk being driven by the firm value process. In order to get the actual probability of default, one goes through three steps:

- Estimation of the market value and volatility of the firm's assets
- Calculation of the distance to default, an index measure of default risk
- Scaling of the distance to actual probabilities of default using a default database

Essentially we are looking at two items: the estimation of firm value V , and the volatility of firm value σ_V . What usually happens is that the price of equity for most public firms is directly observable, and sometimes part of the debt is traded. Typically one has two equations:

$$\text{Equity Value } E = f(V, \sigma_V, K, c, r) \quad (58)$$

$$\text{volatility of equity } \sigma_E = g(V, \sigma_V, K, c, r) \quad (59)$$

Here K denotes the leverage ratio in capital structure, c is the average coupon paid on the long term debt, r is the risk free rate. One usually solves for V and σ_E from these two equations. As an example, suppose the current market value of assets, V_0 is 1000, the net expected growth of assets per annum, $\mu = 20\%$, the expected asset value in one year, $V_T = 1200$, the annualized asset volatility is $\sigma_V = 100$ and the default point is $d^* = 800$. Then the default distance is $d_f = \frac{1200 - 800}{100} = 4$.

Among the population of all firms with $d_f = 4$ at one point in time, supposing there were 5000 firms, of which 20 defaulted in a year. In this case,

$$EDF_1 = \frac{20}{5000} = 0.004 = 40 \text{ bp} \quad (60)$$

Acknowledgments

The author wishes to thank their mentor Rajit Chatterjea, from *The University of Southern California* for guidance on the sections related to probability theory.

References

- 1 C. Donnelly and P. Embrechts, *The Devil is in the Tails: Actuarial Mathematics and the Subprime Mortgage Crisis*, Visited on 06/10/2021).
- 2 N. Bowers, *Actuarial mathematics*.
- 3 D. Cox and D. Oakes, *Analysis of survival data*, <https://lib.ugent.be/en/catalog/rug01:000074230>, visited on 12/11/2019).
- 4 L. Longley-Cook, *Society of Actuaries - Actuarial Application of Monte Carlo Technique* by Russell M.
- 5 R. Nelsen and R. B, *An introduction to copulas*, Springer 1999.
- 6 S. Gane and Lo, *A SIMPLE PROOF OF THE THEOREM OF SKLAR AND ITS EXTENSION TO DISTRIBUTION FUNCTIONS*, URL:<https://arxiv.org/pdf/1803.00409.pdf> (visited on 04/21/2019).
- 7 D. Li, *On Default Correlation: A Copula Function Approach*.
- 8 T. Schmidt, *Coping with Copulas* 2006.
- 9 H. Englund and V. Mostberg, *Probability of Default Term Structure Modeling*, <https://www.diva-portal.org/smash/get/diva2:1667201/FULLTEXT03>, visited on 12/07/2023).
- 10 G. Gupton, *Credit Metrics Technical Document*.
- 11 S. Kealhofer, *Quantifying Credit Risk I: Default Prediction*, . 2501.(Visited on 10/26/2020).