

CreditCruncher - Technical Document
Version 0.9 [R348:355M]



Queen



Drone



Worker

Gerard Torrent Gironella

2 de enero de 2006

Copyright © 2004-2006 Gerard Torrent Gironella.

The image found in cover have been taken from Mark L. Winston. 1987. *The Biology of the Honey Bee* (ISBN: 0-671-07109-2). Harvard University Press. Cambridge, MA. These redrawn figures appear here without permission of Harvard University Press [Ref: 973029].

This file is part of the CreditCruncher software package. For license information, see the COPYING file in the top level directory of the CreditCruncher source distribution.

Índice general

1. Introducción	4
1.1. Acerca de CreditCruncher	4
1.2. Organización del contenido	5
1.3. Consideraciones	5
2. Formulación	6
2.1. Cartera de créditos	6
2.1.1. Ratings	7
2.1.2. Sectores	8
2.1.3. Activos	8
2.2. Tipos de interés	10
2.2.1. Función de transporte	11
2.2.2. Curva spot o cupón cero	12
2.3. Matriz de transición	13
2.3.1. Propiedades	13
2.3.2. Cambio de periodo	14
2.3.3. Función de supervivencia	15
2.4. Matriz de correlación	17
2.5. Valoración del riesgo	19
3. Resolución	22
3.1. Esquema general	22
3.2. El método de Monte Carlo	23
3.3. Variables aleatorias correlacionadas	24
3.4. Simulación del tiempo de fallido	24
3.4.1. Método Rating-Path	25
3.4.2. Método Time-To-Default	26
3.5. Evaluación de la cartera	27
3.6. Valoración del riesgo	28
4. Implementación	29
4.1. Paso 1. Interpretación del fichero de entrada	29
4.2. Paso 2. Particionamiento del tiempo	30

4.3.	Paso 3. Inicialización de los clientes	31
4.4.	Paso 4. Inicialización del método de simulación	31
4.4.1.	Método Rating-Path	31
4.4.2.	Método Time-To-Default	32
4.5.	Paso 5. Inicialización de las cópulas	32
4.5.1.	Método Rating-Path	33
4.5.2.	Método Time-To-Default	33
4.6.	Paso 6. Inicialización de los agregadores	33
4.6.1.	Maqueo del cashflow	35
4.6.2.	Maqueo del recovery	36
4.7.	Paso 7. Simulación	38
4.7.1.	Generación de números aleatorios	38
4.7.2.	Simulación del tiempo de fallido	38
4.7.3.	Valoración del segmento	39
4.8.	Paso 8. Valoración del riesgo	39
A.	Apéndices	41
A.1.	Conceptos básicos de estadística	41
A.2.	La variable aleatoria de Bernoulli	42
A.3.	La variable aleatoria Binomial	43
A.4.	La variable aleatoria Normal	43
A.5.	Estimadores estadísticos	44
A.6.	Error estándar de un quantil (o VaR)	45
A.7.	Error estándar del TCE	46
A.8.	Cálculo de la raíz de una matriz	46
A.9.	Algoritmo de la cópula gaussiana	47
A.10.	Descomposición de Cholesky de una matriz en bloques	49

Capítulo 1

Introducción

Este documento contiene la descripción del método de valoración del riesgo de crédito implementado por el proyecto CreditCruncher. Para su lectura no se presuponen conocimientos avanzados de matemáticas o finanzas. En caso de encontrar un error, sugerir mejoras o no entender algún punto, no dude en ponerse en contacto con el equipo de desarrollo de CreditCruncher¹ que tendrá en cuenta sus aportaciones para futuras versiones de este documento.

1.1. Acerca de CreditCruncher

La valoración del riesgo de crédito no es un tema cerrado, muestra de ello es la multitud de métodos que existen para su valoración. Se recomienda la lectura del artículo *Different strokes* [9] donde se exponen los principales modelos de valoración del riesgo de crédito y sus características.

CreditCruncher valora el riesgo de impago de una cartera de créditos usando la técnica de simulación Monte Carlo. Pretende ofrecer un método de valoración del riesgo de crédito totalmente documentado y soportado por una implementación libre y gratuita. Pertenece a la familia de métodos tipo CreditMetrics².

La mayoría de conceptos y explicaciones que pueden encontrarse en este documento han sido extraídas o inspiradas en el documento *CreditMetrics - Technical Document* [3]. Puede usarse el artículo *Probability models of credit risk* [2] como una introducción corta y clara.

¹<http://www.generacio.com/ccruncher/>

²<http://www.riskmetrics.com/>

1.2. Organización del contenido

Se ha organizado el contenido en tres secciones principales y un bloque de apéndices.

Formulación. Contiene la descripción del problema que se pretende resolver y se introducen los elementos y propiedades considerados claves para la posterior resolución. La lectura de este apartado es necesaria para entender los elementos del fichero de entrada de datos del programa.

Resolución. Se exponen los elementos usados para resolver el problema y se detalla la estructura del método de resolución. La lectura de este apartado es necesaria para la interpretación de los resultados proporcionados por el programa.

Implementación. Se explican los detalles de la implementación. La lectura de este apartado es necesaria para entender alguno de los apartados del fichero de entrada de datos del programa así como para la interpretación de los resultados proporcionados por este.

Apéndices. Contienen elementos necesarios para la comprensión del contenido de las secciones principales, pero que su inclusión en estas oscurecería la explicación.

1.3. Consideraciones

Recomendamos la lectura de las referencias bibliográficas que se incluyen, pueden ayudarle en la comprensión de lo expuesto en este documento.

Los textos contenidos en los gráficos están en inglés debido a que son compartidos por todas las posibles traducciones de este documento.

Si utiliza un visualizador tipo *Adobe* o *gsview* puede que algunos gráficos se muestren en un trazo inadecuado. Imprima el documento para obtener una presentación óptima.

Capítulo 2

Formulación

Dada una cartera de créditos a empresas de tamaño mediano, deseamos valorar el riesgo debido a los impagos al cabo de un tiempo T .

A continuación se introduce los elementos y propiedades básicas que constituyen el marco de trabajo.

2.1. Cartera de créditos

La estructura de la cartera de créditos consiste en un conjunto de clientes agrupados por sectores de actividad. Cada cliente tiene contratado un conjunto de productos de crédito. Cada contrato puede estar cubierto por un número variable de garantías o acuerdos. Puede verse un esquema de la estructura en la figura 2.1.

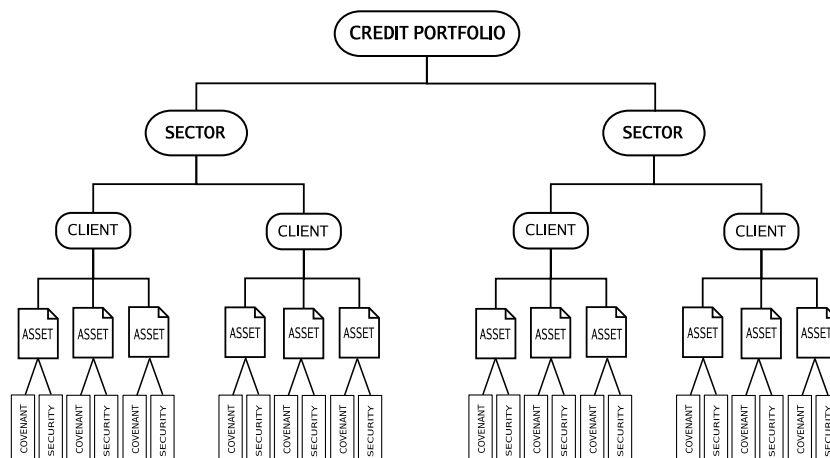


Figura 2.1: Estructura de la cartera de créditos

2.1.1. Ratings

Un *sistema de ratings* es una medida de calidad crediticia usada para valorar creditores. A cada creditor se le asigna una nota discreta (pe. AAA, AA, A, BBB, BB, B, CCC, Default) en función de su calidad crediticia. Los únicos ratings contemplados en este documento son los que tienen una relación estadística directa y cuantificable con la probabilidad de impago del creditor. Ejemplos de este tipo de ratings son los publicados por Moody's Investor Service¹ o Standard & Poors².

La metodología para la generación de un sistema de ratings queda fuera del ámbito de este documento. CreditCruncher presupone que cada empresa de la cartera tiene un rating inicial asignado.

El rating de cada empresa puede variar a lo largo del tiempo (véase figura 2.2). La evolución temporal del rating de una empresa se contempla a través de la matriz de transición o la función de supervivencia (véase la sección 2.3).

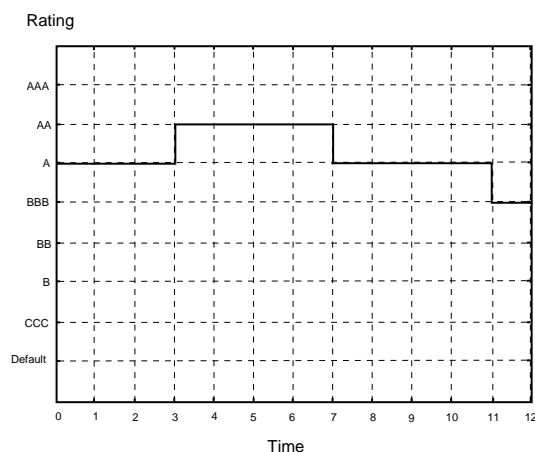


Figura 2.2: Evolución del rating a lo largo del tiempo

Notación. $rating(i, t)$ = rating del cliente i en el tiempo t .

Notación. $P(r_i \rightarrow r_j; t_0; t_1)$ = probabilidad de pasar de un rating inicial r_i en tiempo t_0 a un rating r_j en tiempo t_1 .

¹<http://www.moodys.com>

²<http://www.standardandpoors.com>

2.1.2. Sectores

La correlación de fallidos entre clientes es uno de los conceptos que añaden complejidad de la valoración del riesgo de crédito. No es lo mismo tener una cartera de créditos donde los clientes hacen fallido de forma independiente que una cartera donde los fallidos se encuentran correlacionados. En el primer caso, al cabo de un año tendremos un conjunto limitado de fallidos. En el segundo caso, al cabo de un año la mayoría de clientes habrán hecho fallido o casi ningún cliente habrá hecho fallido.

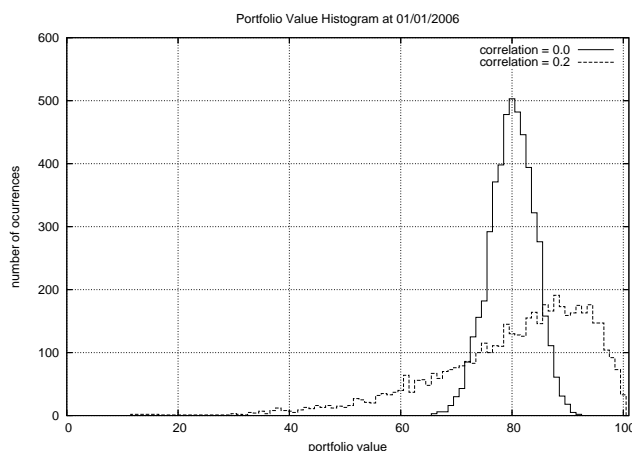


Figura 2.3: Impacto de la correlación intrasectorial

Al no poder asignar una correlación de fallido cliente a cliente, se recurre a la agrupación de estos en *sectores*. Se considera que la cartera de créditos dispone de un conjunto de sectores donde los componentes de cada sector muestran una evolución crediticia similar. O sea, que la mejora o empeoramiento de la calidad crediticia (rating) afecta de forma común a los componentes del sector. En general se identifican estos sectores con los sectores industriales.

Se considera que cada cliente pertenece a un único sector y permanece en el a lo largo del tiempo. La relación entre sectores se contempla a través de la tabla de correlaciones sectoriales (véase la sección 2.4).

Notación. $sector(i)$ = sector al que pertenece el cliente i .

2.1.3. Activos

Cada cliente tiene contratado un conjunto de activos con riesgo de crédito. Caracterizamos un activo por los siguientes elementos (importes positivos

significan que el cliente paga, importes negativos significan que el cliente cobra):

Cashflow. Entregas y devoluciones de dinero a lo largo del tiempo. Incluye las posibles amortizaciones, primas, cupones, comisiones, costes, etc. Usaremos el cashflow para calcular el valor, o precio, de un activo en el instante t .

Recovery. Importe correspondiente a la liquidación de deudas mutuas en caso de fallido. Incluye la posible recuperación, pago de obligaciones contraídas (pe. en el caso de avales), etc. Puede equipararse a $EAD \times (1 - LGD)$ donde EAD es Exposure At Default (importe) y LGD es Loss Given Default (porcentaje).

Ejemplo. Caracterizamos un bono (bond) de 100 € de valor nominal, con fecha emisión 31/12/2006, tipo de interés anual del 4 %, pago anual de cupones y amortización al cabo de 5 años. En caso de fallido se estima que se recupera un 80 % del importe pendiente de abonar.

Date	Cashflow	Recovery
31/12/2006	-100.00	0.00
31/12/2007	4.00	96.00
31/12/2008	4.00	92.00
31/12/2009	4.00	89.60
31/12/2010	4.00	86.40
31/12/2011	104.00	83.20

Ejemplo. Caracterizamos un préstamo hipotecario (mortgage) de 100 €, con fecha de contratación 15/01/2006, tipo de interés anual del 6,5 % y cuota (instalment) mensual y un plazo de 25 años. La vivienda hipotecada está valorada en 80 €. Determinamos la cuota mensual usando la siguiente fórmula (canon vencido o método francés):

$$I = \frac{A \cdot r \cdot (1 + r)^t}{(1 + r)^t - 1} = \frac{100 \cdot (6,5\%/12) \cdot (1 + 6,5\%/12)^{25 \cdot 12}}{(1 + 6,5\%/12)^{25 \cdot 12} - 1} = 0,67521$$

Date	Cashflow	Recovery
15/01/2006	-100.00	0.00
15/02/2006	0.68	80.00
15/03/2006	0.68	80.00
15/04/2006	0.68	80.00
...
15/11/2030	0.68	80.00
15/12/2030	0.68	80.00
15/01/2031	0.68	80.00

Ejemplo. Caracterizamos un aval (endorsement) por un importe avalado de 100 € con fecha de contratación 15/01/2006, durante un periodo de 2 años y una cuota semestral anticipada de 4,5 €.

Date	Cashflow	Recovery
15/01/2006	4.50	0.00
15/07/2006	4.50	-100.00
15/01/2007	4.50	-100.00
15/07/2007	4.50	-100.00
15/01/2008	0.00	-100.00

2.2. Tipos de interés

Definición. Sea C_{t_0} el importe inicial de una operación y C_{t_1} el importe final. Definimos el *tipo de interés efectivo*, r , como:

$$C_{t_1} = C_{t_0} \cdot (1 + r) \quad (2.1)$$

$$r = \frac{C_{t_1} - C_{t_0}}{C_{t_0}} \quad (2.2)$$

Definición. El *tipo de interés simple*, r_s , es un tipo de interés donde para cada periodo de tiempo se incrementa el importe inicial, C_{t_0} por un factor de r_s .

$$C_{t_1} = C_{t_0} \cdot (1 + r_s \cdot (t_1 - t_0)) \quad (2.3)$$

$$r_s = \frac{r}{t_1 - t_0} \quad (2.4)$$

Definición. El *tipo de interés compuesto*, r_c , es un tipo de interés donde en cada periodo de tiempo se incrementa por un factor de r_c el importe acumulado del periodo anterior.

$$C_{t_1} = C_{t_0} \cdot (1 + r_c)^{(t_1 - t_0)} \quad (2.5)$$

$$r_c = (1 + r)^{\frac{1}{t_1 - t_0}} - 1 \quad (2.6)$$

Definición. El *tipo de interés continuo*, r_e , es el caso límite del interés compuesto.

$$C_{t_1} = C_{t_0} \cdot e^{r_e \cdot (t_1 - t_0)} \quad (2.7)$$

$$r_e = \ln(1 + r_c) \quad (2.8)$$

La fórmula exponencial, no el coeficiente, se obtiene considerando el límite del tipo de interés compuesto.

$$\begin{aligned} \lim_{t_1 \rightarrow t_0} 1 + r_c &= \lim_{t_1 \rightarrow t_0} (1 + r)^{\frac{1}{t_1 - t_0}} = \\ \lim_{t_1 \rightarrow t_0} (1 + r_s \cdot (t_1 - t_0))^{\frac{1}{t_1 - t_0}} &= \lim_{n \rightarrow \infty} (1 + \frac{r_s}{n})^n = e^{r_s} \end{aligned} \quad (2.9)$$

Ejemplo. Consideremos una operación que supone una inversión inicial de 100 MM. y que al cabo de 5 años proporciona unos ingresos de 120 MM.

Calculemos los diferentes tipos de interés:

$$\begin{aligned} r &= \frac{120 - 100}{100} = 20 \% \\ r_s &= \frac{r}{5} = \frac{20 \%}{5} = 4 \% \\ r_c &= (1 + r)^{1/5} - 1 = 1,2^{0,2} - 1 = 1,0371 - 1 = 3,71 \% \\ r_e &= \ln(1 + r_c) = \ln(1,0371) = 3,65 \% \end{aligned}$$

Recuperemos el importe final de la operación a partir del importe inicial, el intervalo de tiempo y el tipo de interés:

$$\begin{aligned} \text{tipo efectivo} &\rightarrow C_{t_1} = 100 \cdot (1 + 20 \%) = 120 \\ \text{interés simple} &\rightarrow C_{t_1} = 100 \cdot (1 + 4 \% \cdot 5) = 120 \\ \text{interés compuesto} &\rightarrow C_{t_1} = 100 \cdot (1 + 3,71 \%)^5 = 120 \\ \text{interés continuo} &\rightarrow C_{t_1} = 100 \cdot e^{3,65 \% \cdot 5} = 120 \end{aligned}$$

2.2.1. Función de transporte

Definición. Fijado un tipo de interés, r , y un intervalo de tiempo, $\Delta t = t_k - t_0$, la *función de transporte*, Υ , proporciona el factor que debe aplicarse a un importe en t_0 para obtener el importe equivalente en t_k .

Caso $C_0 \longrightarrow C_k \quad t_0 < t_k$:

$$\begin{aligned} \text{interés simple} &\rightarrow C_0 \cdot \Upsilon(t_0, t_k, r) = C_0 \cdot (1 + r \cdot (t_k - t_0)) = C_k \\ \text{interés compuesto} &\rightarrow C_0 \cdot \Upsilon(t_0, t_k, r) = C_0 \cdot (1 + r)^{(t_k - t_0)} = C_k \\ \text{interés continuo} &\rightarrow C_0 \cdot \Upsilon(t_0, t_k, r) = C_0 \cdot e^{r \cdot (t_k - t_0)} = C_k \end{aligned}$$

Caso $C_k \leftarrow C_0$ $t_k < t_0$:

$$\begin{aligned} \text{interés simple} &\rightarrow C_0 \cdot \Upsilon(t_0, t_k, r) = C_0 \cdot (1 + r \cdot (t_0 - t_k))^{-1} = C_k \\ \text{interés compuesto} &\rightarrow C_0 \cdot \Upsilon(t_0, t_k, r) = C_0 \cdot (1 + r)^{-(t_0 - t_k)} = C_k \\ \text{interés continuo} &\rightarrow C_0 \cdot \Upsilon(t_0, t_k, r) = C_0 \cdot e^{-r \cdot (t_0 - t_k)} = C_k \end{aligned}$$

Notación. En este documento se considera que el tipo de interés aplicado es el tipo de interés compuesto. En este caso, la función de transporte tiene una expresión única, sea cual sea el sentido en el que se aplica:

$$\Upsilon(t_0, t_k, r) = (1 + r)^{(t_k - t_0)} \quad (2.10)$$

2.2.2. Curva spot o cupón cero

Definición. La *curva spot* o *curva cupón cero* es la función S que indica el tipo de interés a aplicar en la función de transporte desde el tiempo t_0 . En el mercado existen productos simples a distintos plazos para los cuales se puede calcular el tipo de interés que proporcionan. Estos tipos de interés solamente se pueden usar en la función de transporte cuando uno de los tiempos sea t_0 y el otro sea superior a t_0 .

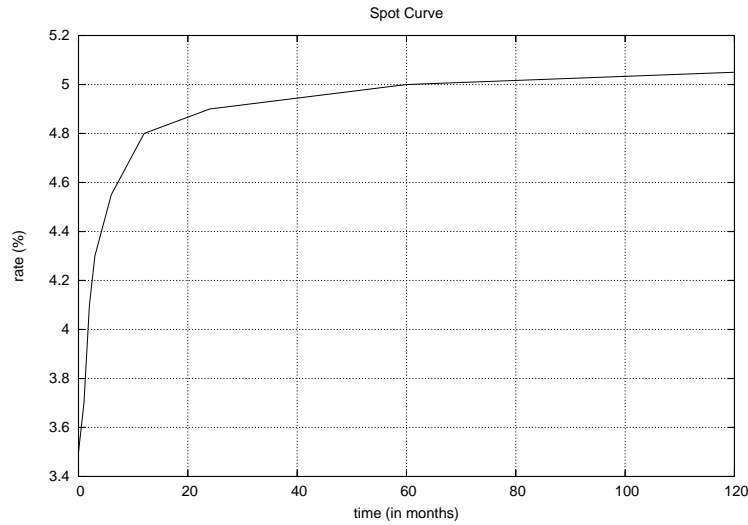


Figura 2.4: Curva Spot

Proposición. Dada una curva spot S_{t_0} en t_0 , podemos calcular el coeficiente de transporte entre t_i y t_j para todo $t_i, t_j \geq t_0$:

$$\Upsilon_S(t_i, t_j) = \Upsilon(t_i, t_0, S_{t_0}(t_i)) \cdot \Upsilon(t_0, t_j, S_{t_0}(t_j)) \quad (2.11)$$

2.3. Matriz de transición

Definición. La *matriz de transición* en el periodo T es una matriz cuadrada que proporciona la probabilidad que un cliente con rating inicial r_i pase a tener, al cabo de un tiempo T , rating r_j . La denotamos de la forma siguiente:

$$M_T = \begin{pmatrix} m_{1,1} & \dots & m_{1,n} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \dots & m_{n,n} \end{pmatrix} \quad m_{i,j} = P(r_i \rightarrow r_j; 0; T)$$

donde n es el número de ratings y $m_{i,j}$ corresponde a la probabilidad de que un cliente con rating r_i pase a tener, al cabo de T tiempo, rating r_j .

Ejemplo. Matriz de transición anual ($T = 1$ año). Las probabilidades están expresadas en tanto por ciento.

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	90,81	8,33	0,68	0,06	0,12	0,00	0,00	0,00
AA	0,70	90,65	7,79	0,64	0,06	0,14	0,02	0,00
A	0,09	2,27	91,05	5,52	0,74	0,26	0,01	0,06
BBB	0,02	0,33	5,95	86,93	5,30	1,17	0,12	0,18
BB	0,03	0,14	0,67	7,73	80,53	8,84	1,00	1,06
B	0,00	0,11	0,24	0,43	6,48	83,46	4,07	5,21
CCC	0,22	0,00	0,22	1,30	2,38	11,24	64,86	19,78
Default	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00

En particular, la probabilidad que un cliente con rating *AA* pase a tener rating *B* al cabo de 1 año es del 0,14%.

2.3.1. Propiedades

Propiedad 1. El valor de los elementos de la matriz de transición se encuentra entre 0 y 1 debido a que los elementos de la matriz son probabilidades.

$$0 \leq m_{i,j} \leq 1 \quad \forall i, j \quad (2.12)$$

Propiedad 2. La suma de los elementos de cualquier fila de la matriz de transición vale 1. De esta forma se está imponiendo que el conjunto de ratings finales solo puede ser el de los ratings contemplados en la matriz.

$$\sum_{j=1}^n m_{i,j} = 1 \quad \forall i \quad (2.13)$$

Propiedad 3. Los elementos de la fila correspondiente al rating *Default* (r_n), son todos 0, excepto el elemento de la columna que corresponde al rating *Default*, $m_{n,n}$, que vale 1. Esta condición indica que cuando se llega al estado de fallido no es posible salir de este estado.

$$\begin{aligned} m_{n,j} &= 0 & \forall j \neq n \\ m_{n,n} &= 1 \end{aligned} \quad (2.14)$$

Propiedad 4. Sea cual sea el rating inicial, existe la posibilidad que realice fallido.

$$\forall i \quad \exists j \quad \text{tq.} \quad m_{i,j} > 0 \quad \text{and} \quad m_{j,n} > 0 \quad (2.15)$$

Propiedad 5. La matriz de transición es diagonalizable. O sea, todos los valores propios son reales e independientes.

$$M_T \quad \text{is diagonalizable} \quad (2.16)$$

Esta propiedad es una condición técnica para garantizar la existencia de las raíces de la matriz de transición.

2.3.2. Cambio de periodo

Deseamos obtener la matriz de transición para periodos distintos (múltiplos o fraccionarios) del periodo proporcionado, T . Esto nos permitirá determinar la probabilidad que un cliente con rating inicial r_i tenga rating r_j al cabo de $k \cdot T$ tiempo o al cabo de T/k tiempo.

Ejemplo. Calculemos la probabilidad de pasar de rating *AA* a rating *B* en un plazo de dos años disponiendo de la matriz de transición anual.

$$\begin{aligned} P(AA \rightarrow B; 0; 2) = & P(AA \rightarrow AAA; 0; 1) \quad \cdot P(AAA \rightarrow B; 1; 2) \quad + \\ & P(AA \rightarrow AA; 0; 1) \quad \cdot P(AA \rightarrow B; 1; 2) \quad + \\ & P(AA \rightarrow A; 0; 1) \quad \cdot P(A \rightarrow B; 1; 2) \quad + \\ & P(AA \rightarrow BBB; 0; 1) \quad \cdot P(BBB \rightarrow B; 1; 2) \quad + \\ & P(AA \rightarrow BB; 0; 1) \quad \cdot P(BB \rightarrow B; 1; 2) \quad + \\ & P(AA \rightarrow B; 0; 1) \quad \cdot P(B \rightarrow B; 1; 2) \quad + \\ & P(AA \rightarrow CCC; 0; 1) \quad \cdot P(CCC \rightarrow B; 1; 2) \quad + \\ & P(AA \rightarrow Default; 0; 1) \quad \cdot P(Default \rightarrow B; 1; 2) \end{aligned}$$

Notamos que se trata del producto de la fila correspondiente al rating *AA* (rating de salida) por la columna correspondiente al rating *B* (rating de llegada).

Proposición. Sean M_{T_1} y M_{T_2} las matrices de transición para los periodos T_1 y T_2 . Entonces, la matriz de transición para el periodo $T_1 + T_2$ es:

$$M_{T_1+T_2} = M_{T_1} \cdot M_{T_2} \quad (2.17)$$

Corolario. Sean M_T la matriz de transición para el periodo T y $k \in \mathbb{N}$. Entonces³:

$$M_{k \cdot T} = M_T^k \quad (2.18)$$

$$M_{\frac{T}{k}} = \sqrt[k]{M_T} \quad (2.19)$$

2.3.3. Función de supervivencia

Definición. La *Tasa de Morosidad Anticipada Acumulada* o *Cumulated Forward Default Rate* (CFDR) del rating r_i en el tiempo t es la probabilidad que una empresa con rating inicial r_i haga fallido en el intervalo de tiempo $(0, t)$.

$$CFDR(r_i, t) = P(r_i \rightarrow Default; 0; t) \quad (2.20)$$

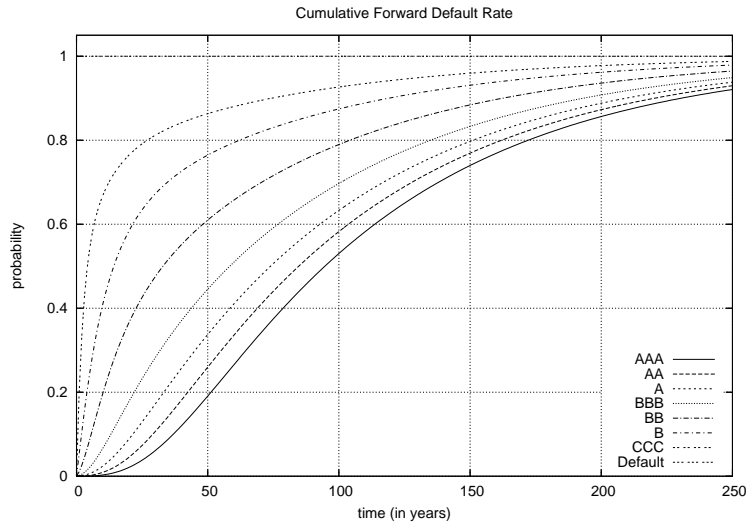


Figura 2.5: Tasa de Morosidad Anticipada Acumulada (CFDR)

Definición. La *Supervivencia* en el tiempo t del rating r_i es la probabilidad que una empresa con rating inicial r_i no haya hecho fallido en el intervalo de tiempo $(0, t)$.

³véase el apéndice A.8 para ver como se calcula la raíz de una matriz

Proposición. La Tasa de Morosidad Anticipada Acumulada se puede expresar en función de la matriz de transición a través de la relación siguiente:

$$CFDR(r_i, k \cdot T) = (M_{k \cdot T})_{i,n} = (M_T^k)_{i,n} \quad (2.21)$$

donde n es el índice del rating Default y T es el periodo de la matriz de transición.

Proposición. La Supervivencia puede expresarse en función de la Tasa de Morosidad Anticipada Acumulada a través de la relación siguiente:

$$Survival(r_i, t) = 1 - CFDR(r_i, t) \quad (2.22)$$

Proposición. Si la matriz de transición es válida, cualquier rating inicial acaba haciendo fallido casi seguramente.

$$\lim_{t \rightarrow \infty} CFDR(r_i, t) = 1 \quad \forall i \quad (2.23)$$

Proposición. Fijado un rating, r_i , la función de supervivencia es monótona decreciente. Si el rating es *Default*, el valor de la función de supervivencia es siempre 0.

$$Survival(r_i, t_j) \geq Survival(r_i, t_k) \quad \forall t_j < t_k \quad (2.24)$$

$$Survival(Default, t) = 0 \quad \forall t \quad (2.25)$$

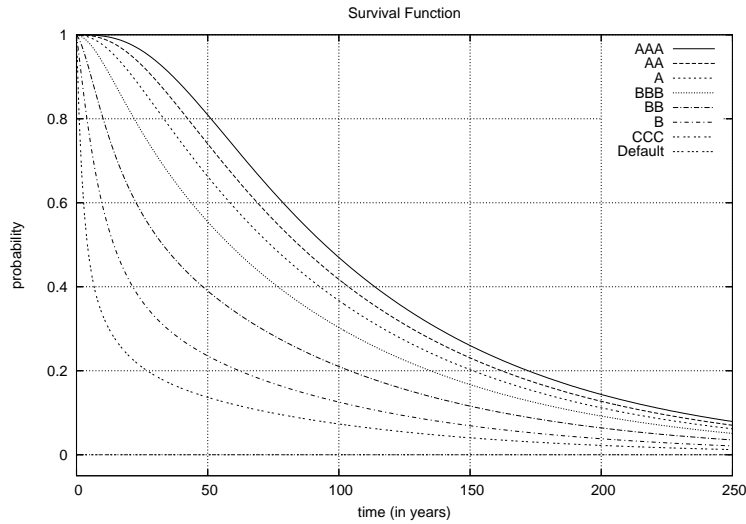


Figura 2.6: Función de Supervivencia

Ejemplo. En las figuras 2.5 y 2.6 se puede observar la Tasa de Morosidad Anticipada Acumulada y la Función de Supervivencia de la matriz de transición usada en este documento.

2.4. Matriz de correlación

Definición. La *tabla de correlaciones sectoriales* proporciona la correlación de los fallidos entre los sectores. La denotamos de la forma siguiente:

	<i>Sector</i> ₁	...	<i>Sector</i> _{<i>m</i>}
<i>Sector</i> ₁	$\gamma_{1,1}$...	$\gamma_{1,m}$
\vdots	\vdots	\ddots	\vdots
<i>Sector</i> _{<i>m</i>}	$\gamma_{1,m}$...	$\gamma_{m,m}$

donde m es el número de sectores, $\gamma_{i,j} = \text{Corr}(\text{Sector}_i, \text{Sector}_j)$ es la correlación entre los fallidos de los sectores i y j y $\gamma_{i,i}$ es la correlación del fallido entre las empresas del sector i . No se trata de una verdadera matriz de correlación debido a que la diagonal puede ser distinta de 1.

Definición. La *matriz de correlación entre clientes* proporciona la correlación de los fallidos entre clientes. La construimos a partir de la tabla de correlaciones sectoriales de la forma siguiente:

$$\Theta = \begin{pmatrix} 1 & \theta_{1,2} & \dots & \theta_{1,p-1} & \theta_{1,p} \\ \theta_{1,2} & 1 & \dots & \theta_{2,p-1} & \theta_{2,p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \theta_{1,p-1} & \theta_{2,p-1} & \dots & 1 & \theta_{p-1,p} \\ \theta_{1,p} & \theta_{2,p} & \dots & \theta_{p-1,p} & 1 \end{pmatrix}$$

donde p es el número de clientes y $\theta_{i,j}$ es la correlación sectorial entre el sector del cliente i y el sector del cliente j . Por construcción, la matriz de correlación entre clientes es simétrica debido a que la correlación entre el sector del cliente i con el sector del cliente j es la misma que la correlación del sector del cliente j con el sector del cliente i .

Observación. Los clientes se acostumbran a ordenar por sectores. En este caso la matriz de correlación entre clientes queda de la forma siguiente:

$$\Theta = \begin{pmatrix} 1 & \dots & \gamma_{p_1,p_1} & \gamma_{1,p_i} & \dots & \gamma_{1,p_i} & \gamma_{1,p_m} & \dots & \gamma_{1,p_m} \\ \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \gamma_{p_1,p_1} & \dots & 1 & \gamma_{1,p_i} & \dots & \gamma_{1,p_i} & \gamma_{1,p_m} & \dots & \gamma_{1,p_m} \\ & & & \ddots & & & & & \\ \gamma_{1,p_i} & \dots & \gamma_{1,p_i} & 1 & \dots & \gamma_{p_i,p_i} & \gamma_{p_i,p_m} & \dots & \gamma_{p_i,p_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ \gamma_{1,p_i} & \dots & \gamma_{1,p_i} & \gamma_{p_i,p_i} & \dots & 1 & \gamma_{p_i,p_m} & \dots & \gamma_{p_i,p_m} \\ & & & & & & \ddots & & \\ \gamma_{1,p_m} & \dots & \gamma_{1,p_m} & \gamma_{p_i,p_m} & \dots & \gamma_{p_i,p_m} & 1 & \dots & \gamma_{p_m,p_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{1,p_m} & \dots & \gamma_{1,p_m} & \gamma_{p_i,p_m} & \dots & \gamma_{p_i,p_m} & \gamma_{p_m,p_m} & \dots & 1 \end{pmatrix}$$

donde p_1, \dots, p_m son el número de clientes que pertenecen a los sectores s_1, \dots, s_m . Con esta ordenación de clientes, la matriz de correlación entre clientes es una matriz con bloques con 1's en la diagonal.

Ejemplo. Supongamos que tenemos dos sectores (S_1 y S_2) cumpliendo la siguiente tabla de correlaciones sectoriales:

	S_1	S_2
S_1	0	0,1
S_2	0,1	-0,2

Supongamos que el sector S_1 tiene 3 clientes y el sector S_2 tiene 2 clientes. Entonces la matriz de correlación entre clientes es:

$$\Theta = \begin{pmatrix} 1 & 0 & 0 & 0,1 & 0,1 \\ 0 & 1 & 0 & 0,1 & 0,1 \\ 0 & 0 & 1 & 0,1 & 0,1 \\ 0,1 & 0,1 & 0,1 & 1 & -0,2 \\ 0,1 & 0,1 & 0,1 & -0,2 & 1 \end{pmatrix}$$

Observación. En general se impone que la matriz de correlación entre clientes sea definida positiva debido a que es una propiedad necesaria para la generación de cópulas gaussianas. El hecho que la matriz de correlación entre clientes deba ser definida positiva no significa que la tabla de correlaciones sectoriales deba ser definida positiva.

2.5. Valoración del riesgo

Llamamos Z a la variable aleatoria que representa las pérdidas de la cartera (portfolio loss) en el tiempo T . Sea F_Z la correspondiente función de distribución (cdf).

Definición. La *Pérdida Esperada* o *Expected Loss* de la cartera en tiempo T es:

$$\text{Expected Loss} = E(Z) \quad (2.26)$$

Definición. El *Valor en Riesgo* o *VAR* en tiempo T es la pérdida máxima esperada dentro de un intervalo de confianza dado. Responde a la pregunta *Cual es la pérdida mínima incurrida en el $\alpha\%$ de los peores casos?*. Se recomienda la lectura del libro *Value at Risk* [6].

$$VAR_\alpha(Z) = \inf\{z | F_Z(z) \geq \alpha\} \quad (2.27)$$

Definición. El *Tail Conditional Expectation* o *TCE* o *Expected Shortfall* es la pérdida esperada en el caso en que la pérdida sea inferior a un cuantil fijado. Responde a la pregunta *Cual es la pérdida media incurrida en el $\alpha\%$ de los peores casos?*. Se recomienda la lectura de los artículos [5] y [1].

$$TCE_\alpha(Z) = E(Z | Z > VAR_\alpha(Z)) \quad (2.28)$$

donde la notación $E(Z|q)$ debe interpretarse como la esperanza de la variable aleatoria Z condicionada a q .

Definición. Definimos el *Capital Económico* al nivel de confianza α en tiempo T como:

$$\text{Economic Capital} = VAR_\alpha(Z) - E(Z) \quad (2.29)$$

Ejemplo. Calculemos el VAR de una cartera de créditos sencilla, de la que se puede obtener la distribución de las pérdidas de forma explícita. Sea una cartera de 1 solo sector con 100 clientes sin correlación alguna entre ellos. Cada cliente tiene un activo que consiste en devolver 1 € al cabo de T tiempo. Supongamos que el sistema de ratings solamente contempla 2 categorías crediticias, no-Default y Default. La probabilidad de hacer fallido al cabo de 1 año es 0,1.

En este caso, se puede modelar la pérdida provocada por el impago del cliente i al cabo de 1 año como una variable aleatoria Bernouilli, $X_i \sim Ber(0,1)$. La pérdida de la cartera al cabo de 1 año es la suma de las

pérdidas de cada cliente, $Z = \sum_{i=1}^{100} X_i$, que por definición es una variable aleatoria Binomial, $Z \sim B(100, 0.1)$.

La Pérdida Esperada al cabo de 1 año es

$$\text{Expected Loss} = E(Z) = 10$$

El VAR al nivel de confianza 99 % al cabo de 1 año es:

$$VAR_{99\%} = \inf\{z | F_Z(z) \geq 99\%\}$$

Expresado en términos de la función de densidad

$$P(Z \leq VAR_{99\%}) = 99\%$$

Consultamos las tablas cdf de la Binomial(100,0.1)

$$P(Z \leq 17) = 98,99927\%$$

Por tanto, podemos considerar que $VAR_{99\%}$ es

$$VAR_{99\%} = 17$$

El TCE al nivel de confianza 99 % al cabo de 1 año es:

$$TCE_{99\%} = E(Z | Z > VAR_{99\%}(Z))$$

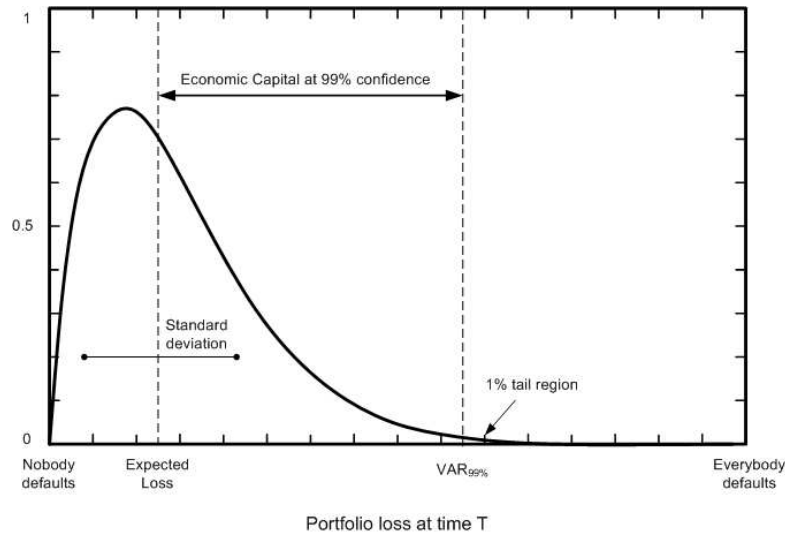


Figura 2.7: Portfolio value at time T

Equivalentemente

$$TCE_{99\%} = \frac{\sum_{i=18}^{100} i \cdot P(Z = i)}{\sum_{i=18}^{100} P(Z = i)} = 18,78458$$

El Capital Económico al nivel de confianza 99 % al cabo de 1 año es:

$$\text{Economic Capital}_{99\%}(Z) = \text{VAR}_{99\%}(Z) - E(Z) = 17 - 10 = 7$$

Este ejemplo no es significativo debido a que se han realizado dos supuestos que en el mundo real no se cumplen: todos los creditores se modelan de la misma forma y los fallidos son independientes. Se obtiene que la distribución de las pérdidas de la cartera al cabo de 1 año es una variable aleatoria Binomial(100,0.1), que puede ser aproximada de forma precisa por una Normal(10,9). Esto no concuerda con las observaciones reales, que muestran que la distribución de las pérdidas es fuertemente asimétrica respecto a la pérdida esperada.

Capítulo 3

Resolución

3.1. Esquema general

El método de resolución implementado por CreditCruncher consiste en aplicar el método de Monte Carlo. En la figura 3.1 se muestra el esquema de un paso del método de Monte Carlo.

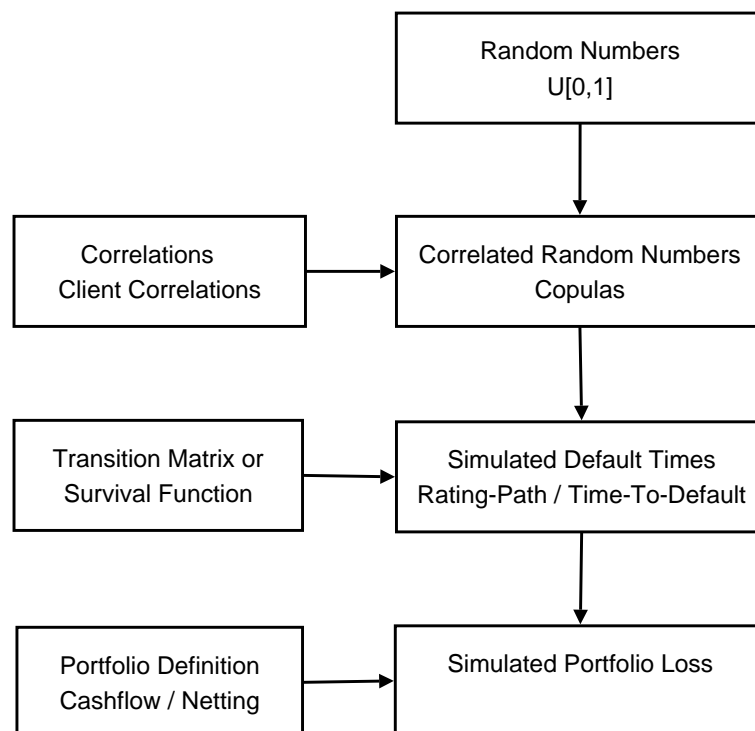


Figura 3.1: Esquema de la simulación

3.2. El método de Monte Carlo

Se recomienda la lectura de la referencia [7]. Se trata de los apuntes para una clase del profesor Mervyn Marasinghe. Se expone el método de Monte Carlo y las técnicas de reducción de la varianza.

Definición. Dado un conjunto de observaciones, x_1, \dots, x_N , de la variable aleatoria X , definimos la *función de distribución empírica* como:

$$\widetilde{F_X}(y) = \frac{1}{N} \sum_{i=1}^N I_{(-\infty, y]}(x_i) \quad I_{(-\infty, y]}(x) = \begin{cases} 1 & \text{if } x \in (-\infty, y] \\ 0 & \text{otherwise} \end{cases}$$

Proposición. La función de distribución empírica tiende a la función de distribución al incrementar el número de observaciones.

$$\lim_{n \rightarrow \infty} \widetilde{F_X} = F_X$$

Definición. Sea X una variable aleatoria con función de distribución conocida, F . El *método de Monte Carlo* consiste en obtener la función de distribución empírica de la variable aleatoria $H(X)$ usando el siguiente método:

$$\begin{array}{ccc} F_X & & \widetilde{F_{H(X)}} \\ \downarrow & & \uparrow \\ \text{simulation} & & \text{empirical cdf} \\ \downarrow & & \uparrow \\ x_1, \dots, x_N & \longrightarrow & H(x_1), \dots, H(x_N) \end{array}$$

Observación. Problemas aparentemente no relacionados con las variables aleatorias pueden reformularse como un problema donde intervenga una variable aleatoria y ser resueltos por el método de Monte Carlo. Normalmente se formula el problema original como el cálculo de un estadístico (pe. la media) sobre la función de distribución. Al método de Monte Carlo se le atribuye una velocidad de convergencia del orden de $1/\sqrt{N}$.

Ejemplo: El ejemplo clásico es obtener el valor de la integral de la función W entre 0 y 1. Lo reformulamos de la siguiente forma:

$$\int_0^1 W(u) du = \int_0^1 W(u) \phi(u) du = E[W(U)]$$

donde $U \sim U[0, 1]$ y $\phi(u) = \text{pdf}(U) = 1$. La última igualdad se establece usando la preposición enunciada en el apéndice A.1. Finalmente la integral se aproxima calculando la media de un conjunto de puntos con distribución $W(U)$.

3.3. Variables aleatorias correlacionadas

Se recomienda la lectura de las referencias [12] y [8]. Se trata de artículos donde se explica que es una cópula, sus propiedades, como simularlas, creencias erróneas, etc.

Definición. Llamamos *cópula* a la función de distribución de una variable aleatoria n -dimensional tal que sus distribuciones marginales son variables aleatorias $U[0, 1]$.

$$C(u_1, \dots, u_n) = P(U_1 \leq u_1, \dots, U_n \leq u_n) \quad U_k \sim U[0, 1]$$

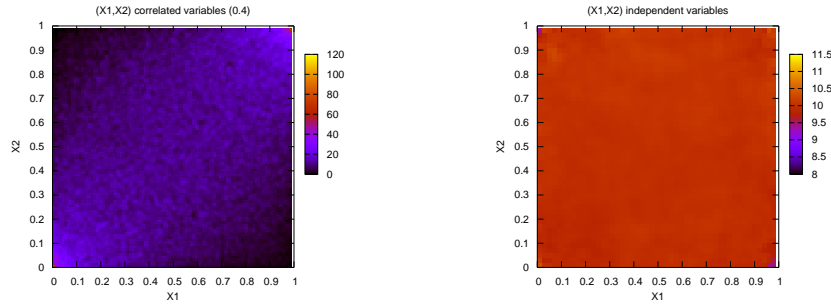


Figura 3.2: Bivariate distribution plot with correlated and uncorrelated variables

Teorema. Toda variable aleatoria n -dimensional puede separarse en las distribuciones seguidas por sus componentes (las distribuciones marginales) y una cópula. Sea F una función de distribución n -dimensional y f_1, \dots, f_n sus marginales. El teorema de Sklar asegura que existe una cópula C tq.

$$F(x_1, \dots, x_n) = C(f_1(x_1), \dots, f_n(x_n))$$

Observación. Una variable aleatoria n -dimensional no está determinada por sus marginales y correlaciones entre estas. Dicho de otra forma, existen infinitas formas de combinar las distribuciones marginales a través de cópulas de forma que cumplan las correlaciones. Las distribuciones elípticas (que incluyen la distribución multinomial) son una excepción.

3.4. Simulación del tiempo de fallido

El objetivo de este apartado es proporcionar métodos para simular los tiempos de fallido, $\vec{t}_w = (t_w^1, t_w^2, t_w^3, \dots, t_w^{n_c})$ de los clientes cumpliendo:

- los fallidos de cada cliente deben satisfacer la matriz de transición M_T (o la función de supervivencia, $Survival(r, t)$).
- la correlación entre los fallidos de los clientes debe cumplir la matriz de correlación entre clientes, Θ

3.4.1. Método Rating-Path

Este método simula la evolución del rating del cliente a lo largo del tiempo. El tiempo de fallido se obtiene al alcanzar el rating *Default*. Para este método es necesario disponer de la matriz de transición, M_T .

Paso 1. Segmentamos el tiempo en intervalos constantes de forma que el intervalo de tiempo entre t_i y t_{i+1} es el cubierto por la matriz de transición, T .

$$t_0, t_1, t_2, \dots, t_k, \dots$$

Paso 2. Para cada intervalo de tiempo k creamos una cópula, Q_k de dimensión nc que satisfaga la matriz de correlación entre clientes, Θ .

$$Q_1, Q_2, Q_3, \dots, Q_k, \dots$$

Paso 3. Simulamos las cópulas $Q_1, Q_2, Q_3, \dots, Q_k, \dots$. Cada realización de Q_k consiste en un vector de componentes $(Q_k(1), Q_k(2), \dots, Q_k(nc))$.

Paso 4. Simulamos el rating del cliente i en t_1 de la forma siguiente: disponemos del rating inicial $rating(i, t_0)$ y un valor $Q_1(i) \in [0, 1]$ proporcionado por el componente i -ésimo de Q_1 . Entonces

$$rating(i, t_1) = M_T^{-1}(rating(i, t_0), Q_1(i))$$

Este paso se encuentra ilustrado en la figura 3.3.

Paso 5. Simulamos el rating del cliente i en t_k de la forma siguiente: disponemos del rating anterior $rating(i, t_{k-1})$ y un valor $Q_k(i) \in [0, 1]$ proporcionado por el componente i -ésimo de Q_k . Entonces

$$rating(i, t_k) = M_T^{-1}(rating(i, t_{k-1}), Q_k(i))$$

Este paso se encuentra ilustrado en la figura 3.3. Finalmente, aproximamos el tiempo de fallido del cliente i de la siguiente forma:

$$t_w^i = \inf\{t_k | rating(i, t_k) = Default\}$$

3.4.2. Método Time-To-Default

Este método consiste en simular directamente el tiempo de fallido de los clientes. Requiere disponer de la función de supervivencia. En caso de disponer solamente de la matriz de transición puede calcularse la función de supervivencia asociada a la matriz de transición usando las fórmulas 2.22 y 2.21.

Paso 1. Creamos una cópula de dimensión nc que satisfaga la matriz de correlación entre clientes, Θ .

$$Q$$

Paso 2. Simulamos la cópula Q . La realización de Q consiste en un vector de componentes $(Q(1), Q(2), \dots, Q(nc))$.

Paso 2. Simulamos el tiempo de fallido del cliente i de la forma siguiente: disponemos del rating inicial, $rating(i, t_0)$ y un valor $Q(i) \in [0, 1]$ proporcionado por el componente i -ésimo de Q . Entonces:

$$t_w^i = Survival^{-1}(rating(i, t_0), Q(i))$$

Este paso se encuentra ilustrado en la figura 3.4.

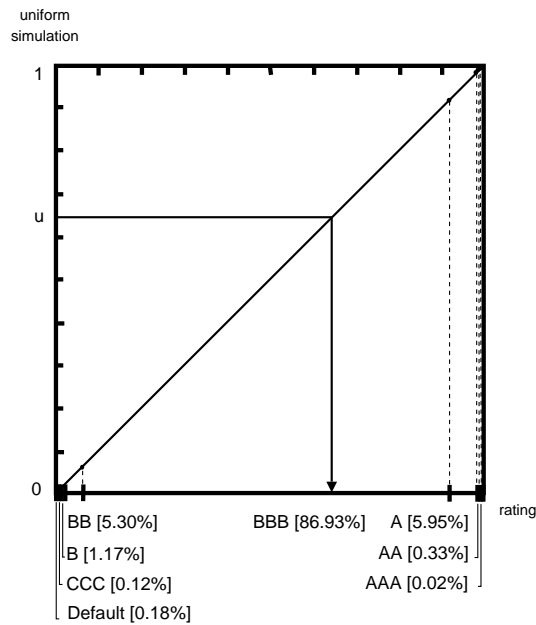


Figura 3.3: Simulación de la evolución del rating *BBB* a T tiempo

3.5. Evaluación de la cartera

En las definiciones siguientes se considera que S es la curva spot en t_0 , na_i es el número de activos del cliente i y nc el número de clientes. Asimismo se considera que el cashflow y recovery de cada activo está fijado y no depende de la evolución del cliente.

Definición. Definimos el valor del activo j del cliente i evaluado en t_K entre el tiempo t_0 y t_K , habiendo hecho fallido en tiempo t_w como:

$$X_{i,j}(t_w; t_0, t_K) = \quad (3.1)$$

$$1_{[t_w \leq t_K]} \cdot \text{recovery}(t_w) \cdot \Upsilon_S(t_w, t_K) + \sum_{t=t_0}^{t_K} 1_{[t < t_w]} \cdot \text{cashflow}(t) \cdot \Upsilon_S(t, t_K)$$

Definición. Definimos el valor de los activos del cliente i evaluados en t_K entre el tiempo t_0 y t_K , habiendo hecho fallido en tiempo t_w como:

$$X_i(t_w; t_0, t_K) = \sum_{j=0}^{na_i} X_{i,j}(t_w; t_0, t_K) \quad (3.2)$$

Definición. Definimos el valor de la cartera evaluada en t_K entre el tiempo t_0 y t_K , siendo $\vec{t_w}$ los tiempos en los que han hecho fallido los clientes como:

$$X(\vec{t_w}; t_0, t_K) = \sum_{i=0}^{nc} X_i(\vec{t_w}; t_0, t_K) = \sum_{i=0}^{nc} \sum_{j=0}^{na_i} X_{i,j}(\vec{t_w}; t_0, t_K) \quad (3.3)$$

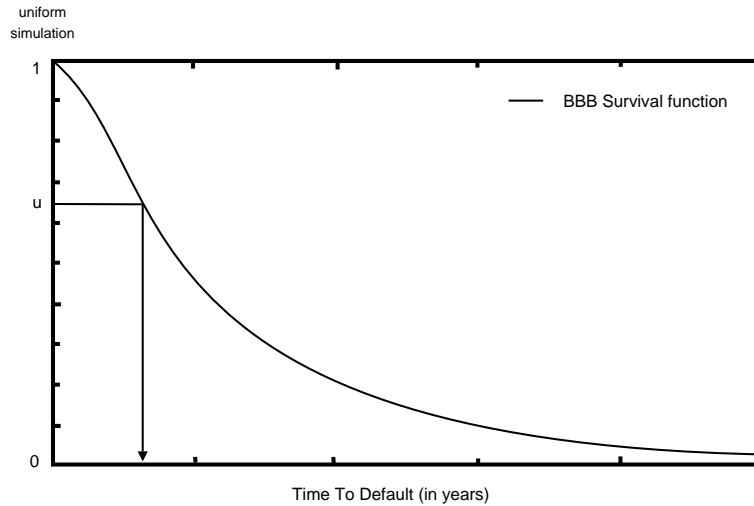


Figura 3.4: Simulación del tiempo hasta el fallido del rating BBB

Definición. Definimos las pérdidas de la cartera evaluada en t_K entre el tiempo t_0 y t_K , siendo \vec{t}_w los tiempos en los que han hecho fallido los clientes como:

$$Z(\vec{t}_w; t_0, t_K) = X(\vec{t}_\infty; t_0, t_K) - X(\vec{t}_w; t_0, t_K) \quad (3.4)$$

3.6. Valoración del riesgo

El método de Monte Carlo genera N realizaciones, $\{z_1, z_2, z_3, \dots, z_N\}$, de la variable aleatoria Z (pérdidas de la cartera) que permiten obtener la función de distribución empírica. A continuación se exponen los estadísticos calculados, sus estimaciones y el intervalo de confianza al nivel de confianza α . CreditCruncher utiliza el entorno R ¹ para realizar los cálculos estadísticos. Para mas información acerca de R véase [10]. Sea $\phi^{-1}(x)$ la inversa de la función de distribución de la Normal(0,1).

Expected Loss. Fórmula extraída de [11] basada en el teorema del límite central.

$$\mu_Z = \widehat{\mu}_Z \pm \phi^{-1}\left(\frac{1-\alpha}{2}\right) \cdot \frac{\widehat{\sigma}_Z}{\sqrt{N}}$$

Consulte A.5 para el cálculo de $\widehat{\mu}_Z$ y $\widehat{\sigma}_Z$.

Desviación estándar. Fórmula extraída de [11] basada en el teorema del límite central.

$$\sigma_Z = \widehat{\sigma}_Z \pm \phi^{-1}\left(\frac{1-\alpha}{2}\right) \cdot \frac{\widehat{\sigma}_Z}{\sqrt{2N}}$$

Consulte A.5 para el cálculo de $\widehat{\sigma}_Z$.

Value at Risk. Fijado un nivel de $VAR = x$

$$VAR_x(Z) = \widehat{q_x(Z)} \pm \phi^{-1}\left(\frac{1-\alpha}{2}\right) \cdot \text{stderr}(q_x(Z))$$

Consulte A.5 y A.6 para el cálculo de $\widehat{q_x(Z)}$ y $\text{stderr}(q_x(Z))$.

TCE o Expected Shortfall. Fijado un nivel de $VAR = x$

$$TCE_x(Z) = \widehat{TCE_x(Z)} \pm \phi^{-1}\left(\frac{1-\alpha}{2}\right) \cdot \text{stderr}(TCE_x(Z))$$

Consulte A.5 y A.7 para el cálculo de $\widehat{TCE_x(Z)}$ y $\text{stderr}(TCE_x(Z))$.

¹<http://www.r-project.org>

Capítulo 4

Implementación

El método de resolución expuesto en el capítulo anterior es computacionalmente muy costoso. Para reducir el número de operaciones se discretiza el tiempo. Esto fuerza a mapear el recovery y el cashflow a los nodos de tiempo. El tiempo de fallido de los clientes debe simularse mapeado sobre los nodos. La unidad de tiempo es el mes.

4.1. Paso 1. Interpretación del fichero de entrada

El formato del fichero de entrada se encuentra descrito en el documento *CreditCruncher - Input File Reference*.

El fichero de entrada puede tener un tamaño considerable (superior a los 100 Mb.). Por este motivo la interpretación del fichero xml se realiza usando un sistema orientado a eventos (tipo SAX). A continuación se describen algunas de las validaciones más importantes:

Formato. Se verifica que se trata de un fichero XML válido que cumple la DTD. Véase *W3C*¹ para más información relativa al formato XML.

Valores. Cada valor tiene un tipo (int, long, double, date, boolean o string), un rango de valores permitidos y un indicador de obligatorio/opcional. Para cada valor se comprueba que se cumplen los criterios descritos en *CreditCruncher - Input File Reference*.

Consistencia. Cuando un valor se refiere a un identificador, se comprueba que el objeto referenciado existe. Por ejemplo, cuando se lee una transición en la definición de la matriz de transición se comprueba que los ratings han sido definidos anteriormente y que los ratings referenciados existen.

¹<http://www.w3.org/XML/>

Matriz de transición. Al leer la matriz de transición se verifica que cumple con las propiedades indicadas en 2.3.1.

Función de supervivencia. En caso de estar definida una función de supervivencia, se comprueba que se trata de una función positiva monótona decreciente que vale 1 cuando $t = 0$, (excepto para el rating *Default* para la que siempre vale 0). Véase las fórmulas 2.24 y 2.25.

Matriz de correlación. Se comprueba que se trata de una matriz simétrica (por comodidad, puede entrarse solamente el triangulo superior, o inferior), con valores comprendidos en $[-1, +1]$.

Criterios de parada. Se comprueba que exista algún criterio de parada y que este sea realizable.

4.2. Paso 2. Particionamiento del tiempo

Para realizar la discretización del tiempo se necesita la fecha inicial, t_0 , la longitud (en meses naturales) de cada intervalo, $StepLength$, y el número de pasos a considerar, $NumSteps$. Los nodos de tiempo son:

$$t_i = t_0 + i \cdot StepLength \quad i \in \{0, 1, 2, \dots, NumSteps\}$$

Entendemos que añadir n meses a una fecha consiste en incrementar el mes de la fecha inicial en n meses, realizando un incremento de año si es preciso. Si el día de la fecha inicial no existe en el mes de la fecha final, consideramos como día de la fecha final el máximo día del mes de la fecha final, en caso contrario, el mismo día de la fecha inicial.

Supongamos que $t_0 = 30/10/2004$, $StepLength = 2$ y $NumSteps = 6$, entonces la partición del tiempo es: $t_0 = 30/10/2004$, $t_1 = 30/12/2004$, $t_2 = 28/02/2005$, $t_3 = 30/04/2005$, $t_4 = 30/06/2005$, $t_5 = 30/08/2005$ y $t_6 = 30/10/2005$. Obsérvese como el día del mes de Febrero es el 28 por no existir los días 29 y 30 de dicho mes.

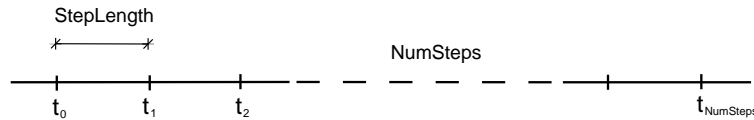


Figura 4.1: Particionamiento del tiempo

4.3. Paso 3. Inicialización de los clientes

Los clientes de la cartera se ordenan en función de su sector y rating de forma creciente en ambos casos. La finalidad de esta ordenación es disponer de una matriz de correlación entre clientes en bloques. Véase las observaciones del apartado 2.4. En el caso que solamente se desee simular los clientes activos (aquellos que tienen algún producto activo durante el periodo simulado), los clientes inactivos son suprimidos.

Supongamos que tenemos 3 sectores, S1, S2 y S3 y un sistema de ratings compuesto por las clasificaciones A, B, C, D, y E. Supongamos que la cartera está compuesta por ²: [S2,D,1], [S3,A,1], [S1,B,1], [S1,B,0], [S3,C,1], [S2,A,1], [S2,B,1], [S3,C,1], [S1,A,1], [S2,D,1].

La cartera ordenada es: [S1,A,1], [S1,B,1], [S1,B,0], [S2,A,1], [S2,B,1], [S2,D,1], [S2,D,1], [S3,A,1], [S3,C,1], [S3,C,1].

Si solamente se desea simular los clientes activos, la cartera ordenada es: [S1,A,1], [S1,B,1], [S2,A,1], [S2,B,1], [S2,D,1], [S2,D,1], [S3,A,1], [S3,C,1], [S3,C,1].

El número de clientes considerado es el número de clientes de la cartera. Si solamente se está simulando los clientes activos, entonces el número de clientes considerado es el número de clientes activos.

4.4. Paso 4. Inicialización del método de simulación

4.4.1. Método Rating-Path

El método Rating-Path requiere disponer de la matriz de transición para el periodo *StepLength*. A partir de la matriz de transición definida por el usuario para el periodo T (normalmente 1 año), se construye la matriz de transición para el periodo *StepLength* usando las indicaciones de 2.3.2.

$$M_{StepLength} = M_T^{\frac{StepLength}{T}} = \sqrt[T]{M_T^{StepLength}}$$

El algoritmo para calcular la raíz de una matriz a partir de la matriz de autovectores y sus autovalores está explicado en el apéndice A.8.

²el primer componente corresponde al sector del cliente, el segundo al rating del cliente y el tercero indica si está activo (1) o inactivo (0)

4.4.2. Método Time-To-Default

El método Time-To-Default requiere disponer de la función de supervivencia. Si el usuario ha definido una se considera la función de supervivencia definida por el usuario. En caso contrario se construye a partir de la matriz de transición usando las indicaciones de 2.3.3.

$$Survival(r_i, t_j) = 1 - (M_{StepLength}^j)_{i,n} \quad j \in \{0, 1, 2, \dots, NumSteps\}$$

donde $(M_{StepLength}^j)_{i,n}$ debe interpretarse como la columna n de la fila i de la matriz $M_{StepLength}$ elevada a j . La matriz de transición para el periodo $StepLength$, $M_{StepLength}$, se determina a partir de la matriz de transición proporcionada por el usuario para el periodo T (normalmente 1 año) usando las indicaciones de 2.3.2.

$$M_{StepLength} = M_T^{\frac{StepLength}{T}} = \sqrt[T]{M_T^{StepLength}}$$

El algoritmo para calcular la raíz de una matriz a partir de sus autovalores y autovectores está explicado en el apéndice A.8.

4.5. Paso 5. Inicialización de las cópulas

Los pasos 1 y 2 de A.9 explican como inicializar una cópula gaussiana. CreditCruncher aplica el mismo algoritmo, pero teniendo en cuenta que la matriz de correlación entre clientes es una matriz en bloques. Esto permite un ahorro de memoria y tiempo de cálculo. En vez de considerar la matriz de

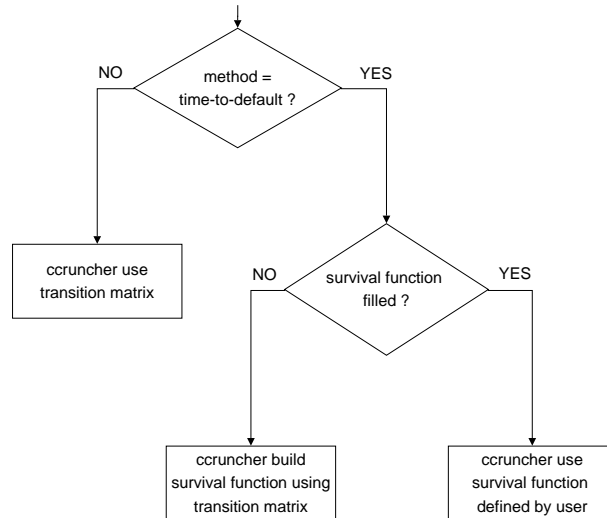


Figura 4.2: Decisión en función del método

correlación entre clientes se considera la matriz de correlación entre sectores y el número de clientes en cada sector. Se aplica el paso 1 de A.9 a la matriz de correlación entre sectores. Se reemplaza el paso 2 de A.9 por A.10.

4.5.1. Método Rating-Path

Si el método de simulación es Rating-Path se construyen $NumSteps$ cópulas. La primera cópula se usará para simular los cambios de rating entre t_0 y t_1 , la segunda cópula para simular los cambios de rating entre t_1 y t_2 , etc.

El generador de números aleatorios de cada cópula se inicializa con la siguiente semilla:

$$Seed(i, q) = Seed(1, 0) + q \cdot 30001 + (i - 1)$$

donde $Seed(i, q)$ es la semilla del generador de números de la cópula i de la instancia q (en el caso de ejecutarse en un cluster $q > 0$, en caso contrario, $q = 0$). $Seed(1, 0)$ es la semilla proporcionada por el usuario (un número aleatorio si el usuario no especifica semilla o especifica una semilla igual a 0). De esta forma se garantiza que todas las instancias usan una semilla distinta (siempre que el número de instancias más $NumSteps$ no supere 30000).

4.5.2. Método Time-To-Default

Si el método de simulación es Time-To-Default se construye 1 cópula. Se usará para simular el tiempo de fallido.

El generador de números aleatorios de cada cópula se inicializa con la siguiente semilla:

$$Seed(1, q) = Seed(1, 0) + q \cdot 30001 + 0$$

donde $Seed(i, q)$ es la semilla del generador de números de la cópula i de la instancia q (en el caso de ejecutarse en un cluster $q > 0$, en caso contrario, $q = 0$). $Seed(1, 0)$ es la semilla proporcionada por el usuario (un número aleatorio si el usuario no especifica semilla o especifica una semilla igual a 0). De esta forma se garantiza que todas las instancias usan una semilla distinta (siempre que el número de instancias no supere 30000).

4.6. Paso 6. Inicialización de los agregadores

Definición. Llamamos *segmentación* a una partición de la cartera en grupos de elementos llamados *segmentos*. Si los elementos son activos, decimos que es una segmentación de activos, si los elementos son clientes, decimos

que es una segmentación de clientes. Si los elementos de una segmentación son clientes, CreditCruncher lo interpreta como los activos de los clientes que componen cada segmento. Se contempla la opción de los clientes por comodidad en la creación de los segmentos.

Observación. Dada una segmentación, la intersección de dos segmentos cualquiera es vacía y la unión de todos los segmentos es el total (la cartera).

Observación. La cartera entera se puede expresar como una segmentación de activos con un único segmento.

Observación. Las segmentaciones de clientes se encuentran incluidas dentro de las segmentaciones de activos.

Ejemplo. La figura 4.3 ilustra la segmentación de la cartera en zonas comerciales. Cada segmento es una zona comercial. Los componentes pueden ser activos (zona de contratación del activo), o clientes (cliente asignado a una oficina perteneciente a la zona X).

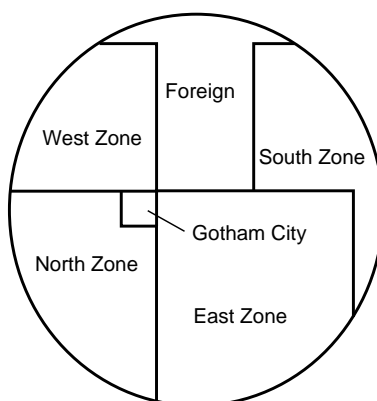


Figura 4.3: Portafolio segmented by comercial zones

CreditCruncher simula segmentos de activos (si son segmentos de clientes lo interpreta como los activos de aquellos clientes). Permite la definición de varias segmentaciones simultáneas. Recordemos que la cartera entera puede expresarse como un único segmento. De esta forma se puede valorar el riesgo teniendo en cuenta la organización jerárquica de la empresa u otro tipo de segmentación (zonas comerciales, tipos de activos, tipos de clientes, sectores, etc.).

Para cada cliente de cada segmento de cada segmentación realizamos el mapeo de cashflow y recovery de los activos pertenecientes a aquel seg-

mento. El objetivo es disponer de un cashflow y recovery evaluados en $t_0, t_1, t_2, \dots, t_{NumSteps}$ para cada cliente de cada segmento.

4.6.1. Mapeo del cashflow

Paso 1

Sea $t_0, t_1, t_2, \dots, t_{NumSteps}$ la partición de tiempo usada. Dado un cashflow (t_i, V) lo mapeamos en la estructura temporal de la siguiente forma:

1. Si $t_i < t_0$, entonces $(t_i, V) \longrightarrow (t_0, 0)$
2. Si $t_{j-1} < t_i \leq t_j$, entonces $(t_i, V) \longrightarrow (t_j, V \cdot \Upsilon_S(t_i, t_j))$
3. Si $t_{NumSteps} < t_i$, entonces $(t_i, V) \longrightarrow (t_{NumSteps}, V \cdot \Upsilon_S(t_i, t_{NumSteps}))$

donde $\Upsilon_S(t_i, t_j)$ es el coeficiente de transporte entre t_i y t_j usando la curva spot S . El cashflow mapeado a $t_0, t_1, t_2, \dots, t_{NumSteps}$ lo denotamos como $V_0, V_1, V_2, \dots, V_{NumSteps}$.

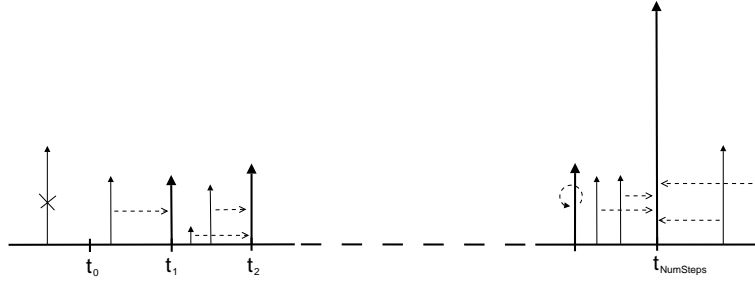


Figura 4.4: CashFlow mapping

Veamos un ejemplo. Consideramos $\Upsilon_S(t_i, t_j) = 1$ para no complicar los cálculos. A la izquierda se representa el cashflow original (sin estar mapeado a los nodos de tiempo). A la derecha se encuentra el mismo cashflow mapeado a los nodos de tiempo.

Date	Cashflow	t_i	Date	Cashflow
17/04/2004	-10.00			
		t_0	30/10/2004	0.00
		t_1	30/12/2004	0.00
05/01/2005	1.00			
20/01/2005	2.00			
		t_2	28/02/2005	1.00 + 2.00
15/03/2005	3.00			
		t_3	30/04/2005	3.00
		t_4	30/06/2005	0.00
30/08/2005	4.00	t_5	30/08/2005	4.00
		t_6	30/10/2005	10.00
15/12/2005	5.00			
25/12/2005	5.00			

Paso 2

Lo que interesa conocer no es el cashflow en t_i , sino el importe percibido en $t_{NumSteps}$ debido al cashflow desde t_0 hasta t_i . Llamamos a este concepto cashflow acumulado y lo calculamos de la forma siguiente:

$$C_k = \sum_{i=1}^k V_i \cdot \Upsilon_S(t_i, t_{NumSteps})$$

Notación. C_k^i = cashflow acumulado del cliente i en el nodo de tiempo k .

4.6.2. Mapeo del recovery**Paso 1**

Sea $t_0, t_1, t_2, \dots, t_{NumSteps}$ la partición de tiempo usada. Calculamos el recovery en el nodo, t_i , de la partición de la siguiente forma:

1. Si existe un recovery anterior, (t_a, W_a) , y otro posterior, (t_b, W_b) , entonces el valor del recovery en t_i es:

$$W_i = A + (B - A) \cdot \frac{t_i - t_a}{t_b - t_a}$$

donde $A = \Upsilon_S(t_a, t_i) \cdot W_a$ y $B = \Upsilon_S(t_b, t_i) \cdot W_b$.

2. Si no existe un recovery anterior a t_i entonces el valor del recovery en t_i :

$$W_i = 0$$

3. Si existe un recovery entre t_{i-1} y t_i , (t_a, W_a) , pero no existe un recovery posterior a t_i , entonces el valor del recovery en t_i es:

$$W_i = \frac{t_a - t_{i-1}}{t_i - t_{i-1}} \cdot W_a \cdot \Upsilon_S(t_a, t_i)$$

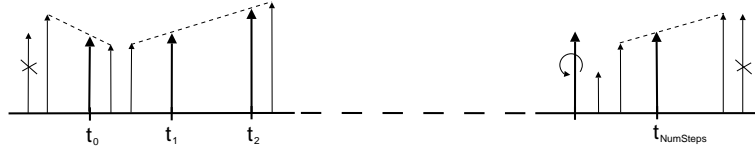


Figura 4.5: Recovery mapping

Veamos un ejemplo. Consideramos $\Upsilon_S(t_i, t_j) = 1$ para no complicar los cálculos. A la izquierda se representa el recovery original (sin estar mapeado a los nodos de tiempo). A la derecha se encuentra el recovery mapeado a los nodos de tiempo.

Date	Recovery	t_i	Date	Recovery
17/04/2004	-10.00	t_0	30/10/2004	$-10,00 + (1 + 10) \cdot \frac{196}{263} = -1,8$
		t_1	30/12/2004	$-10,00 + (1 + 10) \cdot \frac{196}{263} = 0,75$
05/01/2005	1.00			
20/01/2005	2.00			
		t_2	28/02/2005	$2 + (3 - 2) \cdot \frac{39}{54} = 2,72$
15/03/2005	3.00			
		t_3	30/04/2005	$3,00 + (4 - 3) \cdot \frac{46}{168} = 3,27$
		t_4	30/06/2005	$3,00 + (4 - 3) \cdot \frac{46}{168} = 3,64$
30/08/2005	4.00	t_5	30/08/2005	4,00
		t_6	30/10/2005	$4,00 + (5 - 4) \cdot \frac{61}{107} = 4,57$
15/12/2005	5.00			
25/12/2005	10.00			

Paso 2

Lo que interesa conocer no es el recovery en t_i , sino el importe percibido en $t_{NumSteps}$ debido al recovery realizado en t_i . Llamamos a este concepto recovery acumulado y lo calculamos de la forma siguiente:

$$N_k = W_k \cdot \Upsilon(t_k, t_{NumSteps})$$

Notación. N_k^i = recovery acumulado del cliente i en el nodo de tiempo k .

4.7. Paso 7. Simulación

Cada simulación consiste en generar números aleatorios mediante una o varias cópulas (dependiendo del método de simulación escogido), simular el tiempo de fallido de cada cliente (mapeándolos sobre $t_1, t_2, \dots, t_{NumSteps}$ y resto) y evaluar cada segmento de cada segmentación.

Una forma de acelerar la convergencia implementada en CreditCruncher, es la metodología *antithetic*. Esta consiste usar cada generación de números aleatorios (computacionalmente costosa de generar) 2 veces. Sea $(u_1, u_2, \dots, u_{NumClients})$ números aleatorios entre 0 y 1 generados por la realización de una cópula gaussiana. Entonces la realización $(1 - u_1, 1 - u_2, \dots, 1 - u_{NumClients})$ es equiprobable a la anterior.

4.7.1. Generación de números aleatorios

Método Rating-Path

Generamos una realización de cada cópula, $Q_1, Q_2, \dots, Q_{NumClient}$ usando los pasos 3-5 del apartado A.9.

Si se usa metodología *antithetic*, solamente se generan los números de las simulaciones pares. Los números aleatorios usados en las simulaciones impares son \bar{I} menos los de la simulación anterior

Método Time-To-Default

Generamos una realización de la cópula, Q usando los pasos 3-5 del apartado A.9.

Si se usa metodología *antithetic*, solamente se generan los números de las simulaciones pares. Los números aleatorios usados en las simulaciones impares son \bar{I} menos los de la simulación anterior

4.7.2. Simulación del tiempo de fallido

Método Rating-Path

Sean $Q_1(i), Q_2(i), \dots, Q_{NumSteps}(i)$ los componentes i -esimos de las cópulas simuladas. El objetivo es simular la evolución del rating del cliente i -esimo. Conocemos el rating del cliente i en t_0 . Realizamos el método expuesto en 3.4.1 hasta que el rating sea *Default* o el tiempo sea superior a $t_{NumSteps}$.

Método Time-To-Default

Sean $Q(i)$ el componente i -ésimo de las cópula simulada. El objetivo es simular el tiempo de fallido del cliente i -ésimo. Conocemos el rating del cliente i en t_0 . Realizamos el método expuesto en 3.4.2 teniendo en cuenta que el valor retornado por la inversa de la función de supervivencia es un número con decimales; mapeamos este valor (tiempo de fallido) al nodo de tiempo $(t_1, t_2, \dots, t_{NumSteps}, t_{NumSteps+1})$ mas cercano por la derecha. Por ejemplo, si el tiempo de fallido simulado es 3,2 y los nodos de tiempo son 0, 3, 6 y 9, el valor se mapeará a 6.

4.7.3. Valoración del segmento

Los pasos anteriores nos proporcionan el tiempo simulado de fallido para cada cliente, $t_w^1, t_w^2, \dots, t_w^{NumClients}$. Obtenemos el valor de cada segmento de la siguiente forma:

$$X_j = \sum_{i \in \Omega_j} \left\{ \begin{array}{ll} C_{NumSteps}^i & \text{if } t_w^i > t_{NumSteps} \\ C_{t_w^i-1}^i + N_{t_w^i}^i & \text{if } t_w^i \leq t_{NumSteps} \end{array} \right\}$$

$$Z_j = \sum_{i \in \Omega_j} \left\{ \begin{array}{ll} 0 & \text{if } w_i > t_{NumSteps} \\ C_{NumSteps}^i - C_{t_w^i-1}^i - N_{w_i}^i & \text{if } w_i \leq t_{NumSteps} \end{array} \right\}$$

donde X_j y Z_j son el valor y la pérdida del segmento j en tiempo $t_{NumSteps}$, Ω_j es el conjunto de clientes del segmento j , C_k^i y N_k^i son el cashflow y el recovery acumulado en tiempo k del cliente i para el segmento j .

4.8. Paso 8. Valoración del riesgo

Después de realizar N simulaciones se dispone de N valores para cada segmento. Para valorar el riesgo necesitamos que los valores simulados sean las pérdidas de la cartera (Z). Disponemos de z_1, z_2, \dots, z_N . Fijado un nivel de confianza α y un nivel de VAR de x , usamos las indicaciones de 3.6 para calcular el Expected Loss, la Desviación Estándar, el VAR y el TCE y sus correspondientes errores estándares.

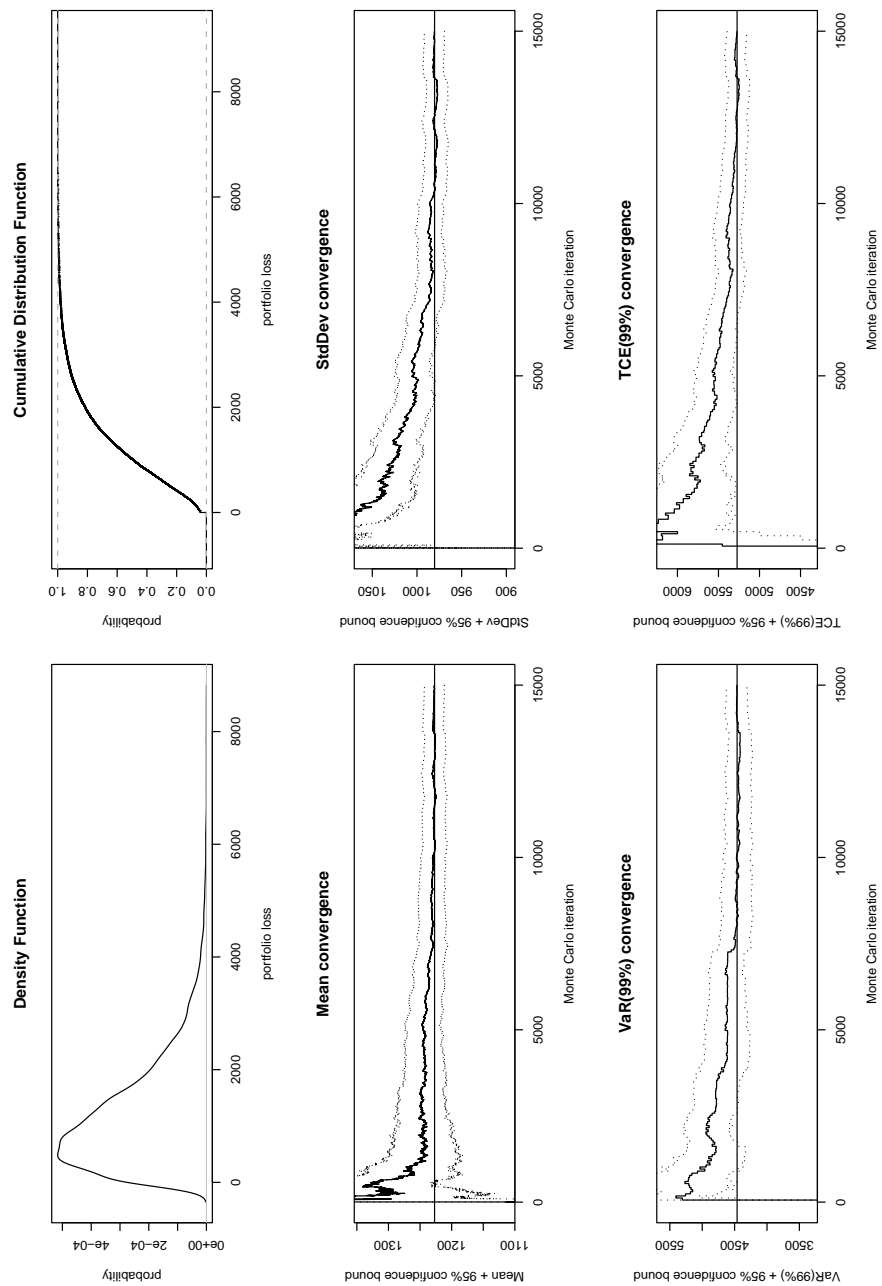


Figura 4.6: Evolución de los estadísticos de riesgo

Apéndice A

Apéndices

A.1. Conceptos básicos de estadística

Definición. Llamamos *función de distribución* o cdf de la variable aleatoria X a la función F que cumple:

$$F(x) = P(X \leq x)$$

Definición. Llamamos *función de probabilidad* o *densidad* o pdf de la variable aleatoria X a la función f que cumple:

$$F(x) = \int_{-\infty}^x f(t)dt$$

Proposición. Sea X una variable aleatoria continua con función de densidad $f_X(x)$. La función de densidad de $Y = H(X)$ siendo $H(\cdot)$ monótona (estrictamente creciente o estrictamente decreciente) es:

$$f_Y(y) = f_X(H^{-1}(y)) \cdot \left| \frac{d(H^{-1}(y))}{dy} \right|$$

Esperanza. Definimos la *esperanza* de una variable aleatoria discreta de la forma siguiente:

$$E(X) = \sum_i i \cdot P(X = i)$$

En el caso de una variable aleatoria continua con función de distribución $f(x)$ la esperanza se expresa como:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

Varianza. Definimos la *varianza* de una variable aleatoria discreta de la forma siguiente:

$$Var(X) = \sum_i (i - E(X))^2 \cdot P(X = i)$$

En el caso de una variable aleatoria continua con función de distribución $f(x)$ la varianza se expresa como:

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x) dx$$

Covarianza. Definimos la *covarianza* entre dos variables aleatorias X e Y de la forma siguiente:

$$Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Correlación. Definimos la *correlación*, ρ , entre dos variables aleatorias X e Y de la forma siguiente:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(x) \cdot Var(Y)}}$$

Proposición. Sea X una variable aleatoria continua con función de densidad $f(x)$ y $H(x)$ una función diferenciable. Entonces:

$$E(H(X)) = \int_{-\infty}^{\infty} H(x) \cdot f(x) dx$$

A.2. La variable aleatoria de Bernoulli

Definición. La variable aleatoria discreta *Bernoulli*, X , se utiliza para modelar fenómenos que solamente pueden tomar dos estados, 0 y 1, con probabilidades p y $(1-p)$ respectivamente. La notaremos como $X \sim Ber(p)$:

$$P(X = 0) = (1 - p) \quad P(X = 1) = p \quad p \in [0, 1]$$

Esperanza. La esperanza de una variable aleatoria Bernoulli $X \sim Ber(p)$ es p . Este resultado es la aplicación directa de la definición de esperanza para una variable aleatoria discreta:

$$E(X) = \sum_i i \cdot P(X = i) = 1 \cdot p + 0 \cdot (1 - p) = p$$

Varianza. La varianza de una variable aleatoria Bernoulli $X \sim Ber(p)$ es $p \cdot (1 - p)$. Este resultado es la aplicación directa de la definición de varianza para una variable aleatoria discreta:

$$Var(X) = \sum_i (i - E(X))^2 \cdot P(X = i) = (1 - p)^2 \cdot p + (-p)^2 \cdot (1 - p) = p \cdot (1 - p)$$

Simulación. La simulación de una variable Bernoulli $X \sim Ber(p)$ la realizamos de la siguiente forma:

$$x = \begin{cases} 0 & u \in [0, 1 - p) \\ 1 & u \in [1 - p, 1] \end{cases} \quad u \sim U[0, 1]$$

A.3. La variable aleatoria Binomial

Definición. La suma de n variables aleatorias Bernoulli, $Ber(p)$, independientes e idénticamente distribuidas es una variable aleatoria discreta, X , que llamamos *Binomial*, $X \sim B(n, p)$.

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \quad \binom{n}{k} = \frac{n!}{k! \cdot (n - k)!} \quad k \in \{0, \dots, n\}$$

Esperanza. La esperanza de una variable aleatoria Binomial $X \sim B(n, p)$ es:

$$E(X) = n \cdot p$$

Varianza. La varianza de una variable aleatoria Binomial $X \sim B(n, p)$ es:

$$Var(X) = n \cdot p \cdot (1 - p)$$

Proposición. El Teorema Central del Límite nos permite, en el caso de n grande, aproximar la distribución discreta Binomial por una distribución continua Normal:

$$B(n, p) \approx N(np, np(1 - p))$$

A.4. La variable aleatoria Normal

Definición. Decimos que una variable aleatoria continua Z es una *Normal* con media μ y desviación estándar σ , $Z \sim N(\mu, \sigma^2)$ si tiene la siguiente función de densidad:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Esperanza. La esperanza de una variable aleatoria Normal $X \sim N(\mu, \sigma^2)$ es:

$$E(X) = \mu$$

Varianza. La varianza de una variable aleatoria Normal $X \sim N(\mu, \sigma^2)$ es:

$$Var(X) = \sigma^2$$

Simulación. Para la generación de una realización, z , de una variable aleatoria normal $Z \sim N(\mu, \sigma^2)$ utilizamos el siguiente algoritmo:

$$z = \mu + \sigma \cdot \sqrt{-2\ln(u_1)} \cdot \cos(2\pi \cdot u_2) \quad u_1, u_2 \sim U[0, 1]$$

Definición. Decimos que una variable aleatoria continua n -dimensional, Z , es una *Normal* con media $\vec{\mu}$ y matriz de covarianza Σ , $Z \sim N(\vec{\mu}, \Sigma)$, si tiene la siguiente función de densidad:

$$f(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu})^\top \Sigma^{-1} (\vec{x} - \vec{\mu}) \right\}$$

donde $|\Sigma|$ es el determinante de la matriz de covarianzas Σ , y Σ^{-1} es la inversa de la matriz Σ .

Simulación. Para la generación de una realización, \vec{z} , de una variable aleatoria normal $Z \sim N(\vec{\mu}, \Sigma)$ utilizamos el siguiente algoritmo:

$$\vec{z} = \vec{\mu} + \Sigma^{1/2} \vec{x} \quad x_i \sim N(0, 1)$$

La matriz $\Sigma^{1/2}$ la calculamos usando el algoritmo de Cholesky. Sabemos que existe solución debido a que Σ es definida positiva por tratarse de una matriz de covarianzas.

A.5. Estimadores estadísticos

Sea $\{x_1, x_2, x_3, \dots, x_N\}$ N realizaciones de la variable aleatoria X .

Proposición. Estimamos la media de X usando el siguiente estimador no sesgado:

$$E(X) = \mu_X \approx \widehat{\mu_X} = \frac{1}{N} \sum_{i=1}^N x_i$$

Proposición. Estimamos la varianza de X usando el siguiente estimador no sesgado:

$$Var(X) = \sigma_X^2 \approx \widehat{\sigma_X^2} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \widehat{X})^2 = \frac{1}{N-1} \left(\sum_{i=1}^N x_i^2 - \frac{\left(\sum_{i=1}^N x_i \right)^2}{N} \right)$$

Proposición. Estimamos el VAR de X al nivel de confianza α usando el siguiente estimador:

$$VAR_\alpha(X) = \inf\{x | F_X(x) \geq \alpha\} \approx \widehat{VAR}_\alpha(X) = x_{k:N}$$

donde,

- F_X es la función de distribución de X
- k cumple $\frac{k}{N} \leq \alpha < \frac{k+1}{N}$
- $x_{k:N}$ es el k -ésimo elemento de la serie ordenada de forma creciente

Proposición. Estimamos el TCE (Tail Conditional Expectation o Expected Shortfall) de X al nivel de confianza α usando el siguiente estimador:

$$TCE_\alpha(X) = E(X | X > VAR_\alpha(X)) \approx \widehat{TCE}_\alpha(X) = \frac{\sum_{i=1}^N x_i \cdot 1_{\{x_i > \widehat{VAR}_\alpha(X)\}}}{\sum_{i=1}^N 1_{\{x_i > \widehat{VAR}_\alpha(X)\}}}$$

A.6. Error estándar de un cuantil (o VaR)

A continuación se describe el método de Maritz-Jarrett basado en el estadístico de orden para calcular el error estándar de un cuantil. Fórmula extraída de [4]. Sea $x_1, x_2, x_3, \dots, x_N$ la muestra sobre la que se desea calcular el error estándar del cuantil α .

$$\begin{aligned} M &= [N\alpha + 0,5] \\ a &= M - 1 \\ b &= N - M \\ W_i &= B(a, b, \frac{i+1}{N}) - B(a, b, \frac{i}{N}) \\ C_k &= \sum_{i=1}^N W_i \cdot x_i \end{aligned}$$

donde $[x]$ es la parte entera de x y $B(a, b, x)$ es la función beta incompleta:

$$B(a, b, x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1} dt$$

entonces,

$$\text{stderr}(q_\alpha) = \sqrt{C_2 - C_1^2}$$

A.7. Error estándar del TCE

El TCE (Tail Conditional Expectation o Expected Shortfall) es la esperanza de una variable aleatoria condicionada a que sea superior a un cierto cuantil. Dado un conjunto de observaciones $x_1, x_2, x_3, \dots, x_N$ de una variable aleatoria X , calculamos el error estándar del TCE siguiendo los siguientes pasos:

Paso 1. Calculamos el estimador de $VAR_\alpha(X)$.

Paso 2. Definimos la variable aleatoria $Y = \{X|X > VAR_\alpha(X)\}$. Generamos un conjunto de observaciones de Y seleccionando las observaciones de X que son superiores a $VAR_\alpha(X)$:

$$y_1, y_2, y_3, \dots, y_M$$

Paso 3. Usamos el teorema central del límite para estimar la media de Y y calcular su error estándar:

$$\begin{aligned} TCE_\alpha(X) &= \widehat{\mu_Y} \\ \text{stderr}(TCE_\alpha(X)) &= \frac{\widehat{\sigma_Y}}{\sqrt{M}} \end{aligned}$$

A.8. Cálculo de la raíz de una matriz

Definición. Diremos que 2 matrices A y B de orden n son *semejantes* si existe una matriz, P , de orden n con $\det(P) \neq 0$ tal que $B = P^{-1} \cdot A \cdot P$.

Proposición. Si dos matrices A y B son semejantes ($B = P^{-1} \cdot A \cdot P$) entonces:

$$\begin{aligned} \det(A) &= \det(B) \\ B^n &= P^{-1} \cdot A^n \cdot P \end{aligned}$$

Definición. Diremos que una matriz A de orden n es *diagonalizable* si es semejante a una matriz diagonal D , o sea, $A = P^{-1} \cdot D \cdot P$ siendo $\det(D) \neq 0$.

Proposición. Para que una matriz A sea diagonalizable es necesario y suficiente que:

- Los valores propios de A sean todos reales
- Los n vectores propios de A sean independientes

Proposición. Si una matriz A es diagonalizable ($A = P^{-1} \cdot D \cdot P$) entonces:

- D es una matriz diagonal compuesta por los valores propios de la matriz A
- P es la matriz formada por los vectores propios de la matriz A

Resultado. Sea A la raíz n -ésima de una matriz diagonalizable B . Entonces:

$$A^n = B = P^{-1} \cdot D \cdot P \implies A = \sqrt[n]{B} = P^{-1} \cdot \sqrt[n]{D} \cdot P$$

Observación. Si D es una matriz diagonal con $d_1, d_2, d_3, \dots, d_n$ en la diagonal, entonces la matriz D^x es una matriz diagonal y sus valores son $d_1^x, d_2^x, d_3^x, \dots, d_n^x$.

Algoritmo. Sea A una matriz diagonalizable. Obtenemos A^x , siendo $x > 0$, calculando los autovectores y autovalores de A :

$$\text{eigenvectors}(A) = P \quad \text{eigenvalues}(A) = D$$

donde D es una matriz con los valores propios en la diagonal y el resto de elementos iguales a 0. Entonces:

$$A^x = P^{-1} \cdot D^x \cdot P$$

Para calcular la inversa de una matriz se calcula la descomposición LU de dicha matriz y se resuelve el sistema $L \cdot U \cdot P^{-1} = Id$ elemento a elemento. El método LU y la inversa de una matriz usando la descomposición LU se encuentra explicada con detalle en el libro *Numerical Recipes in C*¹.

A.9. Algoritmo de la cópula gaussiana

Sea Σ_1 la matriz de correlación a cumplir por la cópula. Observemos que se trata también de la matriz de covarianzas al valer 1 los elementos de la diagonal.

$$\Sigma_1 = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{pmatrix}$$

¹<http://www.nr.com>

Paso 1. Creamos la matriz de covarianzas Σ_2 transformando la matriz Σ_1 componente a componente:

$$\Sigma_2 = \begin{pmatrix} 2\sin(\frac{\pi}{6}) & 2\sin(\rho_{12}\frac{\pi}{6}) & \dots & 2\sin(\rho_{1n}\frac{\pi}{6}) \\ 2\sin(\rho_{21}\frac{\pi}{6}) & 2\sin(\frac{\pi}{6}) & \dots & 2\sin(\rho_{2n}\frac{\pi}{6}) \\ \vdots & \vdots & \ddots & \vdots \\ 2\sin(\rho_{n1}\frac{\pi}{6}) & 2\sin(\rho_{n2}\frac{\pi}{6}) & \dots & 2\sin(\frac{\pi}{6}) \end{pmatrix}$$

Observamos que la matriz Σ_2 vuelve a tener los elementos de la diagonal iguales a 1 debido a que $2\sin(\frac{\pi}{6}) = 1$.

Paso 2. A la matriz Σ_2 le aplicamos el algoritmo de Cholesky para obtener la matriz triangular inferior B cumpliendo $\Sigma_2 = B \cdot B^\top$:

$$B = \begin{pmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix}$$

Paso 3. Simulamos n variables aleatorias $N(0, 1)$ independientes ²:

$$\vec{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad Y_k \sim N(0, 1) \text{ independientes}$$

Paso 4. Simulamos una variable n-dimensional $\vec{Z} \sim N(\vec{0}, \Sigma_2)$ haciendo:

$$\vec{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = \begin{pmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = B \cdot \vec{Y}$$

Paso 5. Finalmente obtenemos la simulación de la cópula, \vec{X} .

$$\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \Phi(Z_1) \\ \vdots \\ \Phi(Z_n) \end{pmatrix}$$

donde $\Phi(x)$ es la función de distribución de la Normal estándar

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

²El algoritmo de simulación de una variable aleatoria Normal(0,1) se encuentra explicado en el apartado A.4

A.10. Descomposición de Cholesky de una matriz en bloques

Dada una matriz cuadrada, simétrica y definida positiva, A , el algoritmo de Cholesky realiza la siguiente descomposición:

$$U^T \cdot U = A$$

donde U es una matriz triangular superior (por tanto, U^T es una matriz triangular inferior). Una descripción e implementación del algoritmo puede encontrarse en *Numerical Recipes in C*³.

Si intentamos realizar la descomposición de Cholesky de una matriz de correlación entre clientes de una cartera de 50,000 clientes nos encontraremos con problemas de tamaño (la matriz ocupará 19 Gb. de memoria) y número de operaciones al multiplicar la matriz por un vector (2,500,000,000 multiplicaciones).

Modificaremos el algoritmo de Cholesky para aprovechar el hecho que la matriz de correlación entre clientes es una matriz en bloques con 1's en la diagonal. Veamos un ejemplo de una cartera de 7 clientes con dos sectores:

$$A = \left(\begin{array}{cccc|ccc} 1 & 0,5 & 0,5 & 0,5 & 0,1 & 0,1 & 0,1 \\ 0,5 & 1 & 0,5 & 0,5 & 0,1 & 0,1 & 0,1 \\ 0,5 & 0,5 & 1 & 0,5 & 0,1 & 0,1 & 0,1 \\ 0,5 & 0,5 & 0,5 & 1 & 0,1 & 0,1 & 0,1 \\ \hline 0,1 & 0,1 & 0,1 & 0,1 & 1 & 0,3 & 0,3 \\ 0,1 & 0,1 & 0,1 & 0,1 & 0,3 & 1 & 0,3 \\ 0,1 & 0,1 & 0,1 & 0,1 & 0,3 & 0,3 & 1 \end{array} \right)$$

Realizamos la descomposición de Cholesky:

$$U = \left(\begin{array}{cccc|ccc} 1,00000 & 0,50000 & 0,50000 & 0,50000 & 0,10000 & 0,10000 & 0,10000 \\ 0 & 0,86603 & 0,28868 & 0,28868 & 0,05774 & 0,05774 & 0,05774 \\ 0 & 0 & 0,81650 & 0,20412 & 0,04082 & 0,04082 & 0,04082 \\ 0 & 0 & 0 & 0,79057 & 0,03162 & 0,03162 & 0,03162 \\ \hline 0 & 0 & 0 & 0 & 0,99197 & 0,28630 & 0,28630 \\ 0 & 0 & 0 & 0 & 0 & 0,94975 & 0,21272 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,92563 \end{array} \right)$$

³<http://www.nr.com>

Observamos que U contiene elementos repetidos. Guardaremos la matriz U en memoria de la forma siguiente:

$$U = \begin{vmatrix} 1,00000 & 0,50000 & 0,10000 \\ 0,86603 & 0,28868 & 0,05774 \\ 0,81650 & 0,20412 & 0,04082 \\ 0,79057 & 0 & 0,03162 \\ 0,99197 & 0 & 0,28630 \\ 0,94975 & 0 & 0,21272 \\ 0,92563 & 0 & 0 \end{vmatrix}$$

o sea, para cada fila guardamos el valor de la diagonal y el valor de cada sector. El tamaño en memoria ahora pasa a ser $N \times (M + 1)$ donde N es el número de clientes y M el número de sectores.

Para reducir el número de operaciones realizadas al multiplicar la matriz por un vector aprovechamos que la matriz U tiene elementos repetidos. Veamos un ejemplo:

$$\begin{aligned} (U \cdot x)_2 &= 0,86603 \cdot x_2 + 0,28868 \cdot x_3 + 0,28868 \cdot x_4 + \\ &0,05774 \cdot x_5 + 0,05774 \cdot x_6 + 0,05774 \cdot x_7 \\ &= 0,86603 \cdot x_2 + 2 \cdot 0,28868 \cdot (x_3 + x_4) + 3 \cdot 0,05774 \cdot (x_5 + x_6 + x_7) \end{aligned}$$

Con estas dos consideraciones se obtiene una algoritmo de Cholesky para matrices en bloques con uso de memoria y número de operaciones (al multiplicar por un vector) del orden $N \times (M + 1)$ en vez de orden N^2 . Con el nuevo algoritmo la memoria necesaria para almacenar la matriz de 50,000 clientes es de 4,2 Mb. y el número de multiplicaciones se reduce a 500,000.

Bibliografía

- [1] Dirk Tasche Carlo Acerbi. Expected shortfall: a natural coherent alternative to value at risk. *BIS*, 2001.
- [2] Paul Glasserman. Probability models of credit risk. *Columbia Business School*, 2000.
- [3] Greg M. Gupton, Christopher C. Finger, and Mickey Bhatia. *Credit-Metrics - Technical Document*. J.P. Morgan & Co. Incorporated, 1997.
- [4] Jonathan E. Grindlay Jaesub Hong, Eric M. Schlegel. New spectral classification technique for x-ray sources: quantile analysis. 2004.
- [5] Stefan R. Jaschke. Quantile-var is the wrong measure to quantify market risk for regulatory purposes. *BIS*, 2001.
- [6] Philippe Jorion. *Value at Risk*. McGraw-Hill, 1997.
- [7] Mervyn Marasinghe. Monte carlo methods. Class notes for Iowa State University, Dept. of Statistics.
- [8] Alexander McNeil Paul Embrechts and Daniel Straumann. Correlation and dependence in risk management: properties and pitfalls. *RiskLab*, 1999.
- [9] Ken Phelan and Colin Alexander. Different strokes. *Risk*, 1999.
- [10] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0.
- [11] Murray R. Spiegel. *Estadística*. Schaum, 1997.
- [12] Shaun S. Wang. Aggregation of correlated risk portfolios: Models & algorithms. *CAS Committee on Theory of Risk*.

Índice alfabético

- Algoritmo de Cholesky, 48, 49
- Algoritmo LU, 47
- Antithetic, 38
- Bernouilli, 42
- Binomial, 43
- Cópula, 24
- Cópula gaussiana, 18, 47
- Capital económico, 19
- Cashflow, 9
- Correlación, 42
- Covarianza, 42
- Cumulated Forward Default Rate, 15
- Curva cupón cero, 12
- Curva spot, 12
- Densidad, 41
- Desviación Estándar, 28
- Distribuciones marginales, 24
- EAD, 9
- Error estándar, 45, 46
- Esperanza, 41
- Expected Loss, 19, 28
- Expected Shortfall, 19, 28, 45
- Función de distribución, 41
- Función de distribución empírica, 23
- Función de probabilidad, 41
- Función de transporte, 11
- LGD, 9
- Método de Monte Carlo, 23
- Método Rating-Path, 25
- Método Time-To-Default, 26
- Matrices semejantes, 46
- Matriz de correlación entre clientes, 17
- Matriz de transición, 13
- Matriz diagonalizable, 46
- Matriz en bloques, 18, 31, 49
- Normal, 43, 44
- Recovery, 9
- Sectores, 8
- Segmentación, 33
- Segmento, 33
- Sistema de ratings, 7
- Supervivencia, 15
- Tabla de correlaciones sectoriales, 17
- Tail Conditional Expectation, 19, 28, 45, 46
- Teorema de Sklar, 24
- Tipo de interés compuesto, 10
- Tipo de interés continuo, 11
- Tipo de interés efectivo, 10
- Tipo de interés simple, 10
- Value At Risk, 19, 28, 45
- Varianza, 42