



UNIVERSIDAD NACIONAL DE COLOMBIA

Modelo estructural de riesgo de crédito con intensidad estocástica de covariables observables y un factor de fragilidad determinado a partir de un proceso de saltos

Luis Alberto Bernal Berrio

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2019

Modelo estructural de riesgo de crédito con intensidad estocástica de covariables observables y un factor de fragilidad determinado a partir de un proceso de saltos

Luis Alberto Bernal Berrio

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magister en Ciencias Estadística

Director(a):
Ph.D., César Augusto Gómez Vélez

Línea de Investigación:
Finanzas y riesgos
Grupo de Investigación:
Estadística

Universidad Nacional de Colombia
Facultad, Escuela de Estadística
Medellín, Colombia
2019

(Dedicatoria)

A Dios primeramente, a mi familia y a todos los que me han brindado su apoyo incondicional en este proceso.

Agradecimientos

Un especial agradecimiento al profesor César Gómez por su dedicación y acompañamiento en este proceso tan enriquecedor. Agradezco a los jurados evaluadores, profesores Norman Diego Giraldo y Jorge Ramírez por su retroalimentación e importantes aportes a la realización de este trabajo. Agradezco a mis grandes amigos Santiago y Juan por su apoyo. Agradezco a la entidad financiera que facilitó la información para la realización de este trabajo y a la Universidad Nacional de Colombia.

Resumen

En este trabajo se estiman los parámetros para la intensidad de default de covariables observables en presencia de un factor de fragilidad no observable. La información observable corresponde a la evolución de algunas variables macroeconómicas en el tiempo, así como la información característica de individuos de un segmento de crédito en una entidad financiera colombiana; se realiza una pequeña modificación al proceso de Cox propuesto para la intensidad en Duffie et al. (2009), con el fin de incluir una componente de saltos a partir de la cual se busca describir los agrupamientos espontáneos de defaults, finalmente se implementa un programa para estimar los parámetros asociados al proceso para la intensidad por medio del algoritmo EM y el muestreador de Gibbs.

Palabras clave: Intensidad de default, Proceso de Cox, Fragilidad, Algoritmo EM, Muestreador de Gibbs.

Abstract

In this work, the parameters for the default intensity of observable covariates in the presence of an unobservable fragility factor are estimated. The observable information corresponds to the evolution of some macroeconomic variables over time, as well as the characteristic information of the individuals of a credit segment in a Colombian financial entity; a small modification to the Cox process proposed for intensity is made in Duffie et al. (2009), in order to include a jump component by means of which it is sought to describe the spontaneous clusters defaults, a program is finally implemented to estimate the parameters associated to the process for intensity by means of the EM algorithm and the Gibbs sampler.

Keywords: Default Intensity, Cox Process, Fragility, EM Algorithm, Gibbs Sampler

Contenido

Agradecimientos	IV
Resumen	V
Lista de figuras	VIII
Lista de tablas	XI
1. Introducción	1
1.1. Formulación del problema	2
1.2. Objetivo General	3
1.3. Objetivos Específicos	3
1.4. Descripción de los datos.	4
1.5. Especificación del modelo	4
1.6. Estado del arte	6
2. Marco teórico	9
2.1. Procesos estocásticos	9
2.1.1. Movimiento Browniano	11
2.1.2. Proceso Ornstein–Uhlenbeck	11
2.1.3. Proceso Ornstein-Uhlenbeck con saltos	13
2.2. Proceso Poisson	14
2.2.1. Proceso de Poisson no homogéneo	15
2.2.2. Proceso de Cox	15
2.2.3. Procesos de Poisson compuestos	16
2.3. Métodos de estimación	17
2.3.1. Función de verosimilitud	17
2.3.2. Resultados asintóticos	17
2.4. Cadenas de Markov Montecarlo	18
2.4.1. Cadenas de Markov	18
2.4.2. Ley de grandes números	21
2.4.3. Algoritmo Metropolis Hasting	22
2.4.4. Muestreador de GIBBS	23
2.4.5. Algoritmo de la Esperanza Maximización (Expectation Maximization)	26

2.4.6.	Enfriamiento simulado	27
2.5.	Procesos vectoriales autorregresivos	29
2.5.1.	Procesos autorregresivos de orden 1	29
2.5.2.	Estimación del proceso vectorial AR	30
3.	Formulación del Modelo y Metodología de estimación	31
3.1.	Proceso de COX o Poisson doblemente estocástico	32
3.2.	Proceso de Estimación	34
3.2.1.	Planteamiento de la Función de Verosimilitud	34
3.2.2.	Esperanza Maximización	38
3.2.3.	Muestreador de Gibbs para generar trayectorias del proceso de fragilidad.	39
3.2.4.	Modelo VAR para las covariables observables	43
4.	Implementación en una entidad financiera colombiana	44
4.1.	Descripción y tratamiento de los datos	45
4.2.	Estimación de ML para el modelo con fragilidad	47
4.3.	Estimación de ML para el modelo sin fragilidad	50
4.4.	Modelo con fragilidad vs modelo sin fragilidad	50
5.	Conclusiones y recomendaciones	52
5.1.	Conclusiones	52
5.2.	Recomendaciones	53
A.	Anexo: Ajuste VAR(1) covariables	55
A.0.1.	Resultados del ajuste VAR(1)	55
B.	Anexo: Breve descripción metodologías de riesgo de crédito	58
B.1.	Modelo de Merton	58
B.2.	Modelo KMV	59
B.3.	Modelos de intensidad	60
B.4.	Modelo Credit Risk Metrics	61
B.5.	Modelos estructurales	61
B.5.1.	Modelo de Cox Ingersoll Ross	61
B.5.2.	Modelo de tres factores	62
B.5.3.	Duffie and Garleanu (2001)	62
B.5.4.	Cópulas	63
C.	Anexo: Programa implementado para el algoritmo EM	64
	Bibliografía	74

Lista de Figuras

2-1. Trayectorias de un proceso estocástico. Tomado de Douglas Rivas (2019)	10
2-2. Una trayectoria del proceso Ornstein Uhlenbeck Y_t , con $Y_0 = 0, \kappa = 2$	13
2-3. Una trayectoria del proceso Ornstein Uhlenbeck Y_t , con $Y_0 = 0, \kappa = 2, \lambda = 2,$ $\sigma = 1$	14
4-1. Aparte resumen base de datos entidad financiera	45
4-2. Número de Defaults entre agosto de 2016 y junio de 2019	47
4-3. Ajuste estimación tasa de intensidad con fragilidad. Elaboración propia.	49
A-1. Ajuste VAR(1) para el IPC	56
A-2. Ajuste VAR(1) para el PIB	56
A-3. Ajuste VAR(1) para la Distancia Default	57
A-4. Ajuste VAR(1) para el logaritmo de los activos	57
B-1. Variables que determinan la probabilidad de default. Tomado de Moodys KMV, Pag 13.	60
B-2. Trayectoria del modelo estructural de difusión de Duffie y Garleanu (2001)	62

Lista de Tablas

- 4-1. Estimadores de máxima verosimilitud para la intensidad con el factor de fragilidad 48
- 4-2. Estimadores de máxima verosimilitud para la intensidad sin el factor de fragilidad . 50

1. Introducción

El crédito ocupa un papel integral en el funcionamiento de un sistema económico organizado, los agentes o individuos con excedentes de liquidez, están en la capacidad de transferirlos a individuos con déficit de estos recursos a cambio de un precio (tasas de interés). Las personas, empresas y gobiernos adquieren obligaciones crediticias con el fin de financiar el gasto público, proyectos de inversión, actividades productivas y de consumo. Sin embargo, los cambios de carácter social, demográfico, ambiental, o político tienen una influencia permanente en la economía, lo que se materializa en una mayor incertidumbre sobre la capacidad de pago de los diferentes agentes prestatarios que están sujetos a cambios en su situación financiera, McNeil et al. (2005).

La tasa de interés refleja la prima por el riesgo de crédito que asumen los agentes prestamistas con excedentes de liquidez; en los últimos años se ha incrementado la relevancia de la investigación en el riesgo de incumplimiento crediticio o *default* debido a la evidencia de su impacto sobre la economía y el bienestar de los hogares. Un fenómeno que motiva el estudio de la dinámica o intensidad del default consiste en los clústeres de incumplimiento crediticio, es decir, en incrementos súbitos de la frecuencia o número de incumplimientos. La evidencia empírica ha mostrado que cuando se presentan fenómenos de desaceleración o recesión económica, se empiezan a presentar incumplimientos que se pueden generalizar provocando pánico financiero, lo que técnicamente se conoce como *contagio default*, fenómeno presente en cada una de las debacles financieras de los últimos años, Spiliopoulos (2015).

La complejidad en la valoración de instrumentos financieros de deuda, no sólo se presenta en productos estructurados de deuda en los mercados de capitales, sino también en la cartera representada por personas o empresas, en cualquier segmento o línea de crédito, ofrecida por ejemplo, por entidades financieras y corporaciones de ahorro y vivienda. La percepción que tienen los inversionistas o entidades financieras del valor y el riesgo de estos instrumentos está fundamentada en gran medida en el concepto o calificación o score crediticio que emiten las diferentes agencias calificadoras de riesgo, en cuánto la calidad y capacidad de pago de los agentes prestatarios. Por su parte los entes de control velan por regular y exigir permanentemente el cumplimiento de las reservas mínimas de capital que deben tener las diferentes entidades financieras, fondos de pensiones y aseguradoras, con el fin de garantizar un fondo de reserva que les permita afrontar una eventual crisis o catástrofe financiera.

Este trabajo tiene como objetivo realizar una pequeña extensión al modelo propuesto en Duffie et al. (2009). El modelo descrito por los autores busca capturar la dinámica con la cual se presenta en el tiempo el surgimiento de incumplimientos de pago o *defaults*; en este trabajo además se identifican aspectos o fenómenos que algunos modelos de riesgo de crédito utilizados en la práctica no capturan y que llevó en su momento a subestimar el riesgo de crédito, algo común en diferentes debacles financieras, entre otras, como las generadas por la *burbuja inmobiliaria* en Colombia (1997) y la *crisis subprime* de EEUU (2008), además de muchos otros casos presentados recientemente, como la crisis europea en (2010).

En el modelo estructural de riesgo crediticio propuesto en Duffie et al. (2009), se estiman los parámetros asociados a la intensidad con la cual se presentan incumplimientos crediticios, allí esencialmente el proceso asociado a la intensidad es afectado por factores de riesgo observables y un factor de riesgo latente (no observable) denominado *fragilidad*, el cual es común al conjunto de préstamos. Se buscará modificar ligeramente la dinámica o especificación del proceso asociado al factor de fragilidad mediante un proceso de difusión con saltos.

Con el fin de determinar la replicabilidad del modelo, se abordan la estimación y ajuste de la intensidad para los eventos de incumplimiento crediticio en una entidad financiera colombiana; también se implementa un programa experimental para estimar los parámetros del proceso por medio del **algoritmo EM** y el **muestreador de GIBBS**, el cual es modificado con el fin de incluir una componente de saltos. Luego de obtener criterio experto de diferentes entidades financieras se establecen las covariables observables en el modelo para la intensidad, teniendo en cuenta el contexto de los individuos considerados en la línea de crédito del análisis.

1.1. Formulación del problema

Para algunas entidades financieras, es común encontrar dificultades en la gestión de la calidad de la cartera de crédito, debido a que trabajan por practicidad con modelos para los que se establece una variedad de supuestos que no replican el comportamiento de pago de los individuos. De otra parte, la información recolectada es susceptible de sesgos o asimetría, lo que induce a un mayor error de estimación de las probabilidades de incumplimiento. La hipótesis planteada, consiste en indagar si la información no observable tiene una incidencia o impacto significativo sobre la dinámica de los incumplimientos en el tiempo. En diferentes acontecimientos y crisis financieras se ha observado que los defaults se presentan en clusters o agrupamientos, sin que los gestores de riesgo de entidades financieras, calificadoras de riesgo y otros agentes, hayan emitido las alertas necesarias, ni tomado las medidas de riesgo pertinentes para evitar deterioros súbitos en la calidad de la cartera.

Debido a la complejidad en la valoración de instrumentos de deuda, se trabaja con modelos que

quizá no capturan todos los elementos representativos que permitan mejorar la estimación de probabilidades de incumplimiento. Cabe entonces mencionar algunos supuestos que se pueden flexibilizar con el fin de tener estimaciones más precisas o ajustadas al comportamiento de pago de un individuo:

- Incorporar una medida de dependencia que no solo esté asociada al comportamiento de un factor de riesgo macroeconómico común a todas las obligaciones, sino también a otros factores de riesgo no observados, Spiliopoulos (2015).
- La existencia de un factor de fragilidad no observado que representa el efecto remanente no explicado por factores observables sobre la intensidad de default de una compañía, Duffie et al. (2009).
- Capturar la dinámica con la cual se presentan agrupamientos o clusters de defaults que finalmente pueden inducir un contagio default, Spiliopoulos (2015).

Las situaciones que acaban de ser planteadas y a ser contempladas en la modelación, son temas que se consideran en cualquier departamento de riesgo crediticio, y van en línea con la gestión de la cartera de cualquier segmento de crédito, con el fin de controlar los cupos y asignar las tasas de interés que permitan compensar el riesgo de colocación, finalmente el objetivo es contar con una valoración razonable de los riesgos al momento de otorgar créditos y ofrecer aumento en los cupos de endeudamiento.

1.2. Objetivo General

Utilizar la metodología propuesta en Duffie et al. (2009), para estimar los parámetros del proceso de Cox para la intensidad de default estocástica de covariables observables y un factor de fragilidad no observado para el cual se incorpora una ligera modificación por medio de un proceso de saltos.

1.3. Objetivos Específicos

- Realizar una pequeña modificación en un modelo utilizado en la literatura para incorporar un componente de saltos.
- Emplear un algoritmo adecuado para estimar los parámetros de la intensidad de default y realizar simulaciones con el modelo calibrado.
- Ajustar el modelo a datos de incumplimiento provenientes de una cartera de crédito de un segmento de personas naturales en una entidad financiera colombiana.

1.4. Descripción de los datos.

Los datos sobre los cuales se ha trabajado corresponden a incumplimientos crediticios interanuales en un segmento de cartera masiva en una institución financiera colombiana. Los datos corresponden al nivel de activos y nivel de deuda para 5.250 individuos registrados entre agosto de 2016 y junio de 2019. La información correspondiente a los factores de riesgo asociados al comportamiento de la economía colombiana son representados por el PIB y las tasas de interés con resolución mensual. Se incorpora además el logaritmo de los activos de cada individuo y un indicador de distancia al default, el cual, mide el número de desviaciones estándar por las cuales los activos del individuo superan los pasivos en un horizonte de tiempo, para ver con mayor detalle la definición de distancia a default, ver (Apéndice A) y página 660 de Duffie et al. (2007).

La fuente de los datos de incumplimientos crediticios y distancia a default de cada compañía en el período de análisis corresponde principalmente a información provista por una entidad financiera reconocida, los datos de las variables macroeconómicas se obtuvieron a partir del DANE y el Banco de la República.

1.5. Especificación del modelo

El instante de incumplimiento del i -ésimo individuo en una cohorte de m prestatarios es modelado como el instante del primer salto o evento de un proceso de Poisson $N_i(t)$ con tasa $\lambda_{i,t}$ que a su vez constituye un proceso de Markov, especificado de la siguiente manera:

$$\lambda_{i,t} = \exp(\beta_0 + \beta_1 V_t + \beta_2 U_{i,t} + \eta Y_t). \quad (1-1)$$

Dónde

V_t es un factor común (macroeconómico) a todos los individuos.

$U_{i,t}$ Es un factor específico del i -ésimo individuo.

Y_t Representa un factor **no observable** o **latente**, común a todo individuo.

El modelo consiste en el Proceso de Poisson $N(t) = \sum N_i(t)$ que resulta de la superposición de Procesos de Poisson con intensidades estocásticas. Dicho de otra forma, condicional en las trayectorias de los procesos $(V_t, U_{i,t}, Y_t)$ el proceso de Poisson $N_i(t)$ es un proceso de Poisson no homogéneo con tasa $\lambda_{i,t}$. Sin embargo, cuando el proceso presenta una intensidad estocástica como en (1-1) se habla de un “**Proceso de Poisson doblemente estocástico**” o “**Proceso de Cox**”, Snyder and Miller (1991).

Los procesos V_t y $U_{i,t}$ son modelados como procesos autorregresivos de orden 1 **gaussianos**, con un vector de parámetros γ que determina su dinámica, para adaptarlos en tiempo continuo, simplemente se interpola entre los periodos, por lo que conservan la estructura de Markov.

El principal objetivo del trabajo consiste en el ajuste y estimación de parámetros, del proceso de Poisson agregado $N(t)$ que resulta de considerar una cohorte de prestatarios. De manera muy particular se abordará el modelo propuesto en Duffie et al. (2009), el cual en lo sucesivo se denominará el modelo **D**, que se caracteriza porque el proceso Y_t en (1-1) consiste en un proceso de Ornstein-Uhlenbeck (en adelante **OU**) de la siguiente forma:

$$dY_t = -\kappa Y_t + dB_t, \quad Y_0 = 0. \quad (1-2)$$

Donde, κ es un parámetro constante y B_t es un movimiento Browniano estándar.

Se pretende realizar una adaptación de la metodología propuesta por los autores en Duffie et al. (2009), para incluir una componente de saltos en el proceso (**OU**). Así en lugar de considerar el proceso Y_t de (1-2) como factor de riesgo en (1-1), se considerará el siguiente proceso:

$$dY_t = -\kappa Y_t + dB_t + dJ_{t-},$$

$$J_t = \sum_{i=1}^{N_t} \xi_i. \quad (1-3)$$

En 1-3 el proceso N_t representa un proceso de Poisson simple con intensidad constante λ_ξ . $\{\xi_i\}$, es una sucesión iid¹ de variables normales $N(0, \nu^2)$ independientes del Movimiento Browniano B_t , y dJ_{t-} denota el proceso:

$$dJ_{t-} = J_t - J_{t-} = \begin{cases} 0 & \text{si } N_t - N_{t-} = 0 \\ \xi_i & \text{si } N_t - N_{t-} = 1 \end{cases} \quad (1-4)$$

Dicho de otra forma J_t es un proceso de Poisson compuesto y la distribución de los saltos es normal $N(0, \nu^2)$. La idea es que el nivel del proceso **OU** en (1-3) es perturbado en tiempos exponencialmente distribuidos por pulsos ξ_i normalmente distribuidos.

¹independientes e idénticamente distribuidas

1.6. Estado del arte

El modelamiento del riesgo crediticio por su naturaleza, continúa siendo un reto en el campo de finanzas cuantitativas, debido a que no se cuenta con información pública disponible y suficiente que permita estimar medidas de riesgo crediticio asociadas a una empresa o individuo. La evidencia empírica ha mostrado que la distribución de pérdida asociada a los incumplimientos asumidos por un prestamista con frecuencia es asimétrica, con defaults en clusters, y valores extremos, McNeil et al. (2005).

En los últimos años se han presentado catástrofes financieras con consecuencias desastrosas que han tenido origen en la mayor incertidumbre y volatilidad experimentadas en el mercado de deuda y que superan significativamente la incertidumbre y volatilidad realizadas en los mercados de renta variable.

Los modelos dinámicos de riesgo crediticio son construídos bajo el supuesto de que una compañía entra en default cuando su nivel de activos cae por debajo de un umbral determinado por su nivel de deuda. Black and Scholes (1973), Edwin O. Fischer (1989) y Leland (1994) consideraron además el movimiento Browniano geométrico para modelar el proceso que gobierna el comportamiento de los activos de una compañía.

Algunas investigaciones en finanzas cuantitativas han llevado a considerar modelos estructurales con tasas de interés estocásticas y procesos de saltos para modelar la evolución de los precios de los activos como en Zhou (2001), umbrales de default variables en el tiempo Hull et al. (2004) o modelos de volatilidad estocástica como en Overbeck and Schmidt (2003). Modelos en tiempo discreto han sido ampliamente investigados por Cox and Rubinstein (1985), Jarrow et al. (1995) y Föllmer and Schied (2004). Por su parte modelos en tiempo continuo han sido desarrollados por Baxter and Rennie (1997), Bingham (2007), Bjork (2009), Duffie (2010) y Lamberton and Lapeyre (2011); otros desarrollos teóricos destacados han sido derivados del aporte de investigadores como Karatzas and Shreve (1998), Musiela and Rutkowski (2006), entre otros.

Grandes entidades no gubernamentales a nivel mundial también han contribuído al desarrollo de este campo, como el modelo KMV de la Moody's, donde se da un tratamiento a la medición del riesgo de crédito en empresas medianas o pequeñas que no cotizan en bolsa, (un mayor detalle se puede encontrar en Gregoriou (2006)), así mismo en 1997 el Credit Suisse Financial Products y McKinsey presentaron *CreditRisk+*, McNeil et al. (2005). En los últimos años se han incorporado modelos de clasificación que utilizan la inteligencia artificial y máquinas de soporte vectorial (SVM) con el fin de predecir la calificación crediticia de una obligación, ver Ahn et al. (2011) y Atzmueller et al. (2016). Sin embargo, la principal desventaja es que estos modelos no reaccionan rápidamente a cambios repentinos en la economía y en la situación financiera de empresas e individuos, McNeil et al. (2005).

Otra clase de modelos de riesgo de crédito ampliamente utilizados son los modelos de migración de rating o scoring crediticio. El modelo estándar en la industria es el propuesto por la Risk Metrics Group como métrica y el cual fue desarrollado conjuntamente con la JP Morgan Company (1996). Estos rating son emitidos por agencias calificadoras de riesgo como la Moody's, la Standard and Poor's o la Fitch. Es usual encontrar este sistema de calificación crediticia en entidades financieras, que por ley deben cuantificar los riesgos y reportar la calidad de sus carteras a partir de la categorización de estas en diferentes niveles de riesgo. La modelación de estos scores crediticios se realiza con el fin de estimar las probabilidades de transición de una calificación a otra, por su parte estas probabilidades dependen de la información histórica de los defaults en las diferentes carteras administradas, McNeil et al. (2005).

En el último lustro se destacan numerosos aportes teóricos por parte de reconocidos investigadores que presentan aproximaciones analíticas para modelar la dinámica de un evento de default. En Bielecki et al. (2014), además de modelar la intensidad y la tasa de recuperación del capital, se agregan estructuras de dependencia para las diferentes obligaciones de un portafolio usando la cópula de Markov. De otro lado Spiliopoulos (2015), adicional a la incorporación de un proceso estocástico para la intensidad, considera otros hechos estilizados como una estructura de correlación para las diferentes obligaciones de un portafolio, para el cual se tiene un crecimiento indefinido del número de nuevos prestatarios; como ocurre normalmente en cualquier línea de crédito de una institución financiera.

En el **anexo B** de este trabajo se describen brevemente algunas de las metodologías de riesgo de crédito mencionadas en este capítulo.

En adelante el trabajo se encuentra estructurado de la siguiente forma: En el capítulo 2 se esbozarán algunos elementos estadísticos y matemáticos que componen los fundamentos teóricos presentados en el problema y la idea de solución, en el capítulo 3 se presenta la adaptación propuesta sobre el artículo de Duffie et al. (2009) objeto de este trabajo y se explica de forma más detallada la metodología de estimación.

De acuerdo con el planteamiento del problema se buscó abordar la replicabilidad del modelo en el contexto de una línea de crédito de una entidad financiera colombiana (capítulo 4). En este punto se hizo un esfuerzo por conseguir datos provenientes de una entidad financiera, con el fin de probar la utilidad en la aplicación de la extensión o idea de solución al modelo estructural de riesgo propuesto por Duffie, en el contexto de una entidad financiera colombiana.

A partir de la información recolectada sobre individuos de una línea de crédito, se estimó la distancia a default, imputando los datos faltantes con el fin de tener información consistente de insumo

para el modelo. Para realizar el ajuste se utilizó el software estadístico R y se programó la función de verosimilitud, posteriormente se programó el algoritmo EM utilizado para la estimación de los parámetros del modelo, realizando una ligera modificación a la metodología de estimación propuesta por Duffie para el paso E debido a la inclusión de la componente de saltos para el proceso de fragilidad.

Finalmente se realizaron corridas del algoritmo EM con diferentes valores para el vector inicial de parámetros y se estimaron los parámetros para ajustar la función de intensidad y analizar los resultados correspondientes.

Como anexo integral del trabajo se presentan los resultados de las estimaciones del proceso autorregresivo asociado a las covariables observables de la intensidad en el apéndice A, en el apéndice B se presenta una revisión exploratoria de la literatura de modelos de riesgo crediticio, además en el anexo C se muestra el programa en R implementado para el modelo propuesto.

2. Marco teórico

La teoría moderna de la probabilidad se ha convertido en una de las principales herramientas en la formulación de métodos cuantitativos para modelar la dinámica en la evolución de activos financieros.

En este capítulo se presentan los principales fundamentos teóricos tras los planteamientos propuestos por Duffie y los elementos utilizados para plantear la propuesta de modificación incluyendo una componente de saltos en el proceso de fragilidad inmerso en la función de intensidad, con la cual se busca explicar la dinámica de los incumplimientos crediticios.

En las secciones 2.2 y 2.3 se introducen brevemente los procesos implementados en la función de intensidad, a saber, el Proceso de Poisson, el Movimiento Browniano y el Proceso Ornstein-Uhlenbeck. Luego en la sección 2.4, se explican detalladamente las metodologías de estimación, partiendo de la definición de la función de verosimilitud y el uso de las propiedades de las Cadenas de Markov, las cuales son fundamentales para garantizar convergencia en los resultados obtenidos tras el proceso de estimación. Posteriormente en las secciones 2.4.6 a la 2.4.9, se describen los principales elementos y algoritmos mediante los cuales se abordó el proceso de estimación de los parámetros, fundamentalmente a partir de los algoritmos EM, Muestreador de Gibbs y el Metropolis Hasting.

Para abordar algunos fundamentos relacionados con procesos estocásticos se presentarán a continuación los elementos estadísticos y matemáticos utilizados, dejando de lado algunas pruebas matemáticas para las que se sugiere al lector remitirse algunos textos guía.

2.1. Procesos estocásticos

Los procesos estocásticos se constituyen como la principal herramienta para modelar la dinámica en el tiempo de una variable aleatoria. La aleatoriedad del fenómeno se captura a partir de un espacio de probabilidad $(\Omega, \mathcal{F}, \mathcal{P})$, donde Ω es el espacio muestral con todos los posibles resultados, \mathcal{F} una σ -álgebra que representa la colección de eventos relevantes para el investigador y \mathcal{P} corresponde a una medida de probabilidad, Dobrow (2016).

Un proceso estocástico $\{X_t\}_{t \in T}$ está conformado por un conjunto de variables aleatorias $\{X_t\}_{t \in T}$, donde T es un conjunto discreto o continuo de números reales que representan el tiempo, ver figura 2-1. En el desarrollo de la teoría de las finanzas cuantitativas y el riesgo, la evolución del precio de acciones, tasas de interés y precios de derivados financieros constituyen ejemplos de variables que pueden ser modeladas por procesos estocásticos, sin embargo, se puede pensar en una infinidad de fenómenos que también se pueden representar por procesos estocásticos, por ejemplo la hidrología en una cuenca hidrográfica, el flujo de partículas en un sistema, entre otros fenómenos físicos, Douglas Rivas (2019).

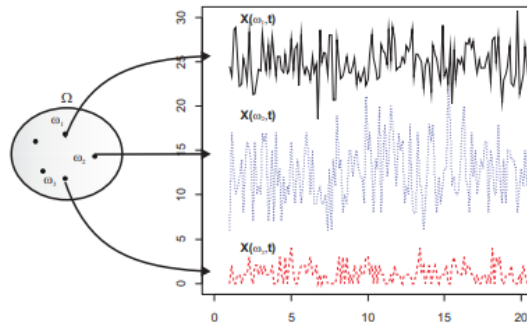


Figura 2-1.: Trayectorias de un proceso estocástico. Tomado de Douglas Rivas (2019)

Definición 1. Si $(\Omega, \mathcal{F}, \mathbb{P})$ es un espacio de probabilidad, y $T > 0$, un proceso estocástico X en $[0, T]$ es una función:

$$X : [0, T] \times \Omega \longrightarrow \mathbb{R} \quad (2-1)$$

Para cada $\omega \in \Omega$, la función $t \longrightarrow X(t, \omega)$ se denomina trayectoria (*path*) o realización de un proceso estocástico, ver figura 2-1. Para cada $t \in [0, T]$ fijo, X_t es una variable aleatoria. El conjunto de índices, como se denomina a $[0, T]$ puede ser discreto, finito ó infinito y de cualquier dimensión.

Definición 2. Un proceso estocástico $(X_t)_{t \in I}$ presenta incrementos independientes si para $r < u < s < t$ con $r, u, s, t \in I$, se tiene entonces que $X_t - X_s$ es independiente de $X_u - X_r$. Un proceso estocástico $(X_t)_{t \in I}$ tiene incrementos estacionarios si para cada $s < t$ con $s+h, t+h \in I$ se tiene que $X_{t+h} - X_{s+h} \sim X_t - X_s$, en otras palabras, la distribución de los incrementos $X_t - X_s$ sólo depende de la diferencia del tiempo $t - s$.

2.1.1. Movimiento Browniano

El movimiento Browniano se utiliza como bloque fundamental para la construcción de modelos que permiten simular, por ejemplo, la evolución en el tiempo del precio de un activo financiero, o la cuenta de balance de un individuo, tal como se plantea en este trabajo para una de las covariables observadas, ver Martínez (2006) pág. 33

Definición 3. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio fijo de probabilidad, el movimiento Browniano estándar es una función:

$$W : [0, \infty) \times \Omega \longrightarrow \mathbb{R} \quad (2-2)$$

Tal que para cada $t \geq 0$, la función $W(t, \cdot) : \Omega \longrightarrow \mathbb{R}$, es una variable aleatoria definida en (Ω, \mathcal{F}) . Mientras que para cada $\omega \in \Omega$ la función $W(\cdot, \omega) : [0, \infty) \rightarrow \mathbb{R}$, es continua en $[0, \infty)$. A $W(\cdot, \omega)$ se le denomina trayectoria y se denotan por $\omega(t)$. Además se cumplen las siguientes propiedades:

1. $W_0 = 0$, c.p1., es decir, el proceso inicia en $t = 0$ con probabilidad uno.
2. Para cualquier conjunto de tiempos $0 < t_1 < t_2 < \dots < t_n$, los incrementos $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ son independientes entre sí.
3. Para cualquier par de tiempos t y s con $0 \leq s \leq t$, $W_t - W_s \sim N(0, t - s)$.

Es decir $W_t - W_s$ posee distribución normal con media cero y varianza $t - s$.

2.1.2. Proceso Ornstein–Uhlenbeck

El proceso Ornstein-Uhlenbeck es conocido también como caminata aleatoria elástica, que corresponde a un Proceso de Markov con incrementos normalmente distribuidos. A diferencia del Movimiento Browniano, el cual es un proceso inestable que después de un largo tiempo tiende a valores infinitos, el proceso Ornstein-Uhlenbeck posee una distribución estacionaria. En finanzas el Proceso Ornstein-Uhlenbeck es adaptado en el modelo de Vasicek y es utilizado para modelar el comportamiento de las tasas de interés, este proceso presenta reversión a la media sobre la tendencia con el fin de buscar una situación de equilibrio, la ecuación diferencial estocástica del proceso Ornstein-Uhlenbeck está dada por, ver Martínez (2006) pág 522 y Tankov (2003) pág 390:

$$dY_t = -\kappa(Y_t - \mu)dt + \sigma dB_t \quad (2-3)$$

Donde κ, μ y $\sigma > 0$ son parámetros. El término de tendencia $-\kappa(Y_t - \mu)$ representa una fuerza que jala al proceso a su media de largo plazo μ , la rapidez con que el proceso revierte a su media es controlada por κ . La componente aleatoria σdB_t hace que el proceso fluctúe de forma irregular alrededor de la media de largo plazo.

El proceso Ornstein-Uhlenbeck también puede expresarse por medio de integrales de la siguiente forma:

$$Y_t = Y_0 + \int_0^t Y_s ds + \int_0^t dB_s. \quad (2-4)$$

Sabemos que (2-4) puede representarse en forma diferencial por:

$$dY_t = -\kappa Y_t dt + dB_t, \quad (2-5)$$

Ahora se expondrá una representación más amena para el proceso Y_t que se obtiene al resolver (2-5).

Reordenando términos y multiplicando por el factor integrante $e^{\kappa t}$ se obtiene:

$$\begin{aligned} e^{\kappa t} (dY_t + \kappa Y_t) &= e^{\kappa t} dB_t, \\ e^{\kappa t} dY_t + \kappa e^{\kappa t} dt &= e^{\kappa t} dB_t, \\ d(e^{\kappa t} Y_t) &= e^{\kappa t} dB_t, \end{aligned} \quad (2-6)$$

Recordando que $Y_0 = 0$, se tiene que:

$$e^{\kappa t} Y_t = \int_0^t d(e^{\kappa s} Y_s) = \int_0^t e^{\kappa s} dB_s, \quad (2-7)$$

De donde finalmente,

$$Y_t = e^{-\kappa t} \int_0^t e^{\kappa s} dB_s, \quad (2-8)$$

En este caso, sabemos que Y_t posee distribución normal con media cero y varianza:

$$\text{Var}[Y_t] = e^{-2\kappa t} \int_0^t e^{2\kappa s} ds. \quad (2-9)$$

Finalmente una trayectoria del proceso tiene la dinámica de la figura **2-2**.

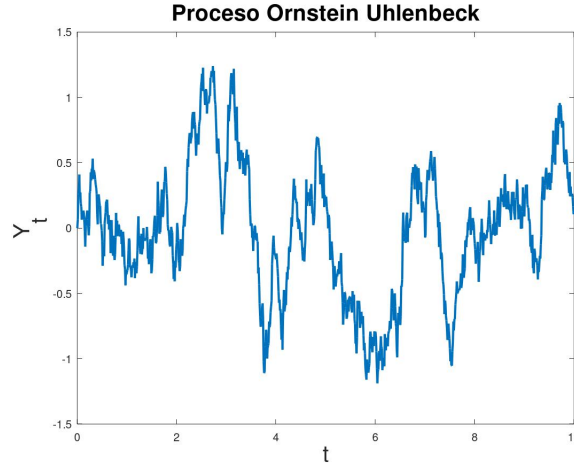


Figura 2-2.: Una trayectoria del proceso Ornstein Uhlenbeck Y_t , con $Y_0 = 0$, $\kappa = 2$.

2.1.3. Proceso Ornstein-Uhlenbeck con saltos

Ahora consideramos una ligera modificación del proceso Ornstein-Uhlenbeck al que se le incorpora una componente de saltos, más específicamente nos concentraremos en el modelo:

$$dY_t = -\kappa Y_t + dB_t + dJ_{t-},$$

$$J_t = \sum_{i=1}^{N_t} \xi_i.$$

(2-10)

En (2-10) el proceso N_t representa un proceso de Poisson simple con intensidad constante λ , $\{\xi_i\}$ es una sucesión iid de variables normales $N(0, \sigma^2)$ independientes del Movimiento Browniano B_t y dJ_{t-} denota el proceso:

$$dJ_{t-} = J_t - J_{t-} = \begin{cases} 0 & \text{si } N_t - N_{t-} = 0 \\ \xi_i & \text{si } N_t - N_{t-} = 1 \end{cases} \quad (2-11)$$

Dicho de otra forma J_t es un proceso de Poisson compuesto y la distribución de los saltos es normal $N(0, \sigma^2)$. La idea en (2-10) es que el nivel del proceso OU es perturbado en tiempos exponencialmente distribuidos por pulsos ξ_i normalmente distribuidos, ver figura 2-3.

Se justificará que Y_t esta vez viene dado por:

$$\begin{aligned} Y_t &= Y_t^c + J_t, \\ &= e^{-\kappa t} \int_0^t e^{\kappa s} dB_s + \sum_{i=1}^{N_t} \xi_i, \end{aligned}$$

(2-12)

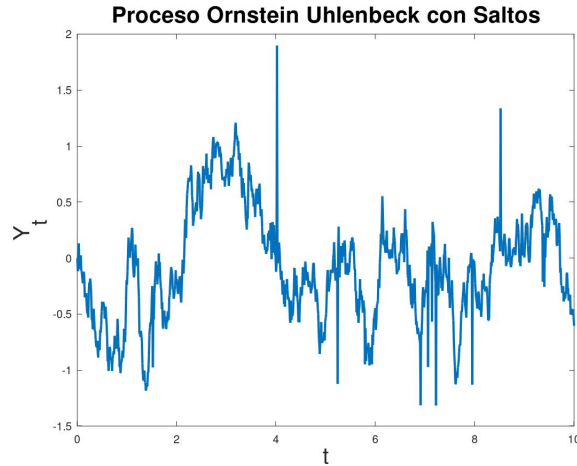


Figura 2-3.: Una trayectoria del proceso Ornstein Uhlenbeck Y_t , con $Y_0 = 0$, $\kappa = 2$, $\lambda = 2$, $\sigma = 1$.

2.2. Proceso Poisson

El proceso de Poisson es un proceso de conteo, representa la ocurrencia de un número determinado de eventos en el tiempo que inicia en $t = 0$, se denota por N_t el número de eventos hasta un instante t , por ejemplo el número de personas que llegan a una unidad de urgencias o el número de incumplimientos de pago en una franquicia. Como se plantea a partir de la función de intensidad del modelo descrito en este trabajo, los tiempos en los que se presentan los default son exponencialmente distribuidos pero el número de incumplimientos proviene de un Proceso de Poisson. A continuación se presentan algunas definiciones y conceptos adyacentes a este proceso, para ver con mayor detalle algunas de las propiedades del Proceso de Poisson, ver la sección 2.5 de la pág 44 de Tankov (2003).

Definición 4. Un proceso de conteo $(N_t)_{t \geq 0}$ es una colección de variables aleatorias, enteras no negativas, tal que si $0 \leq s \leq t$, entonces $N_s \leq N_t$.

El proceso de Poisson es un claro ejemplo de proceso estocástico con trayectorias discontinuas y es ampliamente usado como punto de partida para construir procesos de saltos más complejos. A continuación se presenta formalmente la definición de proceso de Poisson, tomado de Tankov (2003):

Definición 5. Un Proceso de Poisson homogéneo de parámetro λ es un proceso de conteo $(N_t)_{t \geq 0}$ con las siguientes propiedades, Tankov (2003) y Karlin and Taylor (2012):

1. $N_0 = 0$
2. Para cada $t > 0$, N_t tiene una distribución Poisson con parámetro λt

3. (Incrementos estacionarios). Para cada $s, t > 0$ $N_{t+s} - N_s$ tiene la misma distribución de N_t , es decir:

$$\mathbb{P}(N_{t+s} - N_s = k) = \mathbb{P}(N_t = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \text{ para } k = 0, 1, \dots \quad (2-13)$$

4. Para $0 \leq q < r \leq s < t$, $N_t - N_s$ y $N_r - N_q$ son variables aleatorias independientes.

Proposición 1. Superposición. Si $(N_t^{(1)})_{t \geq 0}, \dots, (N_t^{(n)})_{t \geq 0}$ son n procesos de Poisson independientes con parámetros $\lambda_1, \dots, \lambda_n$, donde $N_t = N_t^{(1)} + \dots + N_t^{(n)}$, para $t \geq 0$. Entonces $(N_t)_{t \geq 0}$ es un proceso de Poisson con parámetro $\lambda = \lambda_1 + \dots + \lambda_n$.

2.2.1. Proceso de Poisson no homogéneo

En un proceso de Poisson homogéneo, la ocurrencia de eventos ocurre a una tasa de intensidad constante. Cuando la tasa de ocurrencia o intensidad con la que se presentan eventos cambia con el tiempo se considera un proceso de **Poisson no homogéneo**, tal como ocurre con la tasa de intensidad con la cual se explica la dinámica de los incumplimientos en una cartera de crédito.

Definición 6. Un Proceso de conteo $(N_t)_{t \geq 0}$ es un proceso de Poisson no homogéneo con función de intensidad $\lambda(t)$, si:

1. $N_0 = 0$
2. Para cada $t > 0$, N_t tiene una distribución Poisson con media:

$$E(N_t) = \int_0^t \lambda(x) dx. \quad (2-14)$$

3. Para $0 \leq q < r \leq s < t$, $N_t - N_s$ y $N_r - N_q$ son variables aleatorias independientes.

En contraste con el proceso de Poisson homogéneo el proceso de Poisson no homogéneo también presenta incrementos independientes, pero no necesariamente incrementos estacionarios.

2.2.2. Proceso de Cox

Un proceso de Cox, también conocido como Proceso de Poisson doblemente estocástico, consiste en una generalización del proceso de Poisson donde la tasa de intensidad Λ es en sí misma un proceso estocástico, las propiedades de un proceso de Poisson se mantienen dado que los Procesos Poisson homogéneo y no homogéneo son casos particulares del Proceso de Cox. Si $(N_t)_{t \geq 0}$ es un Proceso de Cox, entonces la distribución condicional de N_t dado $\Lambda = \lambda$, es Poisson con parámetro

λ_t , McNeil et al. (2005).

Claramente el proceso mediante el cual se explica la dinámica de los incumplimientos de crédito y se formula el modelo de partida propuesto en Duffie et al. (2009), cumple con esta especificación, es decir, para una cartera de m prestatarios se tiene que la intensidad se puede expresar como:

$$\lambda_{i,t} = \exp(X_t). \quad (2-15)$$

Donde el X_t es a su vez un proceso estocástico, en este caso gaussiano, aquí se parte de un proceso que depende de dos procesos autorregresivos que explican covariables inherentes al comportamiento de pago de los individuos $U_{i,t}$ y una covariable común a todos que refleja el estado de la economía V_t , y donde además se considera una covariable latente especificada mediante un proceso Ornstein-Uhlenbeck con saltos Y_t .

En este caso se tiene que:

$$X_t = \beta_0 + \beta_1 V_t + \beta_2 U_{i,t} + \eta Y_t \quad (2-16)$$

Para un vector de parámetros $\Theta = (\beta_0, \beta_1, \beta_2, \eta)$.

2.2.3. Procesos de Poisson compuestos

Los procesos de Poisson compuestos son una generalización donde los tiempos de espera entre saltos que son exponencialmente distribuidos, pero de otro lado, los tamaños del salto pueden tener una distribución arbitraria.

Sea N un proceso de Poisson con parámetro λ y $\{\xi_i\}_{i \geq 1}$ una secuencia de variables aleatorias independientes, e idénticamente distribuidas, el proceso,

$$J_t = \sum_{i=1}^{N_t} \xi_i. \quad (2-17)$$

Se denomina proceso de Poisson compuesto, el cual tiene incrementos independientes.

Aquí se considera un proceso de Poisson Compuesto inmerso en el proceso de saltos considerado para una de las componentes del proceso Ornstein-Uhlenbeck.

2.3. Métodos de estimación

En esta sección se esbozan los métodos numéricos utilizados para el programa implementado en el trabajo, el principal objetivo consiste en estimar los parámetros de máxima verosimilitud asociados a la función de intensidad. Para la estimación conjunta de los parámetros en presencia de información observable y latente se hace necesario implementar el algoritmo Esperanza Maximización el cuál se explicará en breve. En primera instancia, se parte de la definición de función de verosimilitud la cual es una función de los parámetros del modelo planteado que permite realizar inferencias acerca de su valor a partir de un conjunto de observaciones.

Al no contar con información sobre la variable latente, se hace necesario muestrear las trayectorias del proceso de fragilidad mediante el muestreador de GIBBS y el algoritmo Metropolis-Hastings, los cuales constituyen dos metodologías de estimación que se implementan en Cadenas de Markov Monte Carlo (Markov Chain Monte Carlo), en lo sucesivo por sus siglas en inglés **MCMC**.

2.3.1. Función de verosimilitud

Considere un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)^T$ el cual tiene función de distribución de probabilidad conjunta $f_X(\mathbf{X}; \Theta)$, indexada con un vector de parámetros $\Theta = (\beta_1, \dots, \beta_n)^T$. Se considera además que la información observable corresponde a una realización de \mathbf{X} para un valor desconocido de Θ . Es decir, se asume que se dispone de observaciones X_1 a X_n en cada componente de \mathbf{X} McNeil et al. (2005).

La función de verosimilitud para el vector de parámetros Θ , dada la información observable es $L(\Theta; \mathbf{X}) = f_X(\mathbf{X}; \Theta)$ y el Estimador de Máxima Verosimilitud (MLE) $\hat{\Theta}$ es el valor de Θ que maximiza $L(\Theta; \mathbf{X})$ o de forma equivalente, el valor que maximiza la función de log-verosimilitud $\mathcal{L}(\Theta; \mathbf{X}) = \ln L(\Theta; \mathbf{X})$. El estimador de Θ se denota como $\hat{\Theta}_n$ cuando se hace hincapié a su dependencia de una muestra de tamaño n . Para n grande se espera que el estimador $\hat{\Theta}_n$, se aproxime a su valor real Θ , los resultados derivados de las propiedades asintóticas dan información acerca de la calidad del estimador en muestras grandes. Para describir la forma en que se presentan los resultados se considera una situación en la cual \mathbf{X} se asume como un vector de componentes iid con función de densidad univariada f tal que, McNeil et al. (2005):

$$\ln L(\Theta; X) = \ln \prod_{i=1}^n f(X_i; \Theta) = \sum_{i=1}^n \ln L(\Theta; X_i), \quad (2-18)$$

2.3.2. Resultados asintóticos

Bajo condiciones fuertes de regularidad (ver Casella and Berger (2002), pág. 516), dado un vector de p parámetros, el estimador de máxima verosimilitud $\hat{\Theta}_n$ de Θ presenta eficiencia asintótica, en

el sentido en que, para $n \rightarrow \infty$ se tiene que:

$$\sqrt{n}(\hat{\Theta}_n - \Theta) \xrightarrow{d} N_p(\mathbf{0}, \mathbf{I}(\Theta^{-1})), \quad (2-19)$$

Dónde $\mathbf{I}(\Theta)$ denota la matriz de información de Fisher, definida por,

$$\mathbf{I}(\Theta) = E \left(\frac{\partial}{\partial \Theta} \ln L(\Theta; \mathbf{X}) \frac{\partial}{\partial \Theta'} \ln L(\Theta; \mathbf{X}) \right) = -E \left(\frac{\partial^2}{\partial \Theta \partial \Theta'} \ln L(\Theta; \mathbf{X}) \right) \quad (2-20)$$

El resultado de la convergencia implica que, para un n suficientemente grande, tenemos que:

$$\hat{\Theta}_n \sim N_p(\Theta, [n\mathbf{I}(\Theta)]^{-1}), \quad (2-21)$$

Este resultado es de utilidad para construir regiones de confianza para Θ o intervalos de confianza para cada componente Θ_j . En la práctica se puede aproximar la matriz de información de Fisher $\mathbf{I}(\Theta)$ a partir de la *matriz de información de Fisher observada*:

$$\bar{\mathbf{I}}(\Theta) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \Theta \partial \Theta'} \ln L(\Theta; X_i) \quad (2-22)$$

Para cada realización obtenida del vector de estados \mathbf{X} . A partir de la ley de los grandes números se alcanzaría convergencia a la matriz de información esperada.

2.4. Cadenas de Markov Montecarlo

Las Cadenas de Markov Monte Carlo (MCMC) comprenden una metodología que utiliza cadenas de Markov para simular de una manera efectiva distribuciones de probabilidad en espacios de alta dimensión que de otra forma serían intratables. La metodología MCMC ha sido descrita como una *revolución*, un *cambio de paradigma* en el campo de la estadística y uno de los 10 algoritmos más importantes del siglo XX, ver Diaconis (2009). En lo que sigue se presenta una breve introducción a las Cadenas de Markov con el fin de abordar las MCMC, algunas definiciones fueron tomadas de Dobrow (2016), véase también Grimmett and Stirzaker (2001).

2.4.1. Cadenas de Markov

A continuación se definen los conceptos subyacentes a las Cadenas de Markov cuyas propiedades son fundamentales para el uso de las Cadenas de Markov Monte Carlo necesarias para implementar el Muestreador de Gibbs. En primera instancia se definirán brevemente y se presentan algunas definiciones y propiedades importantes relevantes.

Definición 7. Una **Cadena de Markov** es una secuencia de variables aleatorias $\{Y_i\}_{i=0}^{\infty}$ que toman valores en un conjunto discreto \mathcal{S} con la propiedad de:

$$\begin{aligned} P(Y_{n+1} = j | Y_0 = y_0, \dots, Y_{n-1} = y_{n-1}, Y_n = y_n) \\ = P(Y_{n+1} = j | Y_n = y_n) \end{aligned} \quad (2-23)$$

para todo $y_0, y_1, \dots, y_n, j \in \mathcal{S}$.

Definición 8. Considere una distribución de probabilidad $\mathbf{p} = (p_i)_{i \in \mathcal{S}}$ en \mathcal{S} , es decir $\mathbf{p} : \mathcal{S} \rightarrow [0, 1]$ es una función de masa de probabilidad que cumple:

1. $\mathbf{p}(i) = p_i \geq 0$ para todo $i \in \mathcal{S}$.
2. $\sum_{i=0}^{\infty} p_i = 1$.

Decimos que \mathbf{p} es una distribución límite de la cadena de Markov Y_n si:

$$\lim_{n \rightarrow \infty} P(Y_n = j | Y_0 = i) = \mathbf{p}(j) = p_j. \quad (2-24)$$

para todo $i, j \in \mathcal{S}$.

Definición 9. Decimos que una distribución de probabilidad \mathbf{p} en el espacio de estados \mathcal{S} de una cadena de Markov Y_n es una **distribución estacionaria** de la cadena, si dado que Y_0 posea distribución \mathbf{p} implica que Y_n también posee distribución \mathbf{p} para todo n , es decir \mathbf{p} debe cumplir:

$$p_j = P(Y_n = j) = \sum_{i \in \mathcal{S}} P(Y_n = j | Y_0 = i) p_i. \quad (2-25)$$

para todo $n = 1, 2, \dots$, y todo $i, j \in \mathcal{S}$, Dobrow (2016).

Para que la distribución de una cadena de Markov Y_n discreta converja a la distribución estacionaria, la cadena necesita satisfacer 3 condiciones técnicas importantes, acá sólo se mencionan, para mayor detalle ver Dobrow (2016):

- Primero la cadena debe ser **irreducible**. Es decir, comenzando desde cualquier punto (estado), debe ser posible para la cadena alcanzar cualquier conjunto no vacío con probabilidad positiva en un número finito de iteraciones. Se trata esencialmente de una propiedad de conexión probabilística.

- Segundo, la cadena necesita ser **aperiódica**. Esto evita que la cadena oscile de forma regular entre conjuntos de diferentes estados.
- Último y más importante, la cadena debe ser **positiva recurrente**, (ver Grimmett and Stirzaker (2001)) para las definiciones pertinentes).

Teorema 1. *Para una cadena irreducible y aperiódica, se tiene que:*

$$\lim_{n \rightarrow \infty} P_{ij}(n) = p_j. \quad (2-26)$$

Dónde p_j es una distribución estacionaria de Y_n .

Observe que para algún estado i con la propiedad que $P_{ii} > 0$ la cadena es necesariamente aperiódica. De esta manera, una condición suficiente para que una cadena de Markov sea aperiódica es que $P_{ii} > 0$ para algún estado i . Esto implica que al menos una entrada de la matriz diagonal no sea igual a 0.

El algoritmo MCMC construye una **cadena de Markov ergódica** para la cual **la distribución límite** es la distribución buscada $p(y|\Theta)$. Una vez se corre un número de iteraciones suficiente para que la cadena converja a su distribución límite, las últimas realizaciones de la secuencia de Markov corresponden a una muestra de una distribución muy próxima a $p(y|\Theta)$.

Las trayectorias son generadas de forma secuencial, donde la distribución de las muestras obtenidas dependen del último valor muestreado; por tanto las trayectorias forman una **Cadena de Markov**. A continuación se describen las principales definiciones y propiedades que justifican los pilares fundamentales del funcionamiento del algoritmo MCMC.

Proposición 2. *Una cadena de Markov es denominada ergódica, si es irreducible, aperiódica, y todos los estados tienen tiempos de retorno esperados finitos. Lo anterior siempre es cierto para cadenas finitas. Así, una cadena de Markov finita es ergódica si ésta es irreducible y aperiódica. Esto es precisamente la clase de cadenas de Markov que tienen distribuciones límite positivas.*

Teorema 2. *Sea Y_0, Y_1, \dots una cadena de Markov ergódica. Entonces, existe una única distribución estacionaria p , la cual es la distribución límite de la cadena. Es decir,*

$$p_j = \lim_{n \rightarrow \infty} P_{ij}^n, \text{ para, todo, } i, j.$$

Definición 10. *Una cadena de Markov irreducible con matriz de transición P y distribución estacionaria p es reversible, o reversible en el tiempo, si:*

$$p_i P_{ij} = p_j P_{ji}, \text{ para, todo, } i, j. \quad (2-27)$$

Las ecuaciones de la forma 2-27 son denominadas ecuaciones de balance. Estas establecen que para una cadena estacionaria,

$$P(Y_0 = i, Y_1 = j) = P(Y_0 = j, Y_1 = i), \text{ para, todo, } i, j. \quad (2-28)$$

Es decir, la frecuencia de las transiciones desde el estado i al estado j es igual a la frecuencia de las transiciones del estado j al estado i .

Si una cadena de Markov estacionaria es reversible, entonces:

$$P(Y_0 = i_0, Y_1 = i_1, \dots, Y_n = i_n) = P(Y_0 = i_n, Y_1 = i_{n-1}, \dots, Y_n = i_0). \quad (2-29)$$

Para, todo, i_0, i_1, \dots, i_n .

Para ampliar remitirse al ejemplo 3.39 de Dobrow (2016).

Proposición 3. *Sea P la matriz de transición de una cadena de Markov. Si y es una función de distribución de probabilidad que satisface,*

$$p_i P_{ij} = p_j P_{ji}, \text{ para, todo, } i, j. \quad (2-30)$$

Entonces y es la distribución estacionaria, y la cadena de Markov es reversible.

Este resultado fundamenta las bases del algoritmo del muestreador de Gibbs que se implementa en la metodología de estimación.

El algoritmo MCMC se fundamenta en que las propiedades de una cadena de Markov ergódica que presentan similitudes a secuencias independientes e idénticamente distribuidas. En particular se conserva la ley fuerte de los grandes números que se describe a continuación y que permite aproximar el valor esperado de una cadena de Markov con distribución estacionaria $p(y|\Theta)$.

2.4.2. Ley de grandes números

La ley de los grandes números es uno de los teoremas límite de la probabilidad. Si Y_1, Y_2, \dots es una secuencia iid con una media común $\mu < \infty$, entonces la ley fuerte de grandes números, establece que, con probabilidad igual a 1,

$$\lim_{n \rightarrow \infty} \frac{Y_1 + \dots + Y_n}{n} = \mu \quad (2-31)$$

De forma equivalente, considere una variable aleatoria Y con la misma distribución de Y_i y considere además que r es una función de valor real. Entonces, $r(Y_1), r(Y_2), \dots$ es también una secuencia iid con media finita, y, con probabilidad igual a 1,

$$\lim_{n \rightarrow \infty} \frac{r(Y_1) + \dots + r(Y_n)}{n} = E(r(Y)). \quad (2-32)$$

$$(2-33)$$

Teorema 3. Considere que X_0, X_1, \dots es una cadena de Markov con distribución estacionaria $p(y|\Theta)$. Considere una función de valor real r . Sea X una variable aleatoria con distribución $p(y|\Theta)$. Entonces, con probabilidad igual 1,

$$\lim_{n \rightarrow \infty} \frac{r(X_1) + \dots + r(X_n)}{n} = E(r(X)) \quad (2-34)$$

Donde $E(r(X)) = \int r(j)p(y|\Theta)dy$.

2.4.3. Algoritmo Metropolis Hasting

El algoritmo de Metropolis Hasting es el algoritmo de una cadena de Markov reversible Y_0, Y_1, \dots cuya distribución estacionaria es $p(y|\Theta) = (p_1, p_2, \dots)$. Considere una matriz de transición q_t para alguna cadena de Markov irreducible con el mismo espacio de estados $p(y|\Theta)$ en que esta soportada, observe que aquí cada sucesión o secuencia, $Y_t, t = 1, 2, 3, \dots$, es generada a partir de un punto inicial Y_0 , y luego, para cada t , se muestrea Y_t a partir de la distribución de transición, $q_t(Y_t^j | Y_{t-1}^{j-1})$. Se supone que a priori se conoce la forma de muestrear trayectorias de q_t . La cadena q_t se utilizará como una cadena propuesta, generando elementos de una sucesión que el algoritmo elige mediante un criterio de aceptación.

Con el fin de describir el proceso, considere una cadena Y_{t-1}^{j-1} en un estado i , el próximo paso de la cadena Y_t^j está determinado por el siguiente proceso iterativo:

1. Genere una trayectoria para el proceso q propuesto, es decir, elija j con probabilidad q_{ij} , el estado j es llamado estado propuesto.
2. Criterio de aceptación de j :

$$\alpha(i, j) = \frac{p_j q_{ji}}{p_i q_{ij}} \quad (2-35)$$

Si $\alpha(i, j) \geq 1$, se acepta a j como el siguiente estado de la cadena, en caso contrario, si $\alpha(i, j) < 1$, se conserva i como el próximo estado de la cadena $Y_t^j = i$. En otras palabras, asuma que $Y_t^j = i$, luego se recurre a la distribución uniforme (0 y 1) para muestrear un número C , donde el próximo estado de la cadena es tal que:

$$Y_t = \begin{cases} j & C \leq \alpha(i, j), \\ i & C > \alpha(i, j). \end{cases} \quad (2-36)$$

Proposición 4. Si la cadena propuesta es ergódica (irreducible y aperiódica) se cumple que la cadena resultante de aplicar al algoritmo también es ergódica con distribución límite igual a $p(y|\Theta)$.

Algunas consideraciones sobre el algoritmo:

- Si la matriz de transición de propuesta q es simétrica, entonces $\alpha(i, j) = \frac{p_j}{p_i}$.
- El algoritmo trabaja para una cadena propuesta irreducible. Es decir, se cuenta con una amplia variedad de alternativas de establecer una cadena propuesta que sea eficiente en el contexto del problema.
- Al ser necesario que la cadena resultante sea estacionaria se deben utilizar muchas iteraciones, sin embargo, siempre existe un sesgo remanente debido a las observaciones iniciales. Por tanto se aplica un procedimiento de quemado o descarte de las observaciones iniciales (*Burn in*), es decir, para algún número m de iteraciones, por la ley fuerte de los grandes números para Cadenas de Markov se tiene que:

$$\lim_{n \rightarrow \infty} \frac{r(X_m) + \dots + r(X_n)}{n - m + 1} = \int r(x)p(y|\Theta)dy. \quad (2-37)$$

La prueba de que la secuencia de las iteraciones Y^0, Y^1, \dots converge a una distribución objetivo tiene dos pasos: Primero, se debe mostrar que la trayectoria o secuencia de valores simulados es una cadena de Markov con única distribución estacionaria, en segunda instancia, que esta distribución estacionaria es igual a esta distribución objetivo.

Para que se cumpla lo primero, la cadena de Markov debe ser irreducible, aperiódica y no transitoria. Salvo algunas excepciones, estas dos últimas condiciones se mantienen para una caminata aleatoria con una distribución apropiada, la propiedad de irreducibilidad se mantiene siempre que las probabilidades de pasar de un estado a otro sean positivas, es decir, la distribución de transición q_t debe eventualmente alcanzar todos los estados con probabilidad positiva, para ver la prueba remitirse a la página 188 de Dobrow (2016).

2.4.4. Muestreador de GIBBS

El muestreador de GIBBS es una adaptación del Metropolis-Hasting, especialmente utilizado en problemas de alta dimensión. En el muestreador de GIBBS la distribución objetivo $p(y|\Theta)$ es una función de distribución conjunta m -dimensional $p(Y) = p(y_1, \dots, y_m)$. Este algoritmo genera elementos de la forma $\mathbf{Y}^0, \mathbf{Y}^1, \mathbf{Y}^2, \dots = (Y_1^0, \dots, Y_m^0), (Y_1^1, \dots, Y_m^1), (Y_1^2, \dots, Y_m^2), \dots$ mediante un proceso iterativo, actualizando cada componente de un vector de estados m -dimensional condicionado en $m - 1$ componentes, ver Dobrow (2016) pág 198.

En el muestreador de Gibbs, la distribución objetivo $p(y|\Theta)$ es una función de densidad de probabilidad conjunta de la forma,

$$p(y|\Theta) = p(Y_1, \dots, Y_m) \quad (2-38)$$

Una cadena de Markov multivariada es construida de forma tal que su distribución límite es $p(y|\Theta)$, y toma valores en un espacio m -dimensional.

Suponga que $Y = (Y_1, \dots, Y_m)$ y se tienen las siguientes distribuciones posteriores condicionales:

$$\begin{aligned} p_1 & (Y_1 | Y_2, \dots, Y_m, \text{datos}) \\ p_2 & (Y_2 | Y_1, Y_3, \dots, Y_m, \text{datos}) \\ p_3 & (Y_3 | Y_1, Y_2, \dots, Y_m, \text{datos}) \\ & \cdot \\ & \cdot \\ p_m & (Y_m | Y_1, Y_3, \dots, Y_{m-1}, \text{datos}), \end{aligned}$$

donde *datos* se refiere a la información observada.

El algoritmo se describe en los siguientes pasos:

1. Inicializar $j = 0$, $Y^{(0)} = (Y_1^{(0)}, \dots, Y_m^{(0)})$.
2. Generar $Y_1^{(j+1)} \sim p_1(Y_1 | Y_2^{(j)}, \dots, Y_m^{(j)})$.
3. Generar $Y_2^{(j+1)} \sim p_2(Y_2 | Y_1^{(j+1)}, Y_3^{(j)}, \dots, Y_m^{(j)})$.
-
-
-
4. Generar $Y_m^{(j+1)} \sim p_m(Y_m | Y_1^{(j+1)}, Y_2^{(j+1)}, \dots, Y_{m-1}^{(j+1)})$.
5. Incremente j y regrese al paso 1.

Para ilustrar el Muestreador de Gibbs considere una distribución normal estándar ($\sigma_x = \sigma_y = 1$) bivariada con coeficiente de correlación ρ . La distribución normal bivariada tiene la propiedad de que sus distribuciones condicionales son normales. Si (X, Y) tiene una distribución bivariada normal estándar, entonces la distribución condicional de Y dado $X = x$ es normal con media ρx y varianza $1 - \rho^2$. De forma similar, la distribución condicional de X dado $Y = y$ es normal con media ρy y varianza $1 - \rho^2$, Dobrow (2016).

El muestreador de Gibbs es implementado para simular Y a partir de una distribución bivariada normal estándar con coeficiente de correlación ρ , En cada paso del algoritmo, un componente de

un vector de dos entradas es actualizado muestreando a partir de su distribución condicional dado el otro componente. Cada actualización cambia de componente a componente. La sucesión resultante de una muestra bivariada converge a la distribución objetivo.

1. Inicializar: $(y_0, x_0) \leftarrow (0, 0)$, $m \leftarrow 1$.
2. Generar y_m a partir de la distribución condicional de X dado $Y = y_{m-1}$. Es decir, simule a partir de una distribución normal con media ρx_{m-1} y varianza $1 - \rho^2$.
3. Generar x_m a partir de la distribución condicional de x dado $Y = y_m$. Es decir, simule a partir de una distribución normal con media ρy_m y varianza $1 - \rho^2$.
4. $m \leftarrow m + 1$.
5. Retornar al paso 2.

Como se mencionó, el muestreador de Gibbs es un caso especial del algoritmo Metropolis-Hastings. Para ver esto, considere una distribución conjunta m -dimensional $p(y|\Theta)$. Se ilustrará solo un paso del muestreador de Gibbs en el caso de la actualización de la primera componente.

Considere que $i = (y_1, y_2, \dots, y_m)$ es el estado actual de una cadena de Markov m -dimensional Y , y $j = (y'_1, y_2, \dots, y_m)$ es el estado propuesto. La distribución de propuesta q es la distribución condicional de Y_1 dado Y_2, \dots, Y_m . La función de aceptación es $\alpha(i, j) = \frac{p_j q_{ji}}{p_i q_{ij}}$, se tiene que,

$$\begin{aligned}
 p_j q_{ji} &= p(y'_1, y_2, \dots, y_m) \left(\frac{p(y_1, y_2, \dots, y_m)}{\int p(y, y_2, \dots, y_m) dy} \right), \\
 &= p(y_1, y_2, \dots, y_m) \left(\frac{p(y'_1, y_2, \dots, y_m)}{\int p(y, y_2, \dots, y_m) dy} \right), \\
 &= p(y_1, y_2, \dots, y_m) \left(\frac{p(y'_1, y_2, \dots, y_m)}{\int p(y, y_2, \dots, y_m) dy} \right), \\
 &= p(y_1, y_2, \dots, y_m) f_{y_1|y_2, \dots, y_m}(y'_1|y_2, \dots, y_m), \\
 &= p_i q_{ij}. \tag{2-39}
 \end{aligned}$$

Así se tiene que, $\alpha(i, j) = 1$. Es decir, el estado propuesto es siempre aceptado. Lo mismo es cierto para cada uno de los m componentes. El algoritmo puede ser implementado bien sea, actualizando de forma sucesiva cada componente de la distribución m -dimensional, o seleccionando componentes para ser actualizadas de manera uniforme y aleatoria. Finalmente cabe decir, que el muestreador de Gibbs es ampliamente versátil y puede ser utilizado en una amplia variedad de problemas de alta complejidad, Dobrow (2016).

2.4.5. Algoritmo de la Esperanza Maximización (Expectation Maximization)

En la literatura estadística se han tratado problemas que presentan datos incompletos, es decir, que implican valores faltantes. El algoritmo (Esperanza-Maximización) a partir de ahora (**EM**), formalizado por Dempster et al. (1977), se ha convertido en una herramienta popular para dar tratamiento a estas dificultades. Muchos problemas de inferencia importantes en estadística, como los modelos de parámetros aleatorios y variables latentes, pueden ser tratados por medio del algoritmo **EM** cuando se formulan como problemas de valor perdido. En estas aplicaciones, no son los datos recopilados en campo los que en cierto sentido están perdidos, más bien, son los parámetros o variables ocultas los que son sistemáticamente no observados. En este trabajo se trata principalmente con este tipo de estructuras datos.

El uso generalizado de **EM** desde Dempster et al. (1977), es de un gran aporte. Los principales atractivos de este algoritmo iterativo radican en primera instancia a que es relativamente fácil de programar y que produce estimaciones de máxima verosimilitud (MLE). Sin embargo, también hay defectos bien documentados de **EM**, por ejemplo, que el algoritmo puede converger a máximos locales o puntos de silla de la función log-verosimilitud, y es a menudo sensible a los valores iniciales. En algunos modelos, el cálculo del **paso E** (E-step) implica integraciones de alta dimensión sobre subconjuntos de un espacio euclideo *k-dimensional*, que pueden ser intratables numéricamente.

El algoritmo EM

Por simplicidad supóngase que los datos *completos* que vienen representados por $x = \{x_1, \dots, x_n\}$, y constituyen realizaciones de variables aleatorias de una distribución con densidad $f(x|\Theta)$ donde Θ es un vector de parámetros.

Se denotará por $x = (W, Y)$, donde Y representa la parte de la muestra correspondiente a variables **no observables** y W representa la parte de la muestra correspondiente a las variables **observables**. Con esta notación, la log-verosimilitud correspondiente a los datos observados es:

$$\mathcal{L}_{\text{Obs}}(\Theta|W) = \log \int f(W, Y|\Theta) dy. \quad (2-40)$$

El cálculo y maximización de $\mathcal{L}_{\text{Obs}}(\Theta|y)$ es de alta complejidad en muchas situaciones de interés. En contraste la log-verosimilitud que corresponde al conjunto completo de datos:

$$\mathcal{L}_{\text{Comp}}(\Theta|x) = \log f(W, Y|\Theta) \quad (2-41)$$

Admite una expresión o fórmula explícita para el correspondiente estimador de máxima verosimilitud (MLE por sus siglas en inglés), $\hat{\Theta}_{\text{ML}}(W, Y)$.

Bajo el modelo (2-40), la información correspondiente al componente de la muestra **no observable** está representada por la densidad condicional $Y|W$,

$$k(Y|\Theta, W) = \frac{f(W, Y|\Theta)}{\int f(W, Y'|\theta) dY'} \quad (2-42)$$

La *idea básica* del algoritmo EM consiste en llevar a cabo la maximización de la log-verosimilitud correspondiente a la parte de la muestra observada $\mathcal{L}_{\text{Obs}}(\Theta|w)$, por medio de maximizaciones sucesivas en Θ de la esperanza condicional $Q(\Theta|\Theta^{(k)})$ de la log-verosimilitud correspondiente a los datos completos $\mathcal{L}_{\text{Comp}}(\Theta|x)$, dada la parte observada W y un ajuste previo $\Theta^{(k)}$ del parámetro,

$$\begin{aligned} Q(\Theta|\Theta^{(k)}) &= \mathbb{E}_{\Theta^{(k)}}[\mathcal{L}_{\text{Comp}}(\Theta|x)|W], \\ &= \int \mathcal{L}_{\text{Comp}}(\Theta|W, Y)k(Y|\Theta^{(k)}, W) dy. \end{aligned} \quad (2-43)$$

Así dado $\Theta^{(k)}$ a la k -ésima iteración del algoritmo EM, la $(k + 1)$ -ésima iteración se desarrolla en 2 pasos:

- **Paso E:** Calcule $Q(\Theta|\Theta^{(k)})$. En este paso se implementa el Muestreador de Gibbs.
- **Paso M:** Actualice $\Theta^{(k)}$, calculando el maximizador $\Theta^{(k+1)}$ de $Q(\Theta|\Theta^{(k)})$.

Este proceso iterativo es repetido hasta una alcanzar una convergencia del algoritmo. Para detalles en las propiedades de convergencia de la secuencia $\Theta^{(k)}$ ver (Dempster et al. (1977)) y (Wu (1983)).

2.4.6. Enfriamiento simulado

El algoritmo de *enfriamiento o recocido simulado* (Simulated Annealing Algorithm - SAA) pertenece a una clase de algoritmos de búsqueda local, también conocidos como algoritmos de Umbral. Hay dos razones por las cuales los algoritmos de umbral resultan interesantes dentro de los algoritmos de búsqueda local, Kirkpatrick et al. (1983) :

1. Trabajan de forma eficiente en una amplia gama de problemas.
2. Algunos algoritmos de umbral, como los de búsqueda local, tienen características que permiten hacer un análisis de la convergencia.
3. A diferencia de otros métodos de optimización no requieren del cálculo del gradiente, que en un funciones complejas como las que hay que maximizar en cada paso E del algoritmo EM pueden presentar dificultades.

El método de enfriamiento simulado es de utilidad para resolver problemas de optimización sin restricciones y con restricciones limitadas. El método modela el proceso físico de calentamiento de una partícula y luego baja lentamente la temperatura para disminuir los defectos, minimizando así la energía del sistema.

En cada iteración del algoritmo, se genera aleatoriamente un nuevo punto. La distancia del nuevo punto desde el punto actual, se basa en una distribución de probabilidad con una escala proporcional a la temperatura. El algoritmo acepta todos los puntos nuevos que maximizan el objetivo, por ejemplo una función de verosimilitud.

Al aceptar puntos que maximizan el objetivo, el algoritmo evita quedar atrapado en los máximos locales y puede explorar globalmente para encontrar más soluciones posibles. Se selecciona un programa de enfriamiento para disminuir sistemáticamente la temperatura a medida que avanza el algoritmo. A medida que disminuye la temperatura, el algoritmo reduce el alcance de su búsqueda para converger al máximo, Kirkpatrick et al. (1983).

2.5. Procesos vectoriales autorregresivos

En esta sección se describe brevemente la función de verosimilitud implementada para estimar los parámetros asociados a la componente observable del proceso, al no ser el objeto central de este trabajo se sugiere al lector ampliar los conceptos subyacentes a la teoría de series de tiempo esencialmente lo que concierne a procesos vectoriales autorregresivos para lo cual puede remitirse a Wei and Tanner (1990) y Hamilton and Press (1994).

2.5.1. Procesos autorregresivos de orden 1

Un proceso de la forma:

$$Y_t = \pi_1 Y_{t-1} + \varepsilon_t \quad (2-44)$$

Donde Y_{t-1} corresponde al primer rezago de Y_t , y ε_t es un ruido blanco, es decir, ε_t es una variable aleatoria que distribuye normal con media cero y varianza 1.

El proceso autorregresivo de orden 1, también se conoce como proceso de Markov porque la distribución de Y_t dado $Y_{t-1}, Y_{t-2}, Y_{t-2}, \dots$ es exactamente la misma distribución de Y_t dado Y_{t-1} . El proceso es invertible y para verificar su estacionariedad, la raíz del polinomio característico $(1 - \pi_1)B = 0$ debe estar por fuera del círculo unitario, es decir $|\pi_1| < 1$.

Proceso Vectorial AR(1)

Un proceso vectorial de orden 1 AR(1) esta dado por:

$$(\mathbf{I} - \mathbf{\Pi}_1 B)Y_t = \varepsilon_t \quad (2-45)$$

Para $m = 2$, se tiene que:

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} - \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \quad (2-46)$$

Donde la matriz de varianza covarianza Σ del proceso es de la forma:

$$\Sigma = E(\varepsilon_t \varepsilon_t^T) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \quad (2-47)$$

2.5.2. Estimación del proceso vectorial AR

La función de log-verosimilitud del proceso VAR se define como:

$$\mathcal{L}(\mathbf{\Pi}, \mathbf{\Sigma} | \mathbf{Y}) = -\frac{nm}{2} \log 2\pi - \frac{m}{2} \log |\Sigma_{\mu}| - \frac{1}{2} \text{tr} (\boldsymbol{\varepsilon}_t \Sigma_{\mu}^{-1} \boldsymbol{\varepsilon}_t^T) \quad (2-48)$$

Una vez se obtiene el estimador de máxima verosimilitud para el vector de parámetros del proceso, se deben validar supuestos de homocedasticidad, incorrelación y normalidad sobre los residuales con el fin de validar el supuesto de ruido blanco y verificar la consistencia del modelo para realizar inferencias o pronosticar. Para ampliar sobre algunos elementos utilizados en este proceso ver, Wei and Tanner (1990) y Hamilton and Press (1994).

3. Formulación del Modelo y Metodología de estimación

En línea con la exposición presentada en la introducción, se parte del modelo propuesto en Duffie et al. (2009), donde se considera un proceso de COX para la intensidad de default, el cual es log-lineal en sus parámetros. La tasa con la que se presentan incumplimientos, depende de un proceso autorregresivo determinado a partir de información observable y de un proceso estocástico de fragilidad no observable, especificado como un proceso Ornstein-Uhlenbeck con una componente de saltos (**OU**).

En este capítulo se presentará de forma más detallada la metodología de estimación propuesta en Duffie et al. (2009), más adelante se presenta la adaptación propuesta en este trabajo, motivada en una de las recomendaciones sugeridas por los autores del mencionado artículo. Se recuerda que la modificación propuesta consiste en modificar levemente el proceso **OU** utilizado para la fragilidad, incorporando un proceso de Poisson compuesto que perturba el proceso **OU** por medio de pulsos o saltos.

Para el proceso de estimación se implementó un programa en el software estadístico R, por medio del cual se intentó replicar el modelo propuesto en Duffie et al. (2009) con la modificación correspondiente que incorpora el proceso de saltos. Se buscará realizar la estimación conjunta de los parámetros asociados a estos procesos a partir del **algoritmo EM** y se utilizará el **muestreador de GIBBS** con el fin de generar las trayectorias asociadas al proceso **OU** con saltos que representa el factor latente.

Con respecto a la información requerida para la estimación de los parámetros del proceso; para la componente observable se utilizó información histórica, característica de cada individuo de una cartera o línea de crédito, de otra parte, se considera una covariable común a todos los individuos, consistente en información de indicadores de la economía, que de alguna manera se relacionan con el comportamiento de pago de los individuos en el sector financiero. Por simplicidad y tal como se propone en Duffie et al. (2009), el proceso que gobierna la dinámica de estas covariables observables es del tipo vectorial autorregresivo (**VAR**). Con el fin de obtener los estimadores de máxima verosimilitud de los parámetros asociados a este proceso se implementa un programa en R, es importante mencionar que en este trabajo se centra la atención sobre la estimación de los parámetros asociados al factor no observable y a la función de intensidad del Proceso de Cox.

3.1. Proceso de COX o Poisson doblemente estocástico

A partir de ahora, se considera una cartera de crédito constituida por m prestatarios o compañías sobre un horizonte de tiempo determinado T . La principal variable de estudio es el tiempo aleatorio τ , que representa el tiempo de *default* de un individuo en relación al compromiso de crédito, de tal forma que la probabilidad de default antes del instante t es $P(\tau \leq t)$.

El instante de incumplimiento de un solo individuo, por ejemplo el i -ésimo entre una cohorte de m prestamistas, es modelado como el instante del primer salto o evento de un proceso de Poisson $N_i(t)$ con tasa $\lambda_{i,t}$ que a su vez constituye un proceso de Markov, especificado de la siguiente manera:

$$\lambda_{i,t} = \exp(\beta_0 + \beta_1 V_t + \beta_2 U_{i,t} + \eta Y_t). \quad (3-1)$$

Acá;

- V_t Es un factor macroeconómico común a todos los individuos, (ver sección 5.1.3).
- $U_{i,t}$ Corresponde a la distancia a default del i -ésimo individuo (se explica en la sección 5.1.3).
- Y_t Representa el factor **no observable** o **latente**, denominado factor de fragilidad y es común a todo individuo.

Con respecto al factor de riesgo adicional Y_t , no observable o latente y al que todos los individuos se encuentran conjuntamente expuestos, cabe decir que se conoce como factor de *fragilidad*. En la literatura estadística de pronóstico de eventos al efecto de una covariable latente o no observable se le denomina ***fragilidad***.

De acuerdo con la sección 1.1 los tiempos de default τ_i también pueden ser caracterizados por medio de procesos indicadores $D_{i,t}$, $i = 1, \dots, m$, dónde:

$$D_{i,t} = \begin{cases} 0 & t < \tau_i \\ 1 & \tau_i \leq t. \end{cases} \quad (3-2)$$

Es decir el proceso indicador de default $D_{i,t}$ salta de “cero” 0 a “uno” 1 justo en el instante del incumplimiento.

Con el fin de simplificar la notación, se asume que $\lambda_{it} = \Lambda(S_i(X_t), \Theta)$, donde $S_i(X_t) = (W_{it}, Y_t)$, $W_{it} = (U_{i,t}, V_i)$ corresponde al vector de covariables observables para el i -ésimo prestatario y $\lambda_{i,t}$ queda expresado por:

$$\Lambda((w, y), \Theta) = e^{\beta_1 w_1 + \dots + \beta_n w_n + \eta y}, \quad (3-3)$$

Para un vector de parámetros $\Theta = (\beta, \eta, \kappa, \lambda_\xi, \nu)$, donde β es el vector de parámetros asociado al proceso W y $(\eta, \kappa, \lambda_\xi, \nu)$ corresponden a los parámetros del proceso OU con saltos Y . Se supone que el factor de fragilidad Y_t es en sí mismo un proceso de Markov en tiempo discreto.

En lo sucesivo se referirá de ahora en adelante al proceso de Poisson $\mathbf{N}(t)$ que consiste en la suma de m procesos de Poisson condicionalmente independientes, dadas las realizaciones de los procesos $V_t, U_{i,t}, i = 1, \dots, m$ y Y_t , como modelo \mathbf{M} . El proceso de Poisson subyacente al modelo \mathbf{M} caracteriza de esta forma la distribución del número de incumplimientos en la cohorte de m prestatarios en el tiempo.

El tiempo de default τ_i de cualquiera de los m prestatarios, como se había mencionado, corresponde al instante del primer evento del proceso de Poisson $N_i(t)$ con intensidad $\lambda_{i,t}$ como en (1-1), es decir:

$$\{\tau_i \leq t\} \iff \{N_i(t) \geq 1\}. \quad (3-4)$$

y así:

$$P(\tau_i \leq t) = P(N_i(t) \geq 1) = \int P(N_i(t) \geq 1 | V, U_i, Y) dP(V, U_i, Y). \quad (3-5)$$

$$= \mathbb{E}^{(V, U_i, Y)} [P(N_i(t) \geq 1 | V, U_i, Y)]$$

$$= \mathbb{E}^{(V, U_i, Y)} [1 - e^{-\int_0^t \lambda_{i,s} ds}].$$

(3-6)

La medida $dP(V, U_i, Y)$ corresponde a la medida conjunta de los tres procesos en el respectivo espacio de trayectorias.

De (3-4) también se deduce que:

$$\begin{aligned}
P(\tau_i \leq t | V, U_i, Y) &= P(N_i(t) \geq 1 | V, U_i, Y) \\
&= 1 - e^{-\int_0^t \lambda_{i,s} ds}, \\
&= 1 - e^{-\int_0^t e^{\beta_0 + \beta_1 V_s + \beta_2 U_{i,s} + \eta Y_s} ds}.
\end{aligned}
\tag{3-7}$$

3.2. Proceso de Estimación

Con el fin de realizar la estimación de los parámetros del modelo **M** descrito en la sección anterior, se sigue casi literalmente la metodología propuesta por Duffie et al. (2009), que se ha utilizado para el ajuste del modelo **D**, esto principalmente porque el modelo de interés (modelo **M**) corresponde a una ligera modificación del modelo que estos investigadores estudian en el mencionado artículo. Según la revisión exploratoria de la literatura, en el estado actual de conocimiento la metodología que será descrita es la única conocida para resolver el problema de estimación, dado que el modelo **M** constituye un proceso mezcla de Procesos de Poisson pero la variable de mezcla es también un proceso estocástico representado por el proceso Y_t .

El método de estimación, se describe a continuación en los siguientes pasos:

3.2.1. Planteamiento de la Función de Verosimilitud

En primera instancia se considera que el vector de parámetros que caracteriza la especificación del modelo **M** es:

$$(\gamma, \beta_0, \beta_1, \beta_2, \eta, \kappa, \lambda_\xi, \nu) \tag{3-8}$$

El parámetro γ es a su vez un vector de parámetros que caracteriza la dinámica conjunta de los procesos exógenos $V_t, (U_{i,t}, i = 1, \dots, m)$ y se estima por método estándar de máxima verosimilitud de la teoría de series de tiempo, pues estos procesos por simplicidad se modelan como procesos autorregresivos de orden 1 Gaussianos.

El objetivo de este trabajo se centra en la estimación de los parámetros restantes.

$$\Theta = (\beta_0, \beta_1, \beta_2, \eta, \kappa, \lambda_\xi, \nu) \tag{3-9}$$

Recuerde que:

$\beta_0, \beta_1, \beta_2, \eta$ parametrizan la función de intensidad log-lineal en (1-1)
 κ corresponde a la tasa de reversión a la media en el proceso (OU) en (1-3)
 λ_ξ, ν corresponden a la frecuencia y desviación estándar de los pulsos del componente de saltos en (1-3).

La función de distribución conjunta en el espacio de trayectorias de los procesos:

$$\mathbf{D}_t = (D_{1,t}, \dots, D_{m,t}), \mathbf{W}_t = (V_t, U_{1,t}, \dots, U_{m,t}), Y_t. \quad (3-10)$$

Se representará por:

$$P(\mathbf{D}, \mathbf{W}, Y; \gamma, \Theta) \quad (3-11)$$

Recordando que el proceso:

$\mathbf{W}_t = (V_t, U_{1,t}, \dots, U_{m,t})$ es observable y por supuesto también lo es: $\mathbf{D}_t = (D_{1,t}, \dots, D_{m,t})$.

Se considerará la siguiente función de verosimilitud para Θ , basada en \mathbf{W} , Y y \mathbf{D} :

$$L(\Theta; \mathbf{D}|Y, \mathbf{W}) \equiv P(\mathbf{D}|Y, \mathbf{W}, \Theta), \quad (3-12)$$

también la función de verosimilitud para γ

$$L(\gamma; \mathbf{W}) \equiv P(\mathbf{W}|\gamma), \quad (3-13)$$

Y la función de verosimilitud para (γ, Θ)

$$\begin{aligned}
 L(\gamma, \Theta; \mathbf{D}, \mathbf{W}|Y) &\equiv P(\mathbf{D}, \mathbf{W}|Y, \gamma, \Theta), \\
 &= P(\mathbf{D}|\mathbf{W}; Y, \Theta)P(\mathbf{W}|\gamma) \\
 &= L(\Theta; \mathbf{D}|Y, \mathbf{W})L(\gamma; \mathbf{W})
 \end{aligned} \quad (3-14)$$

Observe que la factorización en la segunda línea en (3-14), simplemente responde a la regla del producto usual de que una distribución conjunta se obtiene a partir de una condicional por una marginal. Además se tiene en cuenta que \mathbf{W} solo depende del vector de parámetros γ . También, condicional en \mathbf{W} , los procesos \mathbf{D} y Y sólo dependen del parámetro Θ .

La factorización en (3-14) también permite encontrar los estimadores de máxima verosimilitud de γ y Θ por separado, porque la **Log-verosimilitud** corresponde a la suma de 2 términos, uno que

depende sólo de γ y otro que depende solo de Θ .

Como el logaritmo de la verosimilitud en la tercera línea de (3-14) corresponde a la suma de términos separados que envuelven los parámetros γ y θ , se puede descomponer la verosimilitud en dos problemas, Duffie et al. (2007):

$$\begin{aligned} \sup_{\gamma} L(\gamma; W), \quad y \\ \sup_{\Theta} L(\Theta; D|W, Y) \end{aligned} \quad (3-15)$$

Proposición 5. *Los estimadores de máxima verosimilitud γ y Θ hallados para 3-15 forman una solución para*

$$\sup_{\gamma, \Theta} L(\gamma, \Theta; D, W|Y) \quad (3-16)$$

El factor $L(\gamma; W)$ corresponde a la verosimilitud de procesos autorregresivos de orden 1 Gaussianos, el vector de parámetros γ determina el comportamiento de la serie de tiempo asociada a las covariables observables inmersas en el vector de estados X_t , Duffie et al. (2007). Nuevamente se hace hincapié en que la estimación del vector de parámetros γ no es el objeto central de este trabajo, en cambio el interés se centra en el estimador de máxima verosimilitud para Θ para el cual se considerará la verosimilitud $L(\Theta; D|W, Y) \equiv P(D|W, Y, \Theta)$.

En relación a la función de verosimilitud $L(\Theta; D|W, Y) = P(D|W, Y, \Theta)$, se tiene que:

1. Cuando se tiene información **completa** (Y observable), se tiene que

$$\begin{aligned} L(\Theta; D|W, Y) &= P(D|W, Y, \Theta), \\ &= \prod_{i=1}^m \left(e^{-\sum_{t=t_i}^{\tau_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{\tau_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})] \right). \end{aligned} \quad (3-17)$$

Para entender el porqué de (3-17), obsérvese primero, que $P(D|W, Y, \Theta)$ es la distribución conjunta de los procesos indicadores D_{it} que caracterizan cada uno el primer salto del Proceso de Poisson $N_i(t)$ con intensidad λ_{it} . Así considerando una partición Δt del tiempo, se puede aproximar la distribución $P(D_i|W, Y, \Theta)$ (que corresponde a la probabilidad de la realización del proceso D_{it}) por:

$$P(D_i|W, Y, \Theta) \approx e^{-\lambda_{i,t_1} \Delta t} e^{-\lambda_{i,t_2} \Delta t} \dots e^{-\lambda_{i,\tau_i} \Delta t} \lambda_{i,\tau_i} \Delta t, \quad (3-18)$$

$$= e^{-\sum_{t=t_1}^{\tau_i} \lambda_{it} \Delta t} \prod_{t=t_1}^{\tau_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})] \quad (3-19)$$

Recuérdese que:

$$D_{it} = \begin{cases} 0 & \text{si } t < \tau_i \\ 1 & \text{si } \tau_i \leq t. \end{cases}$$

Para entender mejor (3-18) obsérvese que, $e^{-\lambda_i t_j} \Delta t \approx e^{-\int_{t_j}^{t_j+\Delta t} \lambda_{i,s} ds}$ aproxima la probabilidad de que el Proceso de Poisson $N_i(t)$ no salte en el intervalo $[t_j, t_j + \Delta t) = [t_j, t_{j+1})$.

De manera similar $e^{-\lambda_i \tau_i} \Delta t \approx e^{-\int_{\tau_i-\Delta t}^{\tau_i} \lambda_{i,s} ds}$ aproxima la probabilidad de que el Proceso de Poisson $N_i(t)$ salte en el intervalo $[\tau_i - \Delta t, \tau_i]$.

Por lo tanto la expresión en el lado derecho de (3-18)

$$e^{-\lambda_{i,t_1} \Delta t} e^{-\lambda_{i,t_2} \Delta t} \dots e^{-\lambda_{i,\tau_i} \Delta t} \lambda_{i,\tau_i} \Delta t,$$

Aproxima la probabilidad de la realización del proceso D_{it} , $P(D_i|W, Y, \Theta)$ que caracteriza el tiempo hasta el default del i -ésimo prestatario.

Para la corte de m prestatarios los procesos indicadores D_{it} , $i = 1, \dots, m$ son condicionalmente independientes dados \mathbf{W} , \mathbf{Y} y por lo tanto con información completa:

$$\begin{aligned} L(\Theta; D|W, Y) &= P(D|W, Y, \Theta), \\ &= \prod_{i=1}^m P(D_i|W, Y, \Theta), \\ &= \prod_{i=1}^m \left(e^{-\sum_{t=t_i}^{\tau_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{\tau_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})] \right). \end{aligned} \quad (3-20)$$

Con información **incompleta**, es decir, en el caso en que solo se observan D y W en vez de la verosimilitud $L(\Theta; D|W, Y)$, se puede considerar la siguiente log-verosimilitud esperada:

$$\begin{aligned} L(\Theta; D|W) &= \mathbb{E}^Y [L(\Theta; D|W, Y)], \\ &= \int L(\Theta; D|W, y) P_Y(y|\Theta) dy, \\ &= E^Y \left[\prod_{i=1}^m \left(e^{-\sum_{t=t_i}^{\tau_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{\tau_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})] \right) \right], \end{aligned} \quad (3-21)$$

Donde el valor esperado en (3-21) es con respecto a la densidad a priori del proceso Y , $P_Y(y|\Theta)$.

El cálculo de (3-21) por simulación directa implica integración Monte Carlo en un espacio de alta dimensión. Este es un proceso computacionalmente intenso si se lleva a cabo por medio de Monte Carlo básico teniendo en cuenta que se debe optimizar con respecto al vector de parámetros Θ . La metodología sugerida por la literatura consiste en conseguir el estimador de máxima verosimilitud (MLE) de Θ por medio del algoritmo EM (Esperanza Maximización), en el cual, en lugar de la

verosimilitud $L(\Theta; D|W, Y)$ se considera la maximización en Θ de la siguiente log-verosimilitud esperada:

$$\begin{aligned}
Q(\Theta|\hat{\Theta}^{(k)}) &\equiv \mathcal{L}(\Theta; D|W), \\
&= \int \log L(\Theta; D|W, y) P_Y(y|D, W, \hat{\Theta}^{(k)}) dy, \\
&= \mathbb{E}^{Y|D, W, \hat{\Theta}^{(k)}} [L(\Theta; D|W, Y)], \\
&= \mathbb{E}^{Y|D, W, \hat{\Theta}^{(k)}} \left[\prod_{i=1}^m \left(e^{-\sum_{t=t_i}^{\tau_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{\tau_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})] \right) \right], \quad (3-22)
\end{aligned}$$

Donde esta vez el valor esperado en (3-22) es con respecto a la distribución **posteriori** (Es decir teniendo en cuenta los datos observados D, W) del proceso Y , $P_Y(y|D, W, \hat{\Theta}^{(k)})$. Para muestrear de esta distribución se emplea el Muestreador de Gibbs, $\hat{\Theta}^{(k)}$ corresponde a un ajuste previo del parámetro Θ . El algoritmo EM se explicará de forma más amplia a continuación, la exposición del Muestreador de Gibbs se hará más adelante.

3.2.2. Esperanza Maximización

El algoritmo EM consta de dos pasos:

- **Paso Esperanza** Dado $\hat{\Theta}^{(k)}$, mediante el muestreador de Gibbs, se simula una muestra $\{Y^j\}$ de la distribución $P_Y(y; D, W, \hat{\Theta}^{(k)})$ y se estima:

$$Q(\Theta|\hat{\Theta}^{(k)}) = \int \log L(\Theta; D, y|W) P_Y(y|D, W, \hat{\Theta}^{(k)}) dy \quad (3-23)$$

A partir de la ley de los grandes números, y el teorema 3 visto en la sección 2.4.2, del cual se deduce que para la distribución estacionaria $P_Y(y|D, W, \hat{\Theta}^{(k)})$, el estimador de Monte Carlo para el valor esperado en (3-23), se puede aproximar numéricamente de la siguiente manera:

$$\hat{Q}(\Theta|\hat{\Theta}^{(k)}) = \frac{1}{N} \sum_{j=1}^N \log L(\Theta; D, Y^j|W),$$

- **Paso Maximización.**

Maximizar en Θ la función $\hat{Q}(\Theta|\hat{\Theta}^{(k)})$, este paso es numéricamente intenso porque para cada iteración del algoritmo de optimización que se esté utilizando hay que realizar un **paso esperanza**.

Luego de optimizar se tiene:

$$\hat{\Theta}^{(k+1)} = \arg \max_{\Theta} \hat{Q}(\Theta | \hat{\Theta}^{(k)}). \quad (3-24)$$

Con éxito la secuencia $(\Theta^{(k)})$ converge al estimador de máxima verosimilitud de $\mathcal{L}(\Theta; D|W)$.

3.2.3. Muestreador de Gibbs para generar trayectorias del proceso de fragilidad.

Con el fin de completar el paso **E-step** y muestrear las trayectorias para el proceso Y se considera nuevamente la densidad $P_Y(\cdot | W, D, \Theta)$. Una dificultad que debe enfrentarse es que la distribución de las trayectorias de Y , $P_Y(\cdot | W, D, \Theta)$ está soportada sobre un espacio de alta dimensión.

Sin embargo, mediante métodos numéricos como MCMC se puede aproximar la distribución posterior del proceso $P_Y(\cdot | W, D, \Theta)$, simulando una cadena de Markov con densidad de equilibrio dada por $P_Y(\cdot | W, D, \Theta)$.

La estimación vía simulación MonteCarlo para una función $f(\cdot)$ donde su valor esperado esta dado por:

$$\mathbb{E} [f(Y) | W, D, \Theta] = \int f(y) P_Y(y | W, D, \Theta) dy, \quad (3-25)$$

Se puede estimar por medio de:

$$= \frac{1}{N} \sum_{j=1}^N f(Y^{(j)}) \quad (3-26)$$

Acá: $Y^j = (Y_0^j, \dots, Y_T^j)$

1. La idea del muestreador de Gibbs en este caso, consiste en:

Dada una realización (estado) $Y^{(j-1)} = (Y_0^{(j-1)}, Y_1^{(j-1)}, \dots, Y_T^{(j-1)})$, de la cadena, se itera sobre las componentes de $Y^{(j-1)}$ y por medio del algoritmo de **Metropolis-Hastings** se generan nuevas componentes $Y_t^{(j)}$ simulando de la distribución condicional.

$$\begin{aligned} Y_t^j | Y_0^{(j-1)}, Y_1^{(j-1)}, \dots, Y_{t-1}^{(j-1)}, Y_{t+1}^{(j-1)}, \dots, Y_T^{(j-1)}, \\ \equiv Y_t^j | Y_{(-t)}^{j-1}. \end{aligned} \quad (3-27)$$

Obteniendo de esta forma la siguiente realización o estado $Y^{(j)}$.

2. Una cantidad central para estimar la dinámica del histórico de incumplimientos, viene dada por la distribución posteriori $P(Y_t^j | W, D, Y_{(-t)}^{j-1}, \Theta)$. Observe entonces que esta distribución puede expresarse como sigue para efectos de muestreo, (ver pág 2119 en Duffie et al. (2009)):

$$\begin{aligned} P(Y_t | W, D, Y_{(-t)}, \theta) &= \frac{P(W, D, Y, \Theta)}{P_Y(W, D, Y_{(-t)}, \Theta)} \\ &\propto P(W, D, Y, \Theta) \\ &= P(W, D | Y, \Theta)P(Y, \Theta) \\ &= P(D|W, Y, \Theta)P(W|Y, \Theta)P(Y, \Theta) \end{aligned} \quad (3-28)$$

$$\begin{aligned} &\propto P(D|W, Y, \Theta)P(Y, \Theta), \\ &= L(\Theta; D|W, Y)P(Y, \Theta), \\ &= L(\Theta; D|W, Y)P(Y_t|Y_{(-t)}, \Theta)P(Y_{(-t)}, \Theta), \\ &= L(\Theta; D|W, Y)P(Y_t|Y_{(-t)}, \Theta), \end{aligned} \quad (3-29)$$

Donde el paso en (3-28) a la siguiente línea se explica porque W no depende de Y ni de Θ y en consecuencia $P(W|Y, \Theta)$ es igual a $P(W)$ y por tanto puede ser descartado.

Ahora veamos que para el factor $P(Y_t|Y_{(-t)}, \Theta)$ se tiene que:

$$P(Y_t|Y_{(-t)}, \Theta) \propto P(Y_t|Y_{(t-1)}, \Theta)P(Y_{(t+1)}|Y_t, \Theta), \quad (3-30)$$

en efecto, denotando por $P_{j,j+1} = P(Y_{j+1}|Y_j, \Theta)$, se tiene:

$$\begin{aligned} p(Y_t|Y_{(-t)}, \Theta) &= \frac{P(Y_0, Y_1, \dots, Y_{t-1}, Y_t, Y_{t+1}, \dots, Y_T | \Theta)}{P(Y_0, Y_1, \dots, Y_{t-1}, Y_{t+1}, \dots, Y_T | \Theta)}, \\ &= \frac{P(Y_0 | \Theta)P_{0,1}P_{1,2} \cdots P_{t-2,t-1}P_{t-1,t}P_{t,t+1}P_{t+1,t+2} \cdots P_{T-1,T}}{P(Y_0 | \Theta)P_{0,1}P_{1,2} \cdots P_{t-2,t-1}P_{t-1,t+1}P_{t+1,t+2} \cdots P_{T-1,T}}, \\ &= \frac{P_{t-1,t}P_{t,t+1}}{P_{t-1,t+1}}, \\ &\propto P(Y_t|Y_{t-1}, \Theta)P(Y_{t+1}|Y_t, \Theta). \end{aligned} \quad (3-31)$$

Finalmente sustituyendo (3-30) en (3-29) se obtiene que

$$P(Y_t | W, D, Y_{(-t)}, \Theta) \propto L(\Theta, D|W, Y)P(Y_t|Y_{t-1}, \Theta)P(Y_{t+1}|Y_t, \Theta). \quad (3-32)$$

Dónde $P(Y_t | Y_{t-1}, \Theta)$ es la densidad de transición de un paso del proceso **OU** con una componente de saltos. Esta última expresión en (3-32) corresponde a la densidad sobre la cual se desea muestrear trayectorias por medio del sampler de Gibbs. Recuerdese además que:

$$L(\Theta, D|W, Y) = \prod_{i=1}^m \left(e^{-\sum_{t=t_i}^{\tau_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{\tau_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})] \right). \quad (3-33)$$

A partir de este resultado se explica a continuación como se implementa el Muestreador de Gibbs para generar trayectorias que consideren la componente de saltos para el proceso de fragilidad propuesto.

Modificación en el muestreador de Gibbs para tomar cuenta de la componente de saltos

Hasta el momento se ha descrito la idea general del método de estimación propuesto en Duffie et al. (2009) basado en el muestreador de Gibbs y en el algoritmo de Esperanza-Maximización. Ahora se explicará la modificación que se ha planteado en el muestreador de Gibbs para considerar el ajuste de la componente de saltos del proceso (**OU**) en (1-3).

En dicho artículo, en la descripción del muestreador de Gibbs para la simulación de la distribución del proceso Y se propone utilizar la siguiente distribución¹, como distribución de propuesta para el muestreador (ver página 2119 de Duffie et al. (2009)):

$$q(Y_t^{(j)} | Y^{(j-1)}, \hat{\Theta}^{(k)}) = N(Y_t^{(j-1)}, 4), \quad (3-34)$$

La modificación que se ha sugerido para la distribución de propuesta en (3-34) es la siguiente:

- Se simula variable $U \sim \text{Bernoulli}(p)$, es decir:

$$U = \begin{cases} 1 & \text{con prob } p \\ 0 & \text{con prob } 1 - p \end{cases} \quad \text{con } p = 1 - \exp(-\xi \Delta t), \quad (3-35)$$

p aproxima la probabilidad de que se presente al menos un salto de Poisson con tasa ξ en un intervalo de tiempo pequeño Δt .

¹El muestreador de Gibbs consiste en una iteración del algoritmo de Metropolis-Hastings. En este algoritmo siempre hay una función de aceptación.

- Entonces la distribución de propuesta considerada es:

$$q(Y_t^{(j)} | Y^{(j-1), \hat{\Theta}^{(k)}}) = N(Y_t^{(j-1)} + U \cdot Z, 4), \quad (3-36)$$

Dónde Z es una variable aleatoria normal que **simula** la magnitud de los saltos, independiente del movimiento Browniano B_t en (1-3), donde la distribución de los saltos ξ_i corresponde a $Z \sim N(0, \nu^2)$.

La idea es que simplemente se modifica aditivamente la media de la distribución propuesta, pero cabe resaltar que solo en la medida en que los pulsos sean frecuentes y el nivel en que se cambia dicha media corresponde a la magnitud de un posible pulso.

$$q(Y_t^j | W, D, Y^{j-1}, \Theta) = N(Y_t^{(j-1)} + U \cdot Z, 4), \quad (3-37)$$

Cada trayectoria se genera de una media igual a la anterior observación del proceso, obtenida a partir del muestreador de Gibbs y una varianza igual a 2 veces los incrementos de un Movimiento Browniano estándar.

Es de resaltar que frente a la idea propuesta en Duffie et al. (2009), donde se considera un Proceso OU estándar, se intenta capturar los pulsos de la función de intensidad modificando la media pero garantizando que el proceso sigue siendo estacionario. Lo anterior fue verificado numéricamente de forma tal que la convergencia del algoritmo no se viera afectada por presencia de valores extremos o colas pesadas en la densidad propuesta obteniendo los resultados que se presentan en el próximo capítulo.

Finalmente utilizando el algoritmo Metropolis-Hastings se muestrea Y_t en la j -ésima iteración del muestreador de Gibbs, el algoritmo se explica a continuación:

1. Muestrear un candidato tal que:

$$Y_t^j = \begin{cases} y \sim N(Y_t^{(j-1)} + U \cdot Z, 4) & \text{si } U = 1 \\ y \sim N(Y_t^{(j-1)}, 4) & \text{si } U = 0 \end{cases} \quad (3-38)$$

2. Luego se evalúa cada trayectoria para la fragilidad en la siguiente función de aceptación:

$$\alpha(y, Y_t^j) = \min \left(\frac{L(\Theta; D | W, Y_{-t}^{(j-1)}, Y_t = y)}{L(\Theta; D | W, Y^{(j-1)})}, 1 \right) \quad (3-39)$$

3. Se recurre a la distribución uniforme para muestrear un número C entre $(0, 1)$ y considerar:

$$Y_t^j = \begin{cases} y & \text{si } C \leq \alpha(y, Y_t^j) \\ Y_t^{(j-1)} & \text{en otro caso.} \end{cases} \quad (3-40)$$

3.2.4. Modelo VAR para las covariables observables

Considere $R(T_k) = \{k : \tau_i > T_k\}$ los individuos que no han entrado en default antes de T_k , se tiene que el componente de la función de verosimilitud totalmente observable se puede expresar como sigue:

$$\mathcal{L}(\gamma | W) = \prod_{t_k=1}^{t_m} f_{R(t_{k+1}), R(t_k)}(W_{t_{k+1}} | W_{t_k}; \gamma) \quad (3-41)$$

Por practicidad se considera un proceso Gaussiano autorregresivo de primer orden debido a que el proceso de Markov X_t no tiene memoria más allá de un paso.

Según lo propuesto por Duffie et al. (2007), se construirá el vector de covariables V_t a partir de dos variables macroeconómicas, se define a r_{1t} como las tasas de interés, entre tanto r_{2t} corresponde a la variación mensual del PIB. Luego el proceso tiene la siguiente estructura:

$$r_{t+1} = r_t + \beta_r(\theta_r - r_t) + C_r \varepsilon_{t+1} \quad (3-42)$$

Dónde: $\varepsilon_t \sim N(0, 1)$ y C_r es una matriz triangular inferior de 2×2 .

Para el vector de covariables específicas W_{it} se consideran la distancia a default δ_{it} y los activos logarítmicos de cada firma U_{it} , expresadas por:

$$\begin{bmatrix} \delta_{i,t+1} \\ U_{i,t+1} \end{bmatrix} = \begin{bmatrix} \delta_{it} \\ U_{it} \end{bmatrix} + \begin{bmatrix} \beta_\delta & 0 \\ 0 & \beta_U \end{bmatrix} \left(\begin{bmatrix} \theta_{i\delta} \\ \theta_{iU} \end{bmatrix} - \begin{bmatrix} \delta_{it} \\ U_{it} \end{bmatrix} \right) + \begin{bmatrix} b(\theta_r - r_t) \\ 0 \end{bmatrix} + \begin{bmatrix} \sigma_\delta & 0 \\ 0 & \sigma_U \end{bmatrix} \zeta_{i,t+1} \quad (3-43)$$

Dónde: $\zeta_{it} = Az_{it} + Bw_t$

Con respecto a la distancia a default (δ_{it}) cabe decir, que de acuerdo a lo propuesto por Vassalou and Xing 2004, Shumway 2001 y Hillegeist et al. (2004), es construída con base en el número de desviaciones estándar que el logaritmo de los activos de cada compañía excede el valor de sus pasivos (deuda). Lo anterior se puede expresar de la siguiente forma:

$$\delta_{it} = \frac{\ln\left(\frac{V_{it}}{L_{it}}\right) + (\mu_A - \frac{1}{2}\sigma_A^2 T)}{\sigma_A \sqrt{T}}, \quad (3-44)$$

Dónde V_t corresponde al valor de los activos de cada individuo, y L_t el valor de sus deudas en un tiempo t . Por su parte μ_A y σ_A corresponden a la tasa media del crecimiento y la volatilidad de los activos. Para ampliar remitirse al numeral B1 del apéndice B.

Se supondrá independencia entre los errores ε_t y ζ de los modelos econométricos planteados en (3-42) y (3-43). Se puede ver entonces que el vector de parámetros (γ) subyacente a la serie de tiempo del proceso X_t en (3.1) quedaría finalmente dado por los parámetros estimados a partir de un proceso vectorial autorregresivo de primer orden VAR(1).

4. Implementación en una entidad financiera colombiana

Los trabajos de investigación sobre la medición de riesgos son una actividad que en muy poco tiempo ha experimentado un crecimiento significativo. Tanto es así, que a principios de los años 90 no existía, de forma generalizada lo que hoy se llama “Sistemas de administración y Gestión Riesgos”. Mediante el esfuerzo de profesionales del sector financiero, los centros académicos y las empresas de consultoría financiera, se dispone hoy de una rica variedad de modelos y sistemas para la medición y control de riesgos financieros. Actualmente y de manera rutinaria, los sistemas implantados en bancos, sociedades financieras y departamentos de tesorería de grandes empresas no financieras, generan medidas periódicas de los riesgos financieros que son utilizadas como una importante herramienta de gestión administrativa.

Colombia no ha estado ajena estos desarrollos, la Superintendencia Financiera de Colombia (SFC), como ente regulador de las instituciones financieras y adoptando los acuerdos de Basilea, ha emitido una serie de normas que deben cumplir las instituciones sometidas a la inspección y vigilancia de la SFC en materia de administración de riesgos de mercado (SARM), de riesgos crédito (SARC), de riesgos liquidez y riesgos operativos (SARO). Dichos requerimientos están relacionados en la Circular Básica y Contable Financiera (Circular Externa 100 de 1995) los cuales deben ser implementados por las instituciones financieras. Por otra parte, las instituciones deben medir y reportar periódicamente el valor en Riesgo (VaR) de sus operaciones (calculado como mínimo mediante el modelo estándar) y de los niveles de solvencia como una función del valor en riesgo calculado.

El Instituto Colombiano de Normas Técnicas emite la norma técnica NTC-5254 en agosto de 2006, la cual proporciona una guía genérica para la administración y gestión de riesgos. Dicha norma recoge todos los lineamientos del estándar australiano AS/NZS 4360 de 1999 y puede ser aplicada a cualquier industria o sector económico.

A pesar que la norma técnica permite a las entidades financieras dar un paso muy importante en la gestión de riesgos, ya que da las directrices para identificar y medir su impacto en términos económicos, los desarrollos en medición cuantitativa de riesgos plantean ir mas allá, en este sentido, para tener estimaciones razonables de la probabilidad de incumplimiento crediticio en un segmento de negocio, es necesario contar con modelos estructurales más robustos, que a partir de la información conocida de cada individuo permitan incorporar otro tipo de variables y con esto mejorar la

calidad de las estimaciones. El objetivo es mejorar la provisión de la cartera en riesgo, de forma tal que se minimice el error de sobre estimación u subestimación sobre ésta, ya que las implicaciones de incurrir reiteradamente en éste error se reflejan en mayores costos e ineficiencias. Por un lado implicaría la pérdida de oportunidades de negocios por una provisión mucho mayor a la adecuada, y por el otro incrementaría la exposición a pérdidas operacionales, si esta provisión está por debajo del riesgo real de la cartera.

De acuerdo con los argumentos expuestos, se pretende entonces estimar los parámetros asociados a la intensidad de Cox expuesta en el capítulo 3, con el fin de ajustar la intensidad a partir de la cual más adelante se pueden estimar las probabilidades de incumplimiento en un segmento de cartera masiva de una entidad financiera colombiana.

4.1. Descripción y tratamiento de los datos

Se tomaron los datos mensuales de aproximadamente tres años (35 meses) de individuos pertenecientes al segmento de cartera masiva en una entidad financiera. La base de datos utilizada contiene información idiosincrática o personal sobre la evolución de la situación financiera de cada prestatario. Las variables que hacen parte de la información primaria son: Nivel de ingresos, egresos, activos, endeudamiento con el sector real, endeudamiento con entidades financieras y el vector de altura de mora de 1 a 12 meses.

Fecha Cot	Cliente	Total Acti	Total Pasiv	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂
ago-16	Cliente 61	27.350.000	1.430.000	-	15	15	15	-	14	-	-	-	-	-	-
ago-16	Cliente 62	124.900.000	4.256.000	-	-	-	-	-	-	11	11	-	-	-	-
ago-16	Cliente 67	49.550.000	5.482.000	-	-	-	-	-	-	-	-	-	-	-	-
sep-16	Cliente 240	104.250.000	40.462.000	-	-	-	5	5	5	-	-	5	-	-	-
sep-16	Cliente 275	13.249.450	1.193.000	-	89	59	29	29	-	-	-	-	-	-	-
sep-16	Cliente 309	65.600.000	2.205.000	-	5	5	5	5	-	-	-	-	-	-	-
sep-16	Cliente 312	33.965.000	8.647.000	-	29	59	149	119	89	58	29	29	29	-	-
oct-16	Cliente 332	34.900.000	8.494.000	71	41	11	-	131	101	71	40	11	41	11	-
oct-16	Cliente 337	40.800.000	23.231.000	-	46	16	-	16	-	16	-	16	-	-	-
oct-16	Cliente 338	19.200.000	13.607.000	-	-	-	-	-	-	-	-	-	-	-	-
oct-16	Cliente 339	22.030.000	16.503.000	3	-	3	3	3	3	2	-	3	-	-	-
oct-16	Cliente 341	21.525.000	9.345.000	-	-	-	60	30	-	30	-	-	-	-	-
abr-19	Cliente 1886	1.148.000	607.000	-	-	9	9	9	-	-	-	9	-	-	-
abr-19	Cliente 2272	15.905.000	5.276.000	-	-	20	-	20	20	20	-	-	-	-	-
abr-19	Cliente 5027	77.480.000	11.104.000	-	-	-	-	46	16	16	-	-	-	16	-
abr-19	Cliente 3363	55.590.000	23.749.000	-	-	-	-	-	4	-	-	-	4	4	-
abr-19	Cliente 461	30.256.000	5.367.000	-	16	-	-	16	-	16	-	-	-	16	-
may-19	Cliente 4392	52.390.000,00	11.200.000	14	14	14	0	0	74	74	44	14	14	14	0
may-19	Cliente 5043	40.350.000,00	12.024.000	1	1	0	0	0	1	1	1	0	1	0	0
may-19	Cliente 5065	18.400.000,00	8.335.000	13	13	0	13	13	13	13	13	13	0	13	0
may-19	Cliente 5066	35.400.000,00	12.210.000	0	0	0	0	0	0	0	0	0	0	0	0
may-19	Cliente 1386	1.800.000,00	1.500.000	300	270	240	210	180	150	120	90	60	30	0	0
may-19	Cliente 5121	25.820.000,00	2.787.000	83	83	113	83	53	23	53	23	0	0	23	0
jun-19	Cliente 161	10.680.000	2.674.000	0	0	0	0	0	0	7	0	7	0	0	0
jun-19	Cliente 986	35.540.000	3.877.000	0	28	0	28	0	0	28	0	0	28	0	0
jun-19	Cliente 5179	62.330.000	10.950.000	0	0	0	23	0	23	23	0	23	23	0	0
jun-19	Cliente 1665	114.200.000	26.440.000	0	3	3	3	0	3	0	3	3	3	0	0
jun-19	Cliente 5203	19.177.177	6.007.000	0	0	16	16	0	0	0	0	16	16	0	0
jun-19	Cliente 5215	10.300.000	1.400.000	0	52	22	0	52	22	0	22	22	0	22	0
jun-19	Cliente 1203	15.641.000	1.839.000	0	41	11	0	11	11	11	11	11	0	0	0
jun-19	Cliente 788	22.830.000	6.674.000	0	11	0	0	11	0	11	0	11	0	0	0
jun-19	Cliente 5238	6.460.000	1.800.000	0	121	91	61	31	31	1	0	1	0	1	0
jun-19	Cliente 5247	16.100.000	3.917.000	0	60	30	28	30	30	30	0	30	0	0	0

Figura 4-1.: Aparte resumen base de datos entidad financiera

Con el fin de estimar las covariables observadas a partir de información histórica, se organizó la información de forma tal que las filas corresponden a los individuos y las columnas representan el

tiempo en meses, es decir el cambio en el tiempo o discretización es $\Delta t \approx 1,029$ si se considera el año (12 meses) como base. A partir de las series de tiempo de los activos y la deuda financiera se estimó mediante el modelo de Merton (ecuación B-2), la distancia a default para cada uno de los individuos. La información observable de cada prestatario corresponde a la distancia a default y el momento en que se presenta el default. Así, se tiene que $W = (W_1, \dots, W_m)$ es el vector de las covariables observadas de todos los individuos y $D = (D_1, \dots, D_m)$ el vector de procesos indicadores de default, recordando que $D_{it} = 1_{\{\tau_i \leq t\}}$ así D_{it} salta de cero 0 a uno 1 en el instante en que $\tau_i = t$.

Se considera un default si la altura de mora de la obligación sobrepasa 90 días calendario. Se presentaron 528 defaults (ver figura 4-2) para un total de 5250 individuos en el marco de tiempo analizado. Se puede apreciar además, que entre agosto de 2017 y agosto de 2018, se incrementó el número de defaults en el segmento analizado, este podría ser un ejemplo de *contagio default*. Además se puede observar como la desaceleración de la economía colombiana en ese periodo está relacionada de alguna manera con el comportamiento de pago de prestatarios en la cartera para esta entidad financiera.

Para los datos de activos, nivel de deuda y vector de mora se presentaron algunos faltantes. Se consideró utilizar imputación por la mediana, sin embargo, de acuerdo con la literatura, en estos casos, en los cuales se tienen datos de tipo mixto, variables numéricas y categóricas, se sugiere utilizar una técnica de imputación de datos conocida como *Random Forest*, esta técnica permite capturar relaciones no lineales e interacciones de mayor complejidad entre los datos, para conocer más detalles de esta técnica ver Stekhoven and Bühlmann (2011).

Las variables a partir de las cuales se ajustó el modelo para la intensidad de covariables, corresponden a información de indicadores macroeconómicos y específica de cada individuo, estos datos están dados en una resolución mensual entre agosto de 2016 y junio de 2019:

1. **Distancia a default:** Esta covariable fue construída a partir de los datos de activos y nivel de deuda de cada individuo, cabe resaltar que en algunos casos fue necesario imputar datos ya que la información de los individuos no fue cargada o actualizada en el sistema para ciertos periodos. A partir de estas variables se estimó la volatilidad y el rendimiento histórico asociado a la evolución de los activos, con estos datos se utilizó el modelo de Merton B-2, y se obtuvieron las distancias a default de cada individuo de la cartera para los 35 periodos analizados.
2. **Tasas de interés:** Al considerar que los individuos analizados hacen parte de un segmento crediticio propio de personas naturales, esta variable macroeconómica está directamente relacionada con el nivel de liquidez en la economía y las preferencias de liquidez de prestamistas e inversionistas, debido a que impacta directamente la propensión marginal de consumo y por tanto la evolución del nivel de deuda de los individuos, de otra parte esta variable

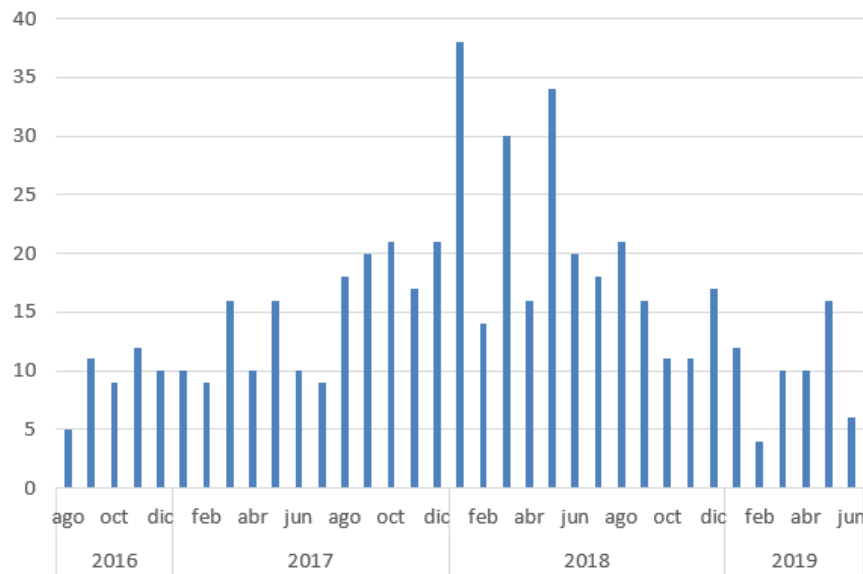


Figura 4-2.: Número de Defaults entre agosto de 2016 y junio de 2019

está inversamente relacionada con el (Índice de Precios al Consumidor) IPC, toda vez que el banco central (Banco de la República) toma decisiones de política monetaria e interviene estas tasas de acuerdo con los niveles de inflación observados, como una de sus principales funciones.

3. **Producto Interno Bruto (PIB):** Este indicador macroeconómico refleja el agregado de bienes y servicios de toda la economía, luego es un indicador que está relacionado por ejemplo con otras variables que pueden incidir sobre el comportamiento de pagos de una persona, como es el caso de los índices de desempleo, el crecimiento de la demanda (consumo) y el comercio.

4.2. Estimación de ML para el modelo con fragilidad

Con el fin de obtener convergencia numérica debido a la complejidad de la función de verosimilitud, un método de optimización tradicional puede presentar dificultades en cada paso de optimización del algoritmo (**M-Step**), debido a que éste debe calcular el gradiente a partir del vector de parámetros actualizado en cada paso con el fin de maximizar el valor esperado de la función de verosimilitud en (3-22). Se abordó la maximización de los parámetros mediante el método de enfriamiento o recocido simulado expuesto en el capítulo 2, obteniendo los resultados de la tabla 4-1.

Las consideraciones para la codificación y parametrización para el algoritmo fueron las siguientes:

- El valor inicial del parámetro $\Theta^{(0)}$ y el frenal de enfriamiento fue definido en 10.

- Para cambiar la temperatura se consideran dos parámetros: $K = \text{número de iteraciones}$, $A = \text{número de iteraciones en que se acepta un cambio}$.
- Factor de calentamiento.
- Si la razón $\frac{A}{K} < \text{Umbral mínimo de enfriamiento}$, se detiene el algoritmo.

El algoritmo buscó el nivel máximo para el valor esperado de la verosimilitud. A partir de 10.000 iteraciones para el Muestreador de Gibbs implementado, se *quemaron* o descartaron las primeras 1.000 trayectorias del factor de fragilidad, el óptimo global fue detectado en el punto límite de enfriamiento (valor de la temperatura aproximadamente cero).

El algoritmo inicio su búsqueda de máximos locales a partir de un vector de parámetros inicializado en $\Theta^{(0)} = (-2, 0,762, 0,964, -0,158, 0, 0,015, 0)$, que corresponden a los estimadores de máxima verosimilitud del modelo sin considerar el factor de fragilidad, por su parte se estableció el valor para los parámetros iniciales de la fragilidad en 0. Finalmente en el límite del enfriamiento se alcanzó un máximo global que corresponde a los parámetros de máxima verosimilitud para el modelo de intensidad, ver tabla 4-1.

Tabla 4-1.: Estimadores de máxima verosimilitud para la intensidad con el factor de fragilidad

Covariable	Coficiente
Constante $\hat{\alpha}$	-1.97
Tasas de interés \hat{r}	-0.2282
PIB \hat{g}	1.9542
Distancia Default $\hat{\beta}$	0.6615
Fragilidad $\hat{\eta}$	0.36
Frecuencia del Salto $\hat{\lambda}_{\xi}$	-0.0150
Varianza salto $\hat{\nu}$	0.35

Con el fin de determinar la calidad del ajuste de la función de intensidad sobre los datos, se superpuso el modelo ajustado para la intensidad de default, sobre la intensidad estimada en el periodo observado, la intensidad histórica u observada fue estimada como $\lambda_{hist}^{\hat{}} = \frac{defaults}{\delta_t}$ (número de eventos por unidad de tiempo).

En la figura 4-3, se presenta el ajuste correspondiente, en el eje de las abscisas se relaciona el tiempo, mientras que el eje vertical corresponde al número de eventos por unidad de tiempo, la línea azul discontinua presenta la intensidad del modelo ajustado, entre tanto, la línea negra corresponde a lo realmente observado entre agosto de 2016 y junio de 2019.

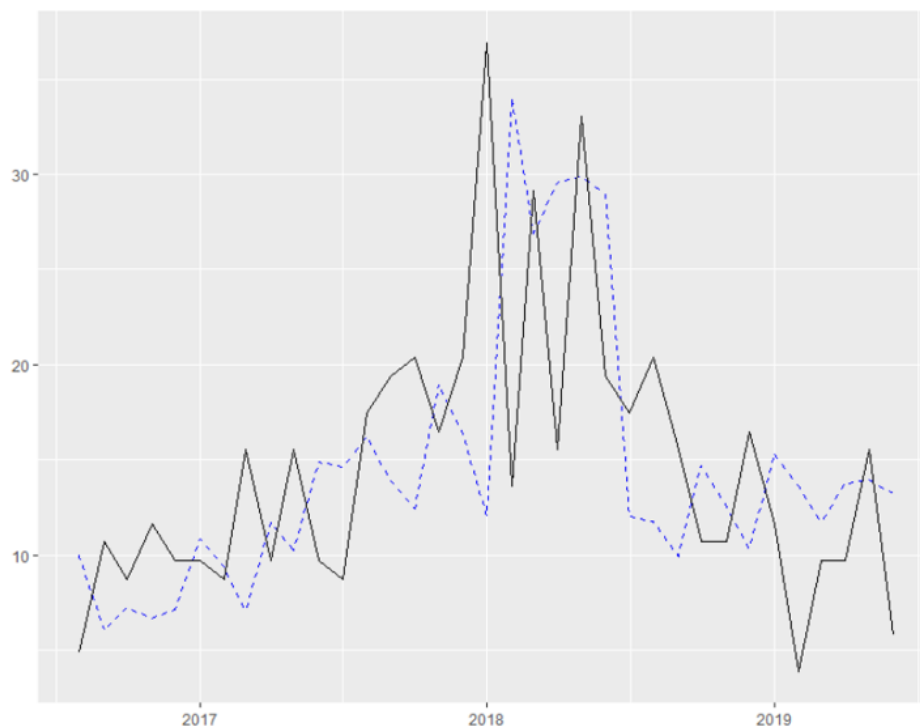


Figura 4-3.: Ajuste estimación tasa de intensidad con fragilidad. Elaboración propia.

A pesar de evidenciarse carencia de ajuste en algunos instantes, en general el modelo sigue los incrementos y disminuciones de la intensidad de los incumplimientos crediticios en el periodo observado, por ejemplo en 2018 se puede apreciar que se incrementaron súbitamente los defaults además de algunos clusters o agrupamientos, este resultado muestra evidencia a favor de incorporar el factor de fragilidad que se propone para la función de intensidad, además se observa un cambio repentino o salto en este periodo por lo cuál también se logra capturar parte de la dinámica de la función de intensidad incorporando el proceso de saltos sobre el factor de fragilidad.

A partir de estos resultados se evidenció que a pesar de no explicar completamente la intensidad de default, según los valores estimados para los diferentes parámetros de las covariables, se puede inferir que los de mayor incidencia sobre la intensidad corresponden a la distancia default y el proceso de fragilidad, por ejemplo el parámetro que acompaña la variable latente η corresponde a la volatilidad mensual del efecto de esta variable, este valor anualizado alcanza el 65.8 %

La covariable macroeconómica que presentó mayor incidencia sobre la intensidad de default para el periodo observado corresponde al PIB, confirmando la relación inversa vista en el periodo analizado, sin embargo, las tasas de interés también presentan un impacto significativo sobre la tasa con la que se presentan incumplimientos. En cuanto la información específica de los individuos, se estimó que un choque negativo de la distancia a default en una desviación estándar incrementa la intensidad de default en aproximadamente $e^{0,7215} - 1 \approx 106$ %.

De otra parte la magnitud del salto con la cual se incrementa o disminuye súbitamente la intensidad de default debido a la dinámica propia de los choques en agrupamientos de incumplimientos representa una volatilidad anual de aproximadamente $0,53\sqrt{12} \approx 184\%$.

4.3. Estimación de ML para el modelo sin fragilidad

En la tabla 4-2 se reportan los parámetros de máxima verosimilitud del modelo de intensidad sin considerar el factor de fragilidad.

Tabla 4-2.: Estimadores de máxima verosimilitud para la intensidad sin el factor de fragilidad

Covariable	Coefficiente
Constante $\hat{\alpha}$	2.8
Tasas de interés \hat{r}	-0.4882
PIB \hat{g}	0.9042
Distancia Default $\hat{\beta}$	-0.9985

Al dejar de lado la inclusión del factor de fragilidad en el modelo para la intensidad, la covariable que mayor incidencia genera en la intensidad de default es la distancia a default, es decir, tiene mayor incidencia la información específica de los individuos. Se estimó que un choque negativo de la distancia a default de una desviación estándar incrementa la intensidad de default en aproximadamente $e^{0,09985} - 1 \approx 170\%$, valor que dobla el impacto de esta covariable cuando se incorpora en el modelo la fragilidad.

De otro lado, se observa que se conservan los signos para los parámetros estimados con excepción de las tasas de interés que reducen su impacto sobre la intensidad de default pero siguen siendo una covariable significativa.

4.4. Modelo con fragilidad vs modelo sin fragilidad

A continuación se presentan los resultados del test de razón de verosimilitudes con el fin de dar cuenta del efecto de la fragilidad en la especificación del modelo para la intensidad. Como en Duffie et al. (2009), se considerará el factor de Bayes con el fin de comparar la calidad del ajuste de ambos modelos.

Asumiendo probabilidades a priori para ambos modelos, a saber: Modelo sin fragilidad ("noF") y Modelo con fragilidad ("F"), la razón de verosimilitudes o también conocida como el *ratio de*

Odds se expresa como:

$$\frac{\mathbb{P}(\mathbf{F}|W, D)}{\mathbb{P}(\mathbf{noF}|W, D)} = \frac{\mathcal{L}_F(\hat{\gamma}_F, \hat{\Theta}_F | W, Y, D)}{\mathcal{L}_{noF}(\hat{\gamma}_{noF}, \hat{\Theta}_{noF} | W, Y, D)} \frac{\mathbb{P}(F)}{\mathbb{P}(noF)}, \quad (4-1)$$

Donde $\hat{\Theta}$ y \mathcal{L} corresponden al vector de parámetros de máxima verosimilitud y la función de verosimilitud para cada modelo. Reemplazando 3-14 en 4-1 y teniendo en cuenta que el proceso VAR para las covariables observables es el mismo en los dos modelos se tiene que:

$$\begin{aligned} \frac{\mathbb{P}(\mathbf{F}|W, D)}{\mathbb{P}(\mathbf{noF}|W, D)} &= \frac{\mathcal{L}(\hat{\gamma}_F | W) \mathcal{L}_F(\hat{\Theta}_F | W, Y, D)}{\mathcal{L}(\hat{\gamma}_{noF} | W) \mathcal{L}_{noF}(\hat{\Theta}_{noF} | W, D)} \frac{\mathbb{P}(F)}{\mathbb{P}(noF)}, \\ &= \frac{\mathcal{L}_F(\hat{\Theta}_F | W, Y, D)}{\mathcal{L}_{noF}(\hat{\Theta}_{noF} | W, D)} \frac{\mathbb{P}(F)}{\mathbb{P}(noF)} \end{aligned} \quad (4-2)$$

Donde el término de la izquierda se conoce como el factor de Bayes, en línea con lo expuesto por Duffie et al. (2009), el criterio de rechazo para el test será de dos veces el logaritmo natural del factor de Bayes. Si el estadístico está entre 2 y 6 hay evidencia positiva a favor del modelo con fragilidad, un valor entre 6 y 10 habla de evidencia fuerte y un valor mayor a 10 es evidencia significativamente fuerte a favor del modelo con fragilidad. Para la especificación de ambos modelos se llegó a al siguiente resultado:

$$\begin{aligned} &= \frac{\mathcal{L}_F(\hat{\Theta}_F | W, Y, D)}{\mathcal{L}_{noF}(\hat{\Theta}_{noF} | W, D)} \\ &= 2 \log \left(\frac{740,83}{289,03} \right) = 2,229173 \end{aligned} \quad (4-3)$$

Como el estadístico se encuentra en 2 y 6 hay evidencia positiva a favor de incluir el factor de fragilidad en el modelo para la intensidad, este resultado corrobora lo obtenido a partir del ajuste en la figura 4-3, donde se puede apreciar que un incremento en la tasa de intensidad histórica es aproximadamente capturado por el proceso estimado a partir de los parámetros MLE estimados y obtenidos a partir del algoritmo **EM**.

5. Conclusiones y recomendaciones

5.1. Conclusiones

- Un fenómeno que motiva el estudio de la dinámica de la intensidad del default consiste en los clústeres o agrupamientos espontáneos en los cuales se presentan estos incumplimientos, es decir, los incrementos súbitos de la frecuencia o número de incumplimientos en cierto periodo de tiempo, a este fenómeno se le conoce como *fragilidad*.
- Se realizó una modificación al modelo propuesto en Duffie et al. (2009), donde el proceso de fragilidad asociado al proceso de Cox para la intensidad es un proceso **OU** donde se incluyó una componente de saltos con el fin de capturar el incremento súbito en el número de defaults y con esto mejorar la calidad en el ajuste sobre el proceso de intensidad.
- El proceso **Ornstein-Uhlenbeck** que incorpora un componente de saltos planteado, podría capturar el efecto acumulado sobre el tiempo de diferentes choques no observados fundamentales a la intensidad de default, además de los choques súbitos con los que pueden presentarse agrupamientos o clusters de estos incumplimientos.
- Mientras existan fallas en la gestión de la información y asimetría en el conocimiento del perfil del cliente por parte de una entidad financiera, a pesar que la componente de fragilidad pueda ser mitigada con mejoras en los procesos correspondientes, pueden originarse nuevas formas de presentarse un incremento común en las intensidades de default lo que justifica el hecho de seguir considerando una fuente de ruido común no observable.
- Se estimaron los parámetros para un modelo estructural de riesgo crediticio, en el cual se ajustó la intensidad de default, a partir de un modelo con factores de riesgo observables y un factor de riesgo latente (no observable), común al conjunto de préstamos para un segmento de cartera masiva de una entidad financiera colombiana.
- Muchos problemas de inferencia importantes en estadística, como los modelos de parámetros aleatorios y variables latentes, resultan ser solucionables por EM. Aquí, se trató principalmente con el problema de estimar los parámetros para la intensidad de default en presencia de un factor latente.
- Debido a la dificultad en el procesamiento del algoritmo **EM** se realizó una variante al proceso de estimación de cada paso de maximización, a partir de un algoritmo conocido como

enfriamiento simulado en lugar de utilizar Newton Raphson como se sugiere en la literatura buscando mejorar la eficiencia en el procesamiento. El método modela el proceso físico de calentamiento de una partícula y luego baja lentamente la temperatura hasta alcanzar una aproximación del punto de máximo global.

- Se aplicó la regla de Bayes con el fin de actualizar la distribución a posteriori del factor de riesgo latente y capturar la dinámica con la que se presentan incumplimientos crediticios en el tiempo, que tienden a presentarse en clusters, cuyo hecho no se espera basándose únicamente en los factores observados.
- Se caracterizaron las covariables idiosincráticas con incidencia en el hábito de pagos para los individuos de una línea de crédito en una entidad financiera: El vector de altura de mora a 90 días, el nivel de endeudamiento reportado en la CIFIN y el nivel de activos, finalmente se estimó la distancia a default para los individuos de la cartera a partir de una adaptación del modelo de Merton.
- Se caracterizaron las covariables macroeconómicas que podrían explicar el comportamiento de pagos en el segmento de banca personas para una entidad financiera colombiana.
- Según los resultados de los parámetros estimados se identificó que el crecimiento de la economía es la covariable macroeconómica con mayor incidencia en los cambios de la intensidad de incumplimientos.
- A partir del test de razón de verosimilitudes y el factor de Bayes, se obtuvo evidencia positiva a favor de la inclusión del factor de fragilidad en el modelo para la intensidad de default.

5.2. Recomendaciones

- El mayor reto en el desarrollo de este trabajo se presentó en el procesamiento del modelo, debido al problema de optimización sobre los valores esperados de una función de verosimilitud con cierta complejidad para la integración numérica. Se proponen varias alternativas que podrían contribuir a mejoras en la estimación de los parámetros para la intensidad de default en trabajos futuros:
 - Explorar otras metodologías para establecer el vector inicial de parámetros.
 - Explorar otros algoritmos o métodos de optimización para mejorar la eficiencia de cómputo en el procesamiento del modelo.
 - Incorporar un proceso para la volatilidad de la fragilidad y ajustar el modelo con otras variantes para el proceso de difusión.
- Se recomienda incorporar criterios de información que den cuenta de la calidad del ajuste. El paquete con el cual se aplicó el algoritmo de enfriamiento simulado *OptimSa*, no cuenta con una rutina integrada que permita generar regiones o intervalos de confianza.

- La cantidad de datos también puede influir en la calidad del ajuste y la estimación, si bien los tres años de información con la cual se contó para realizar el ajuste facilitan el procesamiento, es de interés comprobar la replicabilidad del modelo con más historia que aporte un mayor volumen de información sobre los individuos de la cartera.
- Para el proceso VAR ajustado al componente de la verosimilitud observable se podrían considerar modelos más robustos para ajustar las series de distancia a default y el logaritmo de los activos de los individuos.

A. Anexo: Ajuste VAR(1) covariables

A.0.1. Resultados del ajuste VAR(1)

Respecto a los resultados de la estimación de los parámetros de máxima verosimilitud para el proceso de dos factores, con la información de los indicadores macroeconómicos en el periodo observado se tiene lo siguiente (error estándar en subíndices):

$$\beta_r = \begin{pmatrix} 0,7618_{(0,1127)} & -0,0898_{(0,05)} \\ 0,0514_{(0,0897)} & 0,9642_{(0,06)} \end{pmatrix},$$
$$\theta_r = \begin{pmatrix} 0,0104_{(0,0054)} \\ -0,0003_{(0,0043)} \end{pmatrix}$$

Mientras que la matriz de Cholesky queda dada por:

$$C_r = \begin{pmatrix} 1,658595e - 05 & -1,919835e - 06 \\ -1,919835e - 06 & 1,049275e - 05 \end{pmatrix}$$

En cuanto a los estimadores de máxima verosimilitud para los parámetros con información idiosincrática de todos los individuos del segmento de crédito, se obtuvieron las siguientes estimaciones (con errores estándar entre paréntesis):

$$b = (-0,158_{(0,1795)} - 0,4731_{(0,1531)})',$$
$$\kappa_\delta = 2,3238_{(0,3603)},$$
$$\sigma_\delta = -0,1685_{(0,4834)},$$
$$\kappa_U = 1,473_{(1,1019)},$$
$$\sigma_U = 0,033_{(0,0583)}$$

$$AA' + BB' = \begin{pmatrix} 1 & 0,2381455 \\ 0,2381455 & 1 \end{pmatrix},$$

Mientras que:

$$W_r = \begin{pmatrix} 0,007767747 & 0,005656807 \\ 0,005656807 & 0,072637836 \end{pmatrix}$$

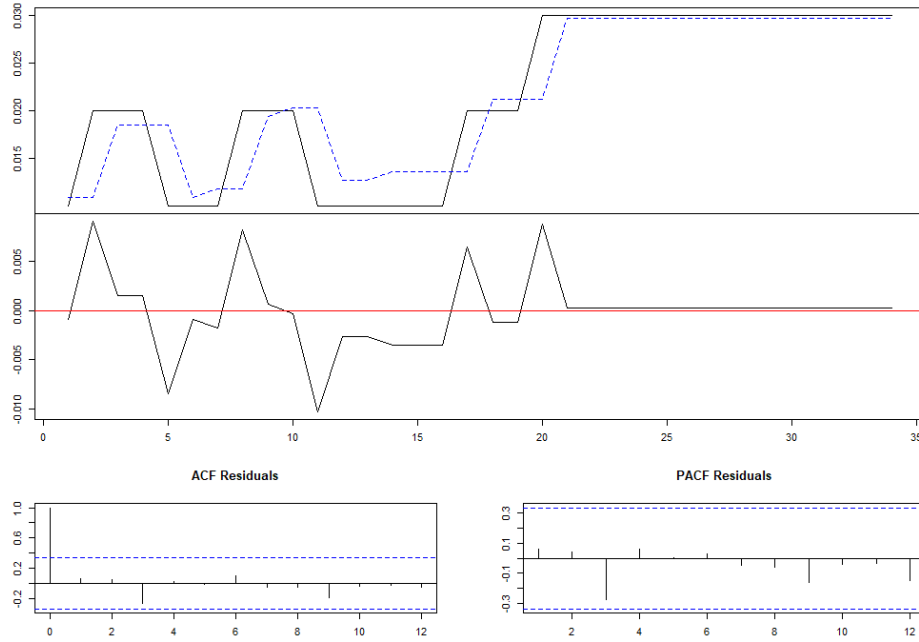


Figura A-1.: Ajuste VAR(1) para el IPC

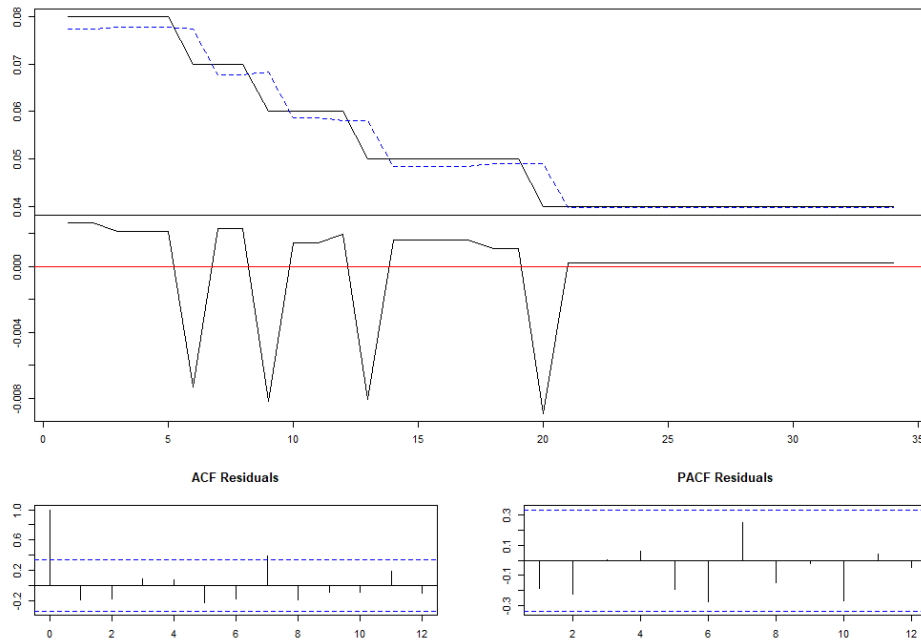


Figura A-2.: Ajuste VAR(1) para el PIB

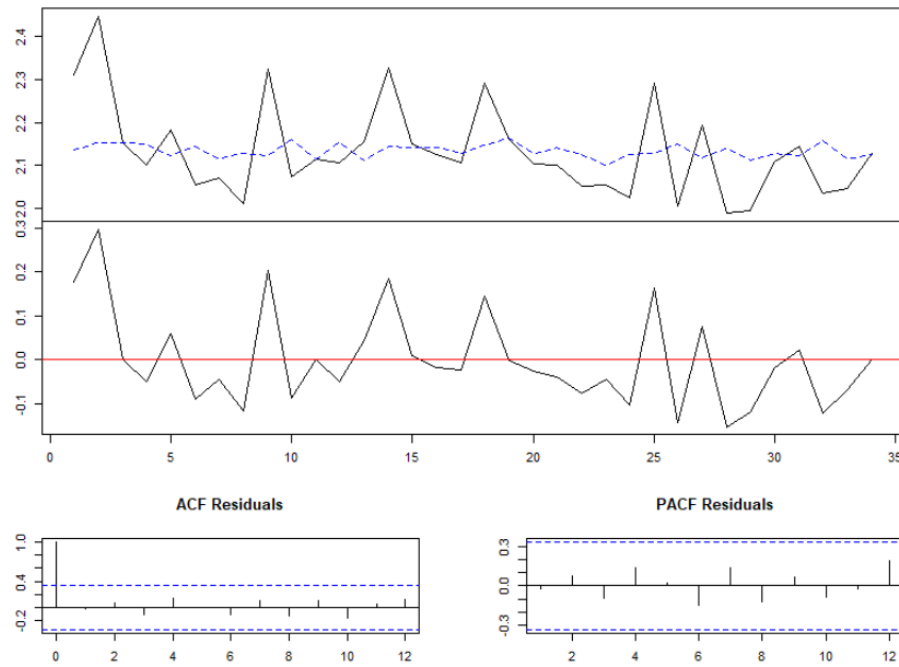


Figura A-3.: Ajuste VAR(1) para la Distancia Default

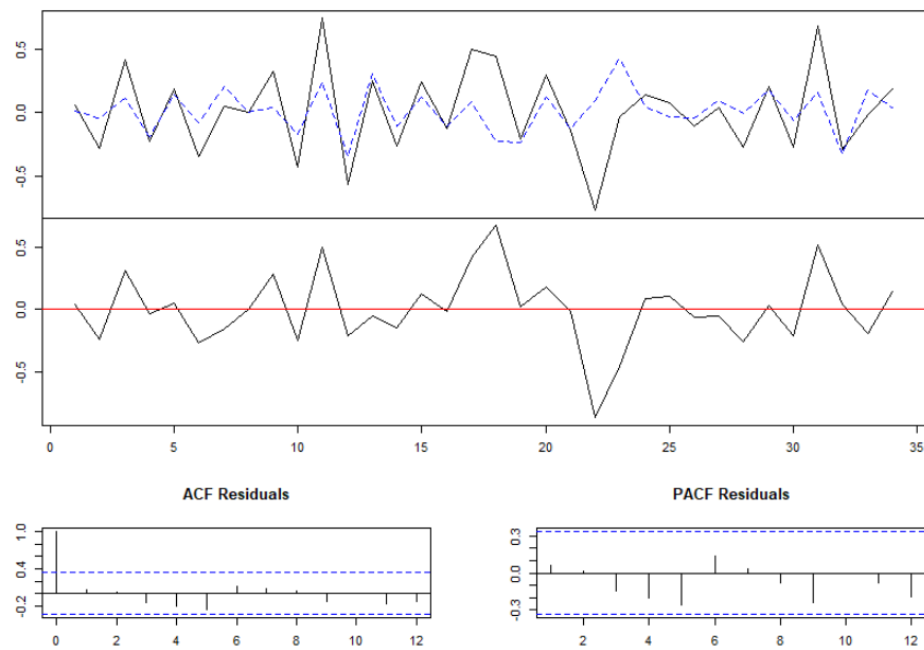


Figura A-4.: Ajuste VAR(1) para el logaritmo de los activos

B. Anexo: Breve descripción metodologías de riesgo de crédito

B.1. Modelo de Merton

Instituciones financieras, bancos de pensiones y compañías de seguros en su sistema de administración de riesgos necesitan evaluar el riesgo de sus carteras activas o pasivas de deuda, por esta razón se utilizan modelos estructurales de riesgo de crédito, este modelo establece una estructura para variables económicas e indicadores financieros para a partir de éstos intentar explicar la probabilidad de default principalmente en carteras de deuda o bonos dónde el prestamista corresponde a una compañía listada en bolsa.

El modelo de Merton es un modelo estructural basado en el modelo de valuación de opciones Black-Scholes. Este modelo puede ser usado tanto para valorar acciones como para valorar el riesgo de crédito. Intuitivamente, el modelo se puede describir a partir de la propiedad residual que tienen los accionistas propietarios en una firma, es decir, sobre el valor de los activos por encima del valor de mercado de la deuda que se asimila al precio de ejercicio, por tanto se podría afirmar que un accionista es propietario de una opción call (de compra) con precio de ejercicio L (valor de la deuda). Si el valor de los activos V_T en un instante T es mayor que L , el pago en T de esta opción es $V_T - L$. Así mismo puede considerarse el tenedor de un bono que tiene una posición corta sobre una opción de venta.

El modelo de Merton es una adaptación del modelo BS para estimar el valor de una compañía en función de sus activos y el valor de mercado de la deuda. Los parámetros implícitos en la respectiva expresión corresponden a: la tasa libre de riesgo r , la volatilidad de los activos σ_v , el tiempo al vencimiento T , por su parte $\Phi(\cdot)$ corresponde al cuantil de la función de distribución acumulada de una normal estándar. Claramente entre más alto sea el valor de los activos en relación con la deuda mayor es el valor de la firma, Hull (2006).

$$W_t = V_t \Phi(d_1) - L_t e^{-rT} \Phi(d_2) \quad (\text{B-1})$$

$$d_1 = \frac{\ln\left(\frac{V_t}{L_t}\right) + \left(r + \frac{1}{2}\sigma_A^2\right)}{\sigma_A \sqrt{T}} \quad (\text{B-2})$$

$$d_2 = d_1 - \sigma_v \sqrt{T} \quad (\text{B-3})$$

Donde V_t corresponde al nivel de activos en un tiempo t , L_t corresponde al valor de las obligaciones o deuda en un tiempo t y W_t corresponde al valor residual o patrimonial en un tiempo t .

Una dificultad con este modelo consiste en que tanto el valor de los activos y la volatilidad de los mismos no son observables. Sólo para las compañías listadas en bolsa se tiene que el precio de su acción es una aproximación del valor de mercado del patrimonio. A partir del modelo de Merton se puede derivar la probabilidad de default, por medio de las probabilidades neutrales al riesgo las cuales expresan la posibilidad de que al vencimiento, el valor de los activos sea menor o mayor que el valor de mercado de la deuda. Por ejemplo en el caso de una opción, la probabilidad de que esta expire al vencimiento sin valor según el modelo Black-Scholes vendría dada por $\Phi(-d_2)$ mientras que la probabilidad de que esta sea ejercida es $\Phi(d_2)$, Hull (2006).

B.2. Modelo KMV

El modelo KMV es otro modelo estructural a partir del cual se busca estimar la frecuencia esperada de default de una firma, al igual que el modelo de Merton es preferiblemente usado para compañías que listan en bolsa donde el valor de mercado de la compañía está implícito en el valor de la acción. La información de mercado de la acción y la información pública de los estados financieros de la compañía son usadas para estimar los riesgos de incumplimiento crediticio, de hecho este modelo y el de Merton presentan ciertas similitudes ya que ambos se derivan del modelo Black-Scholes. En el modelo se plantean seis variables que determinan la probabilidad de default entre un momento inicial y un horizonte de tiempo H .

- El valor actual de los activos.
- La distribución del valor de los activos en un tiempo H
- La volatilidad de los activos al tiempo H .
- El nivel de default dado por el valor en libros de los pasivos.
- El tasa esperada de crecimiento de los activos durante el horizonte.
- El tiempo al vencimiento definido para el horizonte dado H .

Consideremos ahora la figura **B-1**, en el eje x se tiene el tiempo y en el eje y tenemos una posible trayectoria para el valor de mercado de los activos de una compañía, este valor está dado por (1), ahora considere el punto de default (4) definido por el valor en libros de la deuda, en la figura, el numeral (2) corresponde a la distribución futura del valor de los activos en el horizonte H .

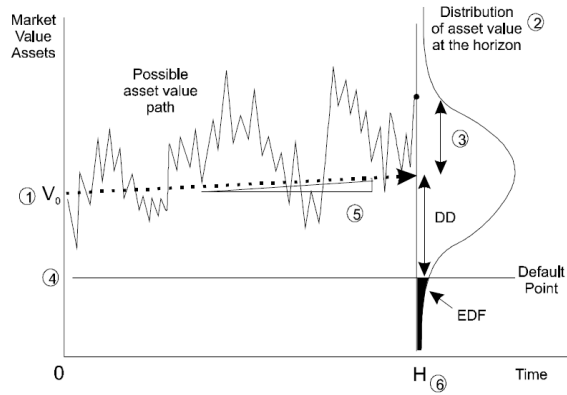


Figura B-1.: Variables que determinan la probabilidad de default. Tomado de Moodys KMV, Pag 13.

Inicialmente las 4 variables con mayor impacto son el valor actual de los activos, la deuda, la volatilidad de los activos y la distribución futura de los rendimientos de los activos de la firma sobre la cuál se quiere estimar la distancia a default. Si el valor de los activos cae por debajo de la deuda entonces la firma entra en default, luego la probabilidad de default es estimada a partir del punto donde inicial el área sombreada (EDF) en la figura **B-1**. La distancia a default DD en el horizonte H está definida como el número de desviaciones estándar a las que se encuentran los activos del punto de default (nivel de deuda) ya que hay evidencia empírica de una alta correlación en el comportamiento de estas dos variables.

B.3. Modelos de intensidad

La principal variable de estudio es un tiempo aleatorio τ que representa el tiempo de *default* de una compañía en relación a un compromiso de crédito, de tal forma que la probabilidad de default antes del instante t es $P(\tau \leq t)$, se caracteriza a partir de la función de distribución de probabilidad de τ , en un escenario de varias compañías se denota por τ_i al tiempo a default de la i -ésima compañía. Si T es el tiempo de vencimiento de una obligación y se tiene que $\tau \geq T$ entonces no hubo incumplimiento por parte del prestatario y se asume que ha cumplido en su totalidad con el plan de pagos acordado.

En los modelos de intensidad se define una función λ , para la cual, si τ_i denota el tiempo a default de la i -ésima compañía, entonces la probabilidad de default antes de t queda definida por, McNeil et al. (2005):

$$F_{\tau_i}(t) = P(\tau_i \leq t) = 1 - e^{-\int_0^t \lambda_{is} ds} \quad (\text{B-4})$$

Dónde λ_t se define a partir de:

$$\lambda_t = \lim_{h \rightarrow 0} \frac{\mathbb{P}(\tau \leq t + h | \tau \geq t)}{h} \quad (\text{B-5})$$

B.4. Modelo Credit Risk Metrics

Este modelo fue implementado por el Banco Credit Suisse en 1997 para estimar la provisión necesaria para cubrir las posibles pérdidas por el riesgo en bonos y créditos. En su concepción es un modelo de mezclas de Bernoulli con probabilidades que dependen de factores de riesgo sistemático, este es ampliamente implementado en el sector financiero.

Se puede caracterizar a partir de probabilidades de default que dependen de factores de riesgo distribuidos Gamma para establecer finalmente una relación lineal entre los factores de riesgo que pueden ser de mercado e incorporar información idiosincrática o referente a los individuos de la cartera. Las variables que incluye el modelo son las siguientes, McNeil et al. (2005):

- Se asume una línea o segmento homogéneo de N créditos. Por ejemplo el segmento de microcréditos en una entidad financiera.
- Se asume un vector $X_t = (Y_1, \dots, Y_m)$ de m factores de riesgo sistemático dados por variables del mercado e inclusive se puede considerar un factor de riesgo idiosincrático con información individual.
- Estas variables poseen distribución Gamma, $Y_j \sim \text{Gamma}(\alpha_j, \beta_j)$ con parámetros $\alpha_j = \frac{1}{\sigma_j^2}$ y $\beta_j = \sigma_j^2$, y son independientes entre sí.
- La tasa o intensidad asociada a los defaults de la cartera en este caso depende de X_t y se establece como $\lambda_i(\mathbf{X}) = p_i w_i' \mathbf{X}$, p_i equivale a una probabilidad de default apriori, aquí w_i es un vector de pesos.
- Finalmente se define una variable D_i , condicionalmente independiente, tales que: $D_i | \mathbf{X} \sim \text{Bin}(1, \lambda_i(\mathbf{X}))$

B.5. Modelos estructurales

Denominados modelos de barrera se caracterizan en explicar la probabilidad de incumplimiento a partir de un modelo no lineal para un vector de estados X_t que esta definido en función de información idiosincrática en presencia de una variable latente. Estos modelos se basan principalmente en procesos de difusión, McNeil et al. (2005).

B.5.1. Modelo de Cox Ingersoll Ross

Este modelo está definido a partir de la siguiente ecuación diferencial estocástica, Dobrow (2016):

$$dX_t = -\kappa(X_t - \mu)dt + \sigma\sqrt{X_t}dB_t \quad (\text{B-6})$$

Este modelo es frecuentemente utilizado para describir la dinámica en la evolución de las tasas de interés (modelos de tasa corta). Se puede apreciar que la difusión conserva el mismo coeficiente de tendencia así como el patrón de reversión a la media del proceso Orstein-Uhlenbeck, sin embargo al considerar una raíz en el término estocástico el proceso deja de tener una distribución normal, de hecho en este caso la distribución corresponde a una χ^2 no central, Martínez (2006).

B.5.2. Modelo de tres factores

Otro modelo de tasa corta, se asume que \mathbf{X}_t es un vector de tres procesos de difusión independientes $\mathbf{X}_t = (X_{t,1}, X_{t,2}, X_{t,3})'$, cada uno con la siguiente dinámica, McNeil et al. (2005):

$$dX_{t,i} = -\kappa_i(X_{t,i} - \mu_{t,i})dt + \sigma\sqrt{X_{t,i}}dB_{t,i} \quad (\text{B-7})$$

Para movimientos Brownianos $B_{t,i}$, $i = 1, 2, 3$ independientes. La tasa corta de interés está dada por: $r_t = r_0 + X_{t,2} + X_{t,1}$ para una constante $r_0 \geq 0$. La tasa de intensidad de la contraparte bajo consideración está dada por $\gamma_t = \gamma_1 X_{t,1} + X_{t,3}$ para $\gamma_1 > 0$. Esta parametrización permite una autocorrelación negativa entre r_t y γ_t .

B.5.3. Duffie and Garleanu (2001)

Este modelo de difusión por saltos es utilizado con frecuencia en la valoración de obligaciones de deuda colaterales, específicamente la dinámica de \mathbf{X}_t viene dada por:

$$dX_{t,i} = -\kappa(\theta - \mathbf{X}_t)dt + \sigma\sqrt{X_t}dB_t + dZ_t \quad (\text{B-8})$$

Donde los parámetros $\kappa, \theta, \sigma > 0$ y el proceso de saltos \mathbf{Z}_t tiene una constante que refleja la magnitud de la intensidad del salto $l^0 > 0$, por tanto los saltos son de igual magnitud y el tiempo entre estos se distribuye exponencial con parámetro $1/\mu$. También conocido como modelo básico estructural de difusión con saltos, donde cada salto siempre es de magnitud positiva, ver **B-2**, McNeil et al. (2005).

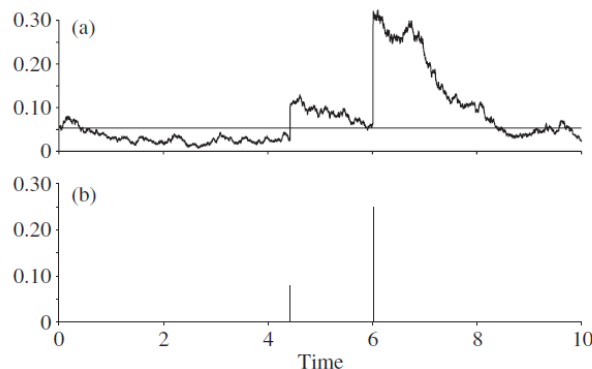


Figura B-2.: Trayectoria del modelo estructural de difusión de Duffie y Garleanu (2001)

En la figura se presenta una trayectoria del modelo de difusión estructural propuesto por Duffie y Garleanu (2001), con parámetros

B.5.4. Cópulas

En teoría de probabilidad una cópula consiste en una función de distribución multivariada cuyas distribuciones marginales provienen de una distribución uniforme. Son especialmente utilizadas para describir la estructura de dependencia entre variables aleatorias. A partir del teorema de Sklar se deriva que función de distribución conjunta puede expresarse en términos de distribuciones marginales, ver McNeil et al. (2005).

Los modelos de cópulas son ampliamente utilizados es la práctica en la valoración de obligaciones de deuda colateralizadas por su facilidad en la implementación, además son usados para modelar el contagio default, fenómeno que inicia con pequeños choques o incumplimientos que inicialmente afectan a unas cuantas entidades financieras pero que gradualmente se empiezan a extender al resto del sector e inclusive hacia otros países. De forma similar a como ocurre con un virus, las cópulas se utilizan especialmente para capturar la estructura de correlación presente en el comportamiento de defaults colectivos.

Estos modelos de factores de cópula asumen una tasa de intensidad $\gamma_i(t)$ y consideran una variable de mezclas de Bernoulli con factor U y probabilidades condicionales de default dadas por $Q_i(\mathbf{v}) = 1 - F_{\tau_i|\mathbf{V}}(t|\mathbf{v})$, en la literatura de análisis de supervivencia el vector \mathbf{V} es una variable latente y es conocida como factor de *fragilidad* del default. En la literatura se pueden encontrar modelos dinámicos de riesgo crediticio de factores que utilizan diferentes tipos de cópulas, a saber: La cópula de Gauss C_P^{Ga} , las cópulas arquimedianas, entre otras, para mayor detalle ver McNeil et al. (2005).

C. Anexo: Programa implementado para el algoritmo EM

```
1 #####CARGUE LIBRERIAS#####
2
3 library(zoo) #####Permite excluir NA's inmersos en los datos
4 library(progress) #####Permite utilizar una barra de progreso
   de las tareas en ejecucion
5 library(optimization) #####Permite implementar la funcion de
   enfriamiento simulado optim_sa
6
7
8 #####DATOS INDIVIDUOS#####
9
10
11 #####Dimension de datos#####
12
13 m = 5250  ###Numero de individuos analizados
14
15 n = 35  ###Numero de periodos
16
17
18 ti <- seq(0,35,length.out = n) #####Resolucion temporal
   mensual#####
19
20 Dt <- ti[2]-ti[1]  #####Calculo Delta t o unidad de cambio en
   el tiempo
21
22 dtc= Dt  #####Se renombra con el fin de no confundir con variable
   Defaults
23
24
25 #####Cargue datos Defaults#####
26
27
```

```
28 D <- read.table(file = "C://Users//Luis Bernal//Desktop//MsC
    Estadistica//Tesis//18082019//Filtrados//Default.csv",header
    = FALSE,dec=".",sep=";",stringsAsFactors=TRUE)
29
30 D <- as.matrix(D[1:m,],nrow = m, byrow = TRUE) ###Se asigna
    como arreglo matricial
31
32 #####La siguiente linea permite darle un manejo matricial a los
    datos de default
33
34 De = matrix(rep(0,n*m),nrow = m,byrow = T)
35
36 for (i in seq(1,m)) {
37
38 De[i,] = as.vector(D[i,])
39
40 }
41
42
43 #####Cargue distancias a default
44
45
46 W <-read.table(file = "C://Users//Luis Bernal//Desktop//MsC
    Estadistica//Tesis//18082019//Filtrados//Distancias.csv",
    header = FALSE,dec=" ",sep=";",stringsAsFactors=TRUE)
47
48
49 W <- as.matrix(W[1:m,],nrow = m, byrow = TRUE) ###Se asigna
    como arreglo matricial
50
51 #####Cargue tasas de interes
52
53 rt <-read.table(file = "C://Users//Luis Bernal//Desktop//MsC
    Estadistica//Tesis//18082019//Filtrados//tasas.csv",header =
    FALSE,dec=" ",sep=";",stringsAsFactors=TRUE)
54
55
56 rt <- as.matrix(rt[1:m,],nrow = m, byrow = TRUE) ###Se asigna
    como arreglo matricial
57
58
59 #####Cargue tasas evolucion del PIB
```

```

60
61 g <-read.table(file = "C://Users//Luis Bernal//Desktop//MsC
    Estadistica//Tesis//18082019//Filtrados//PIB.csv",header =
    FALSE,dec="," , sep=";", stringsAsFactors=TRUE)
62
63
64 g <- as.matrix(g[1:m,],nrow = m, byrow = TRUE) ###Se asigna
    como arreglo matricial
65
66
67 #####Definicion inicial de parametros para el vector asociado
    a la funcion de intensidad#####
68
69 al <- -2          ###Intercepto (beta_0)
70 tasas<- 0.7618   ###Estimador ML modelo sin fragilidad para
    las tasas de interes (beta_1)
71 PIB<- 0.9642     ###Estimador ML modelo sin fragilidad para el
    crecimiento del PIB (beta_2)
72 beta <- -0.1585  ###Estimador ML modelo sin fragilidad para
    las distancias a default de c/individuo (beta_3)
73 eta<- 0          ### Parametro inicial para el proceso de
    fragilidad
74 Zi <- 0.015      ### Parametro inicial para la magnitud del
    proceso de saltos
75 nu <-0           ### Parametro inicial para la varianza de la
    magnitud del salto (gaussiana)
76
77
78 #####Valores iniciales vector de parametros funcion de
    intensidad#####
79
80 theta_in <- c(al,tasas,PIB,beta,eta,Zi,nu)
81
82
83 #####Valor inicial trayectoria fragilidad (no observada)
    #####
84
85 Y <- rep(0,n)
86
87
88 #####Algoritmo EM (Expectation Maximization)
    #####

```

```

89 #####Estimacion conjunta del vector de parametros
    theta
90 #####que depende de covariables observables en
    presencia de una variable latente#####
91
92
93
94 ##### ----- PRIMERA PARTE: VEROSIMILITUD
    ----- #####
95
96
97 ##### El siguiente programa calcula la funcion de
    verosimilitud presentada para el proceso asociado a la
    intensidad#####
98
99 Like <- function(theta_in,Y){
100
101 #####Variable auxiliar para la fragilidad
102
103 Y1 <- matrix(rep(Y,m),nrow = m, byrow = T)
104
105 ##### Funcion de intensidad (log-lineal en los parametros del
    proceso AR
106 #####y el proceso de fragilidad)
107
108 LAMBD_A <- exp(theta_in[1]+rt*theta_in[2]+g*theta_in[3]+W*theta
    _in[4]+Y1*theta_in[5])
109
110
111 #####En el siguiente ciclo se captura el comportamiento
    asociado al proceso de defaults en funcion del proceso de
    intensidad para los m individuos
112
113 Like1 = 1
114
115 for (i in seq(1,m)) {
116
117 if (length(which(D[i,] != 0))) == 0){
118
119 Like1 = Like1*exp(-sum(LAMBD_A[i,])*Dt)
120
121 }

```

```

122
123
124     else {
125
126     Ntau = which(D[i,] > 0 )[1]
127
128
129     Likel = Likel*exp(-sum(LAMBDA_A[i,1:Ntau])*Dt)*LAMBDA_A[i,Ntau]*
        Dt
130
131     }
132
133
134     }
135
136     return(Likel)
137
138     }
139
140     ## Para Chequear la funcion anterior , ejecute la siguiente
        linea..
141
142     Like2 <- (log(Like(theta_in,Y)))
143
144     (Like2)
145
146
147
148     #####-----SEGUNDA PARTE: PASO E(EXPECTATION)
        -----#####
149
150     ##A partir del siguiente programa se obtiene el valor esperado
        de la funcion de verosimilitud que esta dada en funcion de
        los parametros iniciales#####
151
152     Expect_Verol <- function(theta,theta1){
153
154     #####Con el fin de muestrear las trayectorias del proceso
        de fragilidad se implementa a continuacion el muestreador de
        GIBBS, para ello se hace necesario definir el numero de
        trayectorias, asi como un punto donde el proceso alcance
        estacionariedad y se descartan las primeras trayectorias

```

```
155
156
157     Burn = 1000     #####Numero de trayectorias a descartar o "
                        quemar"
158
159     M = 10000      ##### Numero de trayectorias a muestrear para
                        la fragilidad
160
161
162     ##### Debido a que este programa es exigente en el
                        procesamiento es util implementar la siguiente barra de
                        progreso la cual muestra el valor de la funcion del paso E
                        asociado a cada trayectoria del proceso
163
164     ## ----- Expectation Bar progress ----- ##
165
166     pb <- progress_bar$new(
167     format = "(:spin) [:bar] :percent eta: :eta",
168     total = M, clear = FALSE, width = 60)
169
170     ## ----- ##
171
172
173     Esp_Like <- 0 ##### Valor de conteo para el valor esperado
174
175     s <- 1          ##### Valor inicial del contador
176
177
178
179     #####Muestreador de GIBBS#####
180
181     #####Para cada tiempo deltaT se genera un valor para
                        trayectoria de fragilidad
182
183
184     while(s <= M) {
185
186     for (j in seq(1,length(Y))) {
187
188     Y1_aux <- Y     #####Variable auxiliar
189
```



```

190 Prob_jump <- theta_in[6]*Dt #####Probabilidad del salto (
      depende de la magnitud Zi)
191
192 #####A continuacion se define una variable indicadora (
      bernoulli) para los instantes de salto con probabilidad (Zi*
      deltaT)
193
194 Ind_Jump <- sample(c(0,1), size = 1, replace = TRUE, prob = c
      (1-Prob_jump,Prob_jump))
195
196 ##### Se modifica aditivamente la media del proceso OU de
      fragilidad con un salto en los instantes definidos por la
      variable indicadora
197
198 Y_star <- Y1_aux[j] + Ind_Jump*rnorm(1,0,theta_in[7])
199
200
201 ##### En los casos que la variable indicadora toma el valor de
      0 (no hay salto) la fragilidad se muestrea a partir de un
      proceso gaussiano con media igual al valor anterior de la
      cadena y varianza 4
202
203 Y2_aux <- rnorm(1,Y_star,2)
204
205
206 ## -----Algoritmo Metropolis-Hasting----- ##
207
208 #####El objetivo de esta rutina es implementar M-H para
      muestrear finalmente las trayectorias del proceso OU con
      saltos definidas para la fragilidad
209
210
211 Y3_aux <- Y1_aux #####Variable auxiliar para definir el valor
      anterior del proceso
212
213 Y3_aux[j] <- Y2_aux #####Variable auxiliar para definir la
      trayectoria muestreada a partir del proceso de saltos
214
215
216 ##### Funcion de aceptacion: Este corresponde al criterio para
      definir si el valor del proceso se establece a partir de la
      variable muestreada o al valor anterior de la cadena

```

```
217
218     F_acepta <- min(1,Like(theta1,Y3_aux) / Like(theta1,Y1_aux))
219
220     Un <- runif(1)
221
222     if ( F_acepta >= Un){
223
224     Y[j] <- Y2_aux}
225
226
227     }
228
229     ##### A continuacion se tienen en cuenta para el calculo del
        valor esperado unicamente las trayectorias mas alla del
        punto de quemado, el valor que se acumula corresponde a la
        log-verosimilitud
230
231     if (s > Burn){
232
233     Esp_Like = Esp_Like + (log(Like(theta,Y1_aux)))
234
235     }
236
237     #####Esta barra de progreso se utiliza para medir el tiempo
        del calculo del valor esperado
238
239
240     ## ----- Expectation Bar Progress (2) ----- ##
241
242     pb$tick()
243     Sys.sleep(3/100)
244
245     ## ----- ##
246
247     s <- s+1
248
249     print(c(- Esp_Like/s, paste("iter",s)))
250
251     }
252
253     #####Por ley de grandes numeros el valor esperado converge al
        promedio de la funcion de log-verosimilitud evaluada en
```

```
    cada trayectoria del proceso no descartada
254
255
256 return(Esp_Like/(M-Burn))
257
258
259 }
260
261
262 ## Para Chequear la funcion anterior , ejecute la siguiente
    linea..
263
264
265 (Expect_Verol(theta_in,theta_in))
266
267
268
269
270 ## ----- TERCERA PARTE: MAXIMIZATION
    ----- ##
271
272
273 theta_k = theta_in #####Variable auxiliar para cada paso E
    del algoritmo que se pretende maximizar
274
275
276 #####Funcion auxiliar para almacenar el valor esperado (
    funcion a maximizar)#####
277
278
279 Vero_optim <- function(theta){
280
281     theta1 = theta_k
282
283     return(-Expect_Verol(theta,theta1))
284
285 }
286
287
288 #####Algoritmo de Enfriamiento Simulado
    #####
```

```
289 #####Objetivo: Maximizar cada paso E. Se hace
      necesario implementar el algoritmo de enfriamiento simulado
      disponible en la funcion optim_sa de la libreria
      optimization
290
291
292 v <- theta_in #####Valor inicial para el vector de parametros
      theta
293
294
295 #####Para acotar la busqueda de maximos del
      algoritmo se hace necesario definir el siguiente intervalo
      para los valores del vector de parametros theta
296
297 lower <-c(-3,-2,-2,-2,-2,-2,-2) #####Valores minimos
298
299 upper <- c(0,2,2,2,2,2,2) ##### Valores maximos
300
301
302
303 ##### Para encontrar el maximo es necesario
      parametrizar en la funcion optim_sa de base: 1. Una funcion
      a maximizar, un vector inicial de parametros, incluir el
      intervalo de busqueda y ademas un numero minimo para la
      temperatura
304
305 out <- optim_sa(Vero_optim, v, maximization = TRUE, trace =
      TRUE,
306 lower, upper, control = list(nlimit=1000,t_min=10))
307 out
308
309 ## Para Chequear la funcion anterior, y visualizar los
      estimadores de maxima verosimilitud del algoritmo EM
      implementado ejecute la siguiente linea
310
311 (out$par)
```

Bibliografía

- Ahn, J. J., Oh, K. J., Kim, T. Y., and Kim, D. H. (2011). Usefulness of support vector machine to develop an early warning system for financial crisis. *Expert Systems with Applications*, 38(4):2966 – 2973.
- Atzmueller, M., Chin, A., Janssen, F., Schweizer, I., and Trattner, C., editors (2016). *Big Data Analytics in the Social and Ubiquitous Context*, volume 9546 of *Lecture Notes in Computer Science*, Heidelberg, Germany. Springer Verlag.
- Baxter, M. and Rennie, A. (1997). Financial calculus: An introduction to derivative pricing. *British Actuarial Journal*, 3(02):485–495.
- Bielecki, T. R., Cousin, A., Crépey, S., and Herbertsson, A. (2014). Dynamic hedging of portfolio credit risk in a markov copula model. *Journal of Optimization Theory and Applications*, 161(1):90–102.
- Bingham, N. (2007). Regular variation and probability: The early years. *Journal of Computational and Applied Mathematics*, 200(1):357 – 363.
- Bjork, T. (2009). *Arbitrage Theory in Continuous Time*. Oxford University Press, 3 edition.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–54.
- Casella, G. and Berger, R. (2002). *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning.
- Company, M. G. T. (1996). Riskmetrics technical document. Technical Report 2, JP Morgan and Reuters, <https://www.msci.com/documents/10199/5915b101-4206-4ba0-ae2-3449d5c7e95ar>. An optional note.
- Cox, J. and Rubinstein, M. (1985). *Options Markets*. Prentice-Hall.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38.
- Diaconis, P. (2009). The markov chain monte carlo revolution. *Bulletin of the American Mathematical Society*, 46:179textendash205.

- Dobrow, R. (2016). *Introduction to Stochastic Processes with R*. Wiley.
- Douglas Rivas, R. D. (2019). Clases procesos estocásticos.
- Duffie, D. (2010). *Dynamic Asset Pricing Theory*. Princeton Series in Finance. Princeton University Press.
- Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009). Frailty correlated default. *Journal of Finance*, 64(5):2089–2123.
- Duffie, D., Saita, L., and Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3):635 – 665.
- Edwin O. Fischer, Robert Heinkel, J. Z. (1989). Dynamic capital structure choice: Theory and tests. *The Journal of Finance*, 44(1):19–40.
- Föllmer, H. and Schied, A. (2004). *Stochastic Finance: An Introduction in Discrete Time*. De Gruyter studies in mathematics. Walter de Gruyter.
- Gregoriou, G. (2006). *Advances in Risk Management*. Finance and Capital Markets Series. Palgrave Macmillan UK.
- Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Probability and Random Processes. OUP Oxford.
- Hamilton, J. and Press, P. U. (1994). *Time Series Analysis*. Princeton University Press.
- Hillegeist, S. A., Keating, E. K., Cram, D. P., and Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1):5–34.
- Hull, J. (2006). *Options, futures, and other derivatives*. Pearson Prentice Hall, Upper Saddle River, NJ [u.a.], 6. ed., pearson internat. ed edition.
- Hull, J., Predescu, M., and White, A. (2004). The relationship between credit default swap spreads, bond yields, and credit rating announcements. *Journal of Banking & Finance*, 28(11):2789 – 2811. Recent Research on Credit Ratings.
- Jarrow, R., Maksimovic, V., and Ziemba, W. (1995). *Finance*. Handbooks in Operations Research. Elsevier.
- Karatzas, I. and Shreve, S. (1998). *Methods of Mathematical Finance*. Applications of Mathematics: Stochastic Modelling and Applied Probability. Springer.
- Karlin, S. and Taylor, H. (2012). *A First Course in Stochastic Processes*. Elsevier Science.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.

- Lamberton, D. and Lapeyre, B. (2011). *Introduction to Stochastic Calculus Applied to Finance, Second Edition*. Chapman and Hall/CRC Financial Mathematics Series. CRC Press.
- Leland, H. E. (1994). Corporate debt value, bond covenants, and optimal capital structure. *JOURNAL OF FINANCE*, 49(4):1213–1252.
- Martínez, F. (2006). *Riesgos financieros y económicos : productos derivados y decisiones económicas bajo incertidumbre*. THOMSON.
- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management : concepts, techniques and tools*. Princeton series in finance. Princeton University Press, Princeton (N.J.).
- Musiela, M. and Rutkowski, M. (2006). *Martingale Methods in Financial Modelling*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg.
- Overbeck, L. and Schmidt, W. (2003). *Modeling Default Dependence with Threshold Models*. Arbeitsberichte. HfB.
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1):101–124.
- Snyder, D. L. and Miller, M. I. (1991). Random point processes in time and space.
- Spiliopoulos, K. (2015). *Systemic Risk and Default Clustering for Large Financial Systems*, pages 529–557. Springer International Publishing, Cham.
- Stekhoven, D. J. and Bühlmann, P. (2011). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118.
- Tankov, P. (2003). *Financial Modelling with Jump Processes*. Chapman and Hall/CRC Financial Mathematics Series. CRC Press.
- Vassalou, M. and Xing, Y. (2004). Default risk in equity returns. *Journal of Finance*, 59(2):831–868.
- Wei, G. C. G. and Tanner, M. A. (1990). A monte carlo implementation of the em algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704.
- Wu, C. F. J. (1983). On the convergence properties of the em algorithm. *Ann. Statist.*, 11(1):95–103.
- Zhou, C. (2001). The term structure of credit spreads with jump risk. *Journal of Banking & Finance*, 25(11):2015–2040.