



Avances recientes en **valuación de activos** y **administración de riesgos**

VOLUMEN 4

Coordinadores:
Francisco Ortiz Arango
Francisco López Herrera
Francisco Venegas Martínez



Avances recientes en **valuación de activos y administración de riesgos** • VOLUMEN 4 • Francisco Ortiz / Francisco López / Francisco Venegas

**Avances recientes
en valuación de activos y
administración de riesgos**

VOLUMEN 4

Francisco Ortiz Arango

Francisco López Herrera

Francisco Venegas Martínez

Editores

Primera edición: Enero de 2013

Universidad Panamericana

Campus Ciudad de México

Derechos reservados

© Centros Culturales de México, A.C.

Derechos reservados

© Francisco Ortiz Arango (editor)

Derechos reservados

© Francisco López Herrera (editor)

Derechos reservados

© Francisco Venegas Martínez (editor)

ISBN: 978-607-7905-06-6

Prohibida su reproducción total o parcial por cualquier medio, sea electrónico, tradicional o fotomecánico.

Impreso en México

ÍNDICE GENERAL

Introducciónvi

PRIMERA PARTE:

ADMINISTRACIÓN DE RIESGOS

1. Solución a la ecuación diferencial parcial del modelo de volatilidad estocástica de Heston1

Ambrosio Ortiz Ramírez (IPN)

2. Riesgos estratégicos y adquisiciones internacionales: valuación con opciones simultáneas compuestas del crecimiento global de CEMEX19

Alejandra Cabello (UNAM) y Christian Bucio (UNAM)

3. Concentración de la Exposición al Riesgo de los retornos Accionarios en la Bolsa Mexicana de Valores en el Periodo 2004-2012, un Enfoque de Componentes Principales
51

Humberto Valencia (ITESM Cd. de México)

4. Modelos de calificación crediticia: Técnicas estadísticas tradicionales y de aprendizaje maquina67

Adán Díaz Hernández (U. Anáhuac-Norte) y John Goddard Close(UAM-I)

5. El riesgo del Sistema Financiero Internacional sobre el Crecimiento123

Salvador Rivas-Aceves (UAEH), Estrella Castillo Ramírez (UAEH) y José Francisco Martínez Sánchez (UAEH)

6. Determinación de la probabilidad de incumplimiento para empresas que cotizan en la BMV utilizando el modelo de Merton con activos con características de persistencia
143

Guillermo Sierra Juárez (CUCEA-U de G) y Katia García Loredó (CUCEA-U de G)

7. Normas Internacionales de Información Financiera (IFRS) en la Administración de Riesgos177

Noemí Vásquez Quevedo (ITESM Cd. de México)

SEGUNDA PARTE:

VALUACIÓN DE ACTIVOS

8. Aplicación de los modelos de valuación de bonos con incumplimiento a dos empresas mexicanas193

José Benito Díaz Hernández (UAM-I) y Patricia Saavedra Barrera (UAM-I)

9. Modelo de valuación de opciones con volatilidad estocástica mediante simulación Monte Carlo 217

Adrián Hernández del Valle (ESE-IPN) y Claudia I. Martínez García (ESE-IPN)

10. Análisis de la variación de precios mediante caminatas aleatorias: Weierstrass y fluctuaciones libres de escala 233

F. Javier Almaguer (UANL), Homero de la Fuente (UANL), Javier Morales (UANL), Jonás Velasco (UANL) y Roberto Soto (UANL)

11. Un Modelo GARCH con Asimetría Condicional Autorregresiva para modelar series de tiempo: Una aplicación para los rendimientos del Índice de Precios y Cotizaciones de la BMV 247

Rocío Durán-Vázquez (U. de las Américas), Arturo Lorenzo-Valdés (U. de las Américas) y Antonio Ruiz-Porras (CUCEA-U. de G.)

12. Sobre la no unicidad de soluciones de problemas de decisión de consumo y portafolio en presencia de saltos 263

Francisco Venegas-Martínez (ESE-IPN), Abigail Rodríguez Nava (UAM-X) y Francisco López-Herrera (FCA-UNAM)

13. Una nota sobre el uso de Martingalas en el modelo estándar de mercado para valorar derivados de tasas de interés 285

Carlos Alexander Grajales Correo (UA-Colombia) y Francisco Venegas-Martínez (ESE-IPN)

TERCERA PARTE:

NUEVAS PERSPECTIVAS

14. Utilidad Diferencial Recursiva Estocástica (UDRE) vs. Hamilton-Jacobi-Bellman (HJB) 293

Francisco Venegas-Martínez (ESE-IPN), Isela Elizabeth Téllez León (ESE-IPN) y Francisco Ortiz-Arango (ECEE-UP)

15. Seguimiento y pronóstico de precios de tres acciones que cotizan en la BMV empleando Redes Neuronales Diferenciales 311

Francisco Ortiz-Arango (ECEE-UP), Agustín I. Cabrera-Llanos (UPIBI-IPN) y Fernando Cruz-Aranda (ECEE-UP)

16. Uso de un Modelo de Redes Bayesianas para Medir el Riesgo Operacional de una Empresa Financiera 327

Griselda Dávila Aragón (ECEE-UP)

Introducción

En esta ocasión el Volumen Cuatro la obra Avances Recientes en Valuación de Activos y Administración de Riesgos, se enriquece de manera sustancial con la incorporación en la coordinación editorial del connotado investigador mexicano Francisco Venegas Martínez, ampliamente reconocido por sus grandes aportaciones al área de la Economía Financiera, tanto en el ámbito académico como en el de la práctica financiera. En este cuarto volumen puede palpase la madurez que ha ido adquiriendo esta obra a lo largo de sus distintas ediciones. La obra está integrada por **dieciseis capítulos**, en los cuales se ha tratado de incluir algunos de los grandes temas de mayor interés teórico y práctico en las áreas de valuación de activos financieros y la administración de riesgos.

La obra sigue su tradición fundacional de incluir los trabajos de colegas de distintas corrientes y temas de especialidad en el estudio de las Finanzas, así como de distintas instituciones educativas donde colaboran como investigadores quienes participan como autores y coautores de los capítulos que integran este volumen, dando cuenta con ello de la importancia que se le ha atribuido a la investigación en estos tiempos recientes, trascendiendo las paredes físicas de dichas instituciones. De esta forma la obra se ha convertido en un recinto en el cual se presenta una muestra de los resultados que se han alcanzado como frutos de los esfuerzos que se llevan a cabo, día con día, dentro de los distintos grupos de investigación en México e incluso colegas del extranjero, con el afán de lograr avances en el conocimiento que enriquezcan la comprensión sobre el funcionamiento cotidiano del medio financiero y sobre las decisiones que se toman en este medio.

Por la razón anterior ha sido posible tener la participación como autores de los capítulos de este cuarto volumen, a académicos de distintas instituciones mexicanas de enseñanza superior tanto públicas como privadas: el Instituto Politécnico Nacional, la Universidad Nacional Autónoma de México, la Universidad de Guadalajara, la Universidad Panamericana, la Universidad de las Américas, la Universidad Autónoma Metropolitana, el Instituto Tecnológico y de Estudios Superiores de Monterrey, EGADE del Tecnológico de Monterrey, la Universidad Anáhuac, la Universidad Autónoma de Estado de Hidalgo, Universidad Autónoma de Nuevo León y de la hermana de República de Colombia, la Universidad de Antioquía.

La respuesta de nuestros colegas a la convocatoria para integrar el cuarto volumen de Avances Recientes en Valuación de Activos y Administración de Riesgos, constituye la evidencia palpable de que indubitablemente existe en el medio académico un genuino y profundo interés por contribuir tanto al desarrollo de un medio financiero sano, eficiente y capaz de jugar un papel importante en la economía, así como el ánimo de participar en el desarrollo de herramientas útiles y adecuadas para la toma de mejores decisiones en el terreno práctico de las finanzas.

El interés de los investigadores que concurrieron a nuestro llamado se plasma en trabajos de investigación en los cuales se analizan diversos problemas de importancia desde el punto de vista de los practicantes del mercado, el público inversionista y las autoridades reguladoras. El resultado del gran interés mostrado por nuestros colegas investigadores en Finanzas se plasma en la calidad de los trabajos de investigación presentados, buscando que cada uno de los participantes en el libro proponga en sus trabajos, modelos y técnicas novedosas que faciliten las tareas de análisis y medición que son necesarias para la administración de riesgos y valuación de activos.

Cabe resaltar el hecho de que la relevancia de los trabajos incluidos en este volumen no se agota en los elementos que aportan para la práctica profesional de las finanzas en sus distintas vertientes, estos trabajos incluyen aportaciones significativas para otros investigadores e inclusive para la docencia en materias propias de las ciencias financieras y las disciplinas afines que también se ocupan del estudio de las finanzas en niveles de posgrado.

Siguiendo la pauta de los volúmenes anteriores, de acuerdo con la temática abordada, las contribuciones de los investigadores para este cuarto volumen se han agrupado en tres partes:

Primera Parte: Administración de Riesgos

Segunda Parte: Valuación de Activos

Tercera Parte: Nuevas Perspectivas

Los capítulos que integran la **primera parte** se enfocan en el estudio de la problemática de la Administración de Riesgos. En el primer capítulo se plantea a detalle la solución de la ecuación diferencial parcial del modelo de Heston usado para medir volatilidad estocástica; la relación entre los riesgos estratégicos con las adquisiciones internacionales se pone de manifiesto en el capítulo 2, donde se ejemplifica utilizando opciones reales de crecimiento para el caso de la empresa CEMEX; en el capítulo 3 se presentan las condiciones y consecuencias de la concentración de la exposición al riesgo de los rendimientos obtenidos por la BMV durante el periodo 2004-2012; utilizando técnicas estadísticas tradicionales y lo que se denomina aprendizaje maquina se lleva a cabo un esquema de calificación crediticia (capítulo 4); se desarrolla un planteamiento en el cual se pone de manifiesto cómo el Riesgo Financiero pone en riesgo al Crecimiento (capítulo 5); utilizando el modelo de Merton se llevó a cabo la determinación de la probabilidad de incumplimiento para empresas que cotizan en la BMV, considerando los activos que presentaban características de persistencia (capítulo 6) y en el capítulo 7 se presenta un análisis detallado de las Normas Internacionales de Información Financiera y sus consecuencias en la Administración de Riesgos.

La **segunda parte** enfocada a la Valuación de Activos, se inicia con una aplicación de modelos de valuación de bonos con incumplimiento, ejemplificándolo con dos empresas mexicanas (capítulo 8); el capítulo 9 desarrolla un modelo de valuación de opciones con volatilidad estocástica empleando simulación de Monte Carlo; posteriormente en el capítulo 10 se lleva a cabo un análisis detallado de la variación de precios empleando caminatas aleatorias con dos variantes, la de Weierstrass y la de fluctuaciones libres de escala; para modelar los rendimientos del IPC de la BMV, (capítulo 11) se emplea un Modelo GARCH con Asimetría Condicional Autorregresiva; en el capítulo 12, se desarrolla un análisis teórico sobre la no unicidad de las soluciones del problema de decisión de consumo y portafolios con la presencia de saltos; el capítulo 13 desarrolla un planteamiento analítico sobre la conveniencia del uso de Martingalas en un modelo de mercado estándar en que se pretenda valorar derivados de tasas de interés.

En la **tercera parte** dedicada a las nuevas perspectivas del tratamiento de problemas financieros, se muestra como primera aplicación novedosa un estudio comparativo entre la técnica de Utilidad Diferencial Recursiva Estocástica (UDRE) contra el método clásico de optimización en control óptimo estocástico formulado por Hamilito-Jacobi-Bellman para analizar las decisiones de consumo e inversión en un portafolio que contiene un activo riesgoso (capítulo 14); en el capítulo 15 se vuelve a utilizar la metodología de las Redes Neuronales Diferenciales para llevar a cabo en proceso de seguimiento y pronóstico de precios de cierre de tres acciones que cotizan en la BMV; el último capítulo de este libro muestra el desarrollo de un modelo de Riesgo Operacional de una empresa financiera mexicana, el cual está basado en la metodología de las Redes Bayesianas, se presenta una breve introducción a esta metodología e incluso se llevan a cabo "corridos" numéricos para obtener valores reales de aplicabilidad de este modelo en la institución donde se implementó.

Además de nuestro agradecimiento a los autores por haber confiado en nuestra capacidad para editar sus textos, queremos expresar nuestro agradecimiento al grupo de arbitros que dictaminaron cada uno de los capítulos de este cuarto volumen, porque no obstante su atareada agenda de trabajo, aceptaron gustosamente este encargo, poniendo de esta forma su grano de arena para que estas páginas vieran la luz. También agradecemos el apoyo brindado por el MBA Antonio Castro D'Franchis, Director de la Escuela de Ciencias Económicas y Empresariales de la Universidad Panamericana, quien ha apoyado fervientemente la elaboración de este libro desde el primer volumen hace tres años. Para ser justos, agradecemos también a nuestras respectivas familias el apoyo que nos brindaron mediante su comprensión por el tiempo que dejamos de dedicarles en aras de concluir las tareas editoriales requeridas por esta obra.

Finalmente, para que podamos considerar que nuestra labor editorial ha sido completa falta que resulte de provecho y del agrado de sus lectores, por lo que esperamos que sean ustedes, los

lectores, quienes nos ayuden a completar la tarea que hemos llevado a cabo, la cual por cierto fue placentera y esperamos que ahora ustedes compartan dicho placer.

Francisco Ortiz Arango
Francisco López Herrera
Francisco Venegas Martínez
México, D.F., diciembre de 2012

Capítulo 1

Solución a la ecuación diferencial parcial del modelo de volatilidad estocástica de Heston

Ambrosio Ortiz-Ramírez

Sección de Estudios de Posgrado e Investigación, Escuela Superior de Economía, I.P.N.

Resumen

En este capítulo por medio de portafolios replicantes se plantea la ecuación diferencial parcial del modelo de volatilidad estocástica de Heston (1993), asimismo al hacer los cambios de variable apropiados la ecuación es expresada en términos del logaritmo del subyacente, por lo que una vez obtenidas las funciones características de que la opción termine Dentro del dinero, las probabilidades neutrales al riesgo de la opción son determinadas con ayuda del Teorema de inversión. De acuerdo con Heston se supone que tales funciones características tienen una forma log-lineal, entonces como una aplicación del Teorema de Feynman-Kač y condiciones de frontera adecuadas se resuelve la ecuación diferencial parcial.

1.1 Introducción

A pesar de su popularidad entre practicantes e investigadores, es ampliamente conocido que el modelo de Black y Scholes (ByS) refleja una pequeña parte de la complejidad de los mercados de derivados financieros y produce un sesgo persistente en la determinación de precios de opciones. Un supuesto claramente inconsistente de este modelo es que los rendimientos del subyacente se distribuyen normalmente. De hecho, existe en la literatura evidencia empírica de que la distribución de rendimientos presenta asimetría negativa y exceso de curtosis. Además, el supuesto de volatilidad constante es una limitante significativa y contrasta con las sonrisas de volatilidad (*volatility smile*) y las muecas (*smirk*) implícitas en los precios de opciones que se negocian en los mercados.

En las últimas dos décadas la investigación ha sido pródiga en lo que respecta a modelos que descansen en supuestos más realistas que perfeccionen algunos de los supuestos restrictivos del modelo de ByS. Se han propuesto un conjunto de modelos sofisticados al suponer diferentes procesos estocásticos que conducen la dinámica de los activos subyacentes como acciones, índices bursátiles, tasas de interés, tipos de cambio, materias primas, y su correspondiente prima de riesgo, véase Gatheral (2006). Un inconveniente de estos modelos es pocas veces se dispone de fórmula en forma cerrada y la ventaja de contar con supuestos más realistas se compensa por los costos de implementación y calibración. Entre dichos modelos alternativos los modelos de volatilidad estocástica (VE) son de particular interés. En un enfoque de volatilidad estocástica se supone que la volatilidad cambia en el tiempo de acuerdo a un proceso estocástico usualmente especificado

mediante una ecuación diferencial estocástica. De esta manera, los modelos de VE presentan resultados empíricos que reproducen regularidades empíricas mostradas por la densidad neutral al riesgo implícita en las cotizaciones de un mercado de opciones.

Los trabajos seminales de Hull y White (1987), Stein y Stein (1991) y Heston (1993) condujeron al desarrollo de los modelos de volatilidad estocástica, estos modelos son de dos factores; en los cuales un factor conduce la dinámica estocástica de la volatilidad. Al respecto, se han propuesto diferentes modelos para el proceso de volatilidad, incluyendo el movimiento geométrico Browniano y procesos con reversión a la media del tipo Ornstein-Uhlenbeck como en Schöbel y Zhu (1999) que supone que los dos movimientos Brownianos estén correlacionados y obtiene la función característica del logaritmo del precio con una expresión compuesta por un sistema de ecuaciones diferenciales estocásticas complicado. Posteriormente Zhu (2010) propone un modelo de volatilidad estocástica de doble raíz. En el Cuadro 1.1 mostramos un resumen de modelos de volatilidad estocástica de dos factores:

Modelo	Autor
(i) $dv_t = \kappa v_t dt + \sigma dW_t$	Johnson/Shano (1987)
(ii) $d \ln v_t = \kappa(\theta - \ln v_t) dt + \sigma dW_t$	Wiggins (1987)
(iii) $dv_t^2 = \kappa v_t^2 dt + \sigma v_t^2 dW_t$	Hull/White (1987)
(iv) $dv_t^2 = (a + \beta v_t^2) dt + \sigma \sqrt{v_t^2} dW_t$	Hull/White (1988)
(v) $dv_t = \kappa(\theta - v_t) dt + \sigma dW_t$	Stein/Stein (1991), Schöbel/Zhu (1999)
(vi) $dv_t^2 = \kappa(\theta - v_t) dt + \sigma \sqrt{v_t} dW_t$	Heston (1993)
(vii) $dv_t^2 = \kappa(\theta - v_t^2) dt + \sigma v_t^{3/2} dW_t$	Lewis (2000)
(viii) $dv_t^2 = \kappa(\theta - \sqrt{v_t} - \lambda v_t) dt + \sigma \sqrt{v_t} dW_t$	Zhu (2010)

Cuadro 1.1: Cuadro resumen de modelos de volatilidad estocástica de dos factores.

Otros modelos que consideran saltos en la dinámica del subyacente son Bates (1996) y Bakshi, Cao, y Chen (1997), los cuales son una extensión con volatilidad estocástica del modelo de difusión con saltos de Merton (1976), la lista no es exhaustiva, pero si abrumadora si se debe elegir entre varias alternativas. Para hacer la revisión más difícil el número de posibles modelos de valuación de opciones es prácticamente infinito. Observemos que todos los modelos de valuación de opciones tienen tres supuestos básicos: el proceso para el precio del subyacente (generalmente una distribución), el proceso de la tasa de interés y la prima de riesgo de mercado, para cada hipótesis hay muchas opciones posibles¹, por lo cual este trabajo se concentra en el modelo de Heston (1993), dejando en la agenda de investigación otros modelos para su análisis respectivo.

El presente capítulo plantea la ecuación diferencial parcial del modelo de Heston y mediante cambios de variable, la aplicación del Teorema de inversión y el Teorema de Feynman-Kač se obtiene la solución a dicha ecuación analíticamente. Este capítulo está organizado como sigue, en la siguiente sección planteamos un marco general de volatilidad estocástica, en la tercer sección discutimos las características teóricas del modelo de Heston junto con su función de densidad marginal con diferentes valores para los parámetros de correlación y volatilidad de la volatilidad. En la cuarta y quinta sección planteamos y resolvemos la ecuación diferencial parcial que satisface el precio de la opción de compra en este modelo respectivamente, en la sexta sección recurrimos a la ecuación diferencial ordinaria de Riccati para resolver la ecuación diferencial parcial, posteriormente presentamos un resumen de los resultados obtenidos. Concluimos en la octava sección.

1.2 Un marco de volatilidad estocástica

¹ El lector interesado en una compilación selecta de los modelos de volatilidad estocástica y su relación con econometría financiera puede consultar Shephard, N. (2005).

De acuerdo con la literatura los modelos de volatilidad estocástica con sus respectivas extensiones resultan ser más apropiados para la determinación de precios de opciones en tiempo continuo con tal supuesto, aunque también pueden ser planteados en tiempo discreto como se describe en Harvey, Ruiz y Shepard (1994), pero no son tan populares en la valuación de derivados como los modelos GARCH y sus extensiones.

Un modelo general de volatilidad estocástica es planteado como un sistema de dos ecuaciones diferenciales estocásticas como sigue:

$$\begin{aligned} dS_t &= \mu_t S_t dt + \sqrt{v_t} S_t dW_t \\ dv_t &= \alpha(v_t) dt + \beta(v_t) dW_t^v. \end{aligned}$$

El efecto de apalancamiento se genera si el rendimiento de activo y las innovaciones de la volatilidad se correlacionan:

$$\mathbb{E}(dW_t dW_t^v) = \rho dt.$$

El precio del derivado se compone ahora de una función que depende de la volatilidad, el tiempo y el precio del activo subyacente:

$$P_t = P(t, S_t, v_t).$$

Cuando introducimos volatilidad estocástica relajamos dos de los supuestos del modelo de Black y Scholes, el primero que un solo movimiento Browniano genera incertidumbre y además de que los mercados son completos. Los modelos de volatilidad estocástica consideran dos movimientos Brownianos, y es por ello que el precio del derivado no se puede replicar, por lo que el mercado no es completo. Esto significa que existe un número infinito de precios derivados que invalidan el supuesto de no arbitraje.

De esta manera, si incluimos otro instrumento de cobertura con un derivado entonces el mercado es completo. Por lo tanto los derivados ahora deberán valuarse en relación unos con otros, así como el activo subyacente. Esta es una de las razones de que los modelos de volatilidad estocástica se puedan calibrar dado un conjunto de precios de mercado de opciones, es decir, dado un conjunto de parámetros iniciales y mediante un algoritmo numérico, determinar los parámetros del modelo que minimicen la distancia entre los precios de mercado y los obtenidos con el modelo teórico.

Otra manera de tratar con este detalle es por medio de la idea de riesgo de volatilidad, introducido por el segundo Browniano W_t^v (y en particular por la parte de \widetilde{W}_t de W_t^v que es ortogonal a W_t , ya que se puede escribir $W_t = \rho W_t^v + \sqrt{1 - \rho^2} \widetilde{W}_t$). El activo subyacente no depende de este Browniano, y por lo tanto el riesgo generado por este Browniano no está incluido en el precio del activo. Por otro lado, por supuesto, el riesgo de W_t está incrustado en la prima de riesgo $\mu_t - r_t$. Un inversionista podría ser adverso al riesgo, y aunque esta aversión al riesgo no se exhibe en el mercado para el activo subyacente, si aparecerá en el mercado de opciones, ya que estos contratos dependen directamente de v_t . Mediante un derivado podemos identificar la prima de riesgo, y entonces obtener precios de otros derivados.

1.3 El modelo de volatilidad estocástica de Heston

En 1993 Steve L. Heston en su artículo “A Closed-Form Solution for Options with Stochastic Volatility” publicado en el *Review of Financial Studies* valúa una opción sobre una acción con volatilidad estocástica. Una característica relevante en el artículo de Heston es que obtiene las funciones características de las probabilidades neutrales al riesgo como soluciones de una ecuación diferencial parcial de segundo orden. Por medio de estas probabilidades neutrales al riesgo se obtiene una fórmula similar a la de Black y Scholes para valorar una opción europea de compra, el precio de la opción de venta se puede obtener con la paridad *put-call*.

En el modelo de volatilidad estocástica de Heston (1993) la varianza es conducida por un proceso de raíz cuadrada (también denominado como proceso de Feller, desarrollado en Feller en

1951), también se utiliza en el modelo de tasa corta Cox, Ingersoll, y Ross (1985). Si denotamos por $v_t = \sigma_t^2$ como la varianza, la dinámica del modelo de Heston está dada sistema de dos ecuaciones diferenciales estocásticas:

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_{1,t} \\ dv_t &= \kappa(\theta - v_t)dt + \sigma \sqrt{v_t} dW_{2,t}^v \\ \mathbb{E}(dW_{1,t} dW_{2,t}^v) &= \rho dt. \end{aligned} \quad (1.1)$$

El parámetro de volatilidad de la varianza σ controla la curtosis, mientras que el parámetro de correlación ρ determina la asimetría de la densidad de los rendimientos del activo subyacente. El proceso de la varianza exhibe reversión a la media, teniendo como atractor el parámetro de varianza de largo plazo θ . El parámetro κ define la fuerza de reversión a la media, y establece la rapidez con que el *skew* de volatilidad se aplana.

Como señala Heston, la volatilidad v_t al inicio sigue un proceso de Ornstein-Uhlenbeck con nivel reversión a la media cero, esto es:

$$d\sqrt{v_t} = -\beta\sqrt{v_t}dt + \delta dW_{2,t}^v, \quad (1.1a)$$

al aplicar el lema de Itô al cuadrado de la volatilidad, denominado como la varianza del rendimiento instantáneo del subyacente, este sigue un proceso de raíz cuadrada dado por:

$$dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}dW_{2,t}^v, \quad (1.1b)$$

donde:

$$\beta = \frac{\kappa}{2}, \delta = \frac{\sigma}{2}, \theta = \frac{\delta^2}{\kappa}.$$

Si κ, θ y σ satisfacen las siguientes condiciones:

$$2\kappa\theta > \sigma^2, v_0 > 0, \quad (1.1c)$$

se puede demostrar que la varianza v_t es siempre positiva y su proceso dado en (1.1b) está bien definido bajo estas condiciones. La condición (1.1c) se conoce también como condición de proceso de raíz cuadrada de Feller.

El modelo de Heston es utilizado en la industria y muy citado en la literatura debido a sus características especiales², además pertenece a una clase general de modelos afines propuestos por Duffie *et al* (2000), y la función característica de logaritmo del precio está dada analíticamente. En particular, tiene una forma afín-exponencial.

Por otro lado, sea $x_t = \ln(S_t/S_0) - \mu t$, podemos expresar el modelo de Heston en (1.1) en términos de (log-) rendimiento centrado x_t y v_t (Čížek *et al.* (2011)). El proceso es caracterizado por la transición $P_t(x, v|v_0)$ para tener el (log-) rendimiento x y la varianza v al tiempo t dado un rendimiento inicial $x = 0$ y varianza v_0 al tiempo $t = 0$. La dinámica temporal de $P_t(x, v|v_0)$ es conducida por la siguiente ecuación de Fokker-Planck (o forward Kolmogorov):

$$\begin{aligned} \frac{\partial}{\partial t} P &= \kappa \frac{\partial}{\partial v} \{(v - \theta) P\} + \frac{1}{2} \frac{\partial}{\partial x} (vP) + \rho \sigma \frac{\partial^2}{\partial x \partial v} (vP) \\ &+ \frac{1}{2} \frac{\partial^2}{\partial x^2} (vP) + \frac{\sigma^2}{2} \frac{\partial^2}{\partial v^2} (vP). \end{aligned} \quad (1.2)$$

² Su popularidad proviene del hecho de que determina precios de opciones europeas y exóticas eficientemente por medio de la Transformada Rápida de Fourier denominado algoritmo FFT.

Al resolver esta ecuación se obtiene la siguiente fórmula semi-analítica para la densidad de los rendimientos centrados de x , dado cambio en el precio en un intervalo de tiempo t , de acuerdo con Dragulescu y Yakovenko (2002):

$$P_t(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\xi x + F_t(\xi)} d\xi, \quad (1.3)$$

donde:

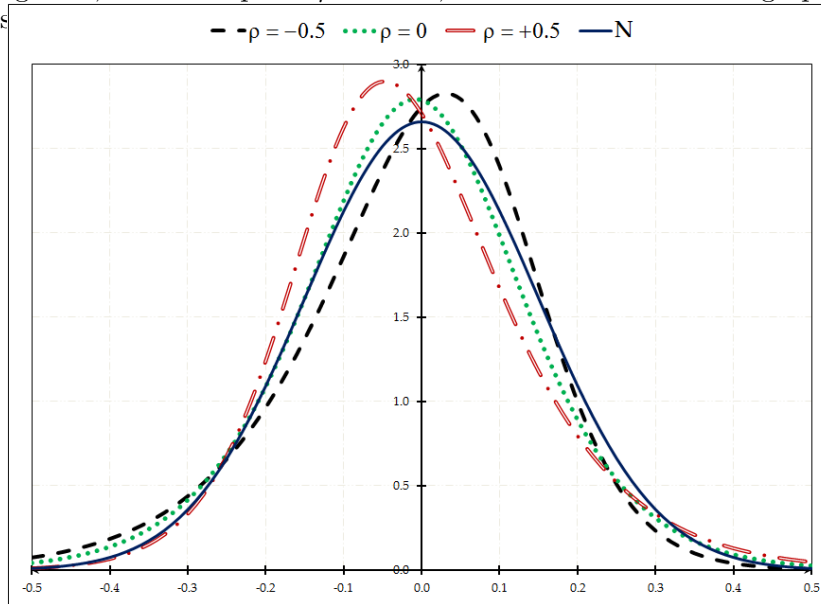
$$F_t(\xi) = \frac{\kappa\theta}{\sigma^2}\gamma t - \frac{2\kappa\theta}{\sigma^2} \log \left(\cosh \frac{\Omega t}{2} + \frac{\Omega^2 - \gamma^2 + 2\kappa\gamma}{2\kappa\Omega} \sinh \frac{\Omega t}{2} \right),$$

$$\gamma = \kappa + i\rho\sigma\xi, \quad y \quad \Omega = \sqrt{\gamma^2 + \sigma^2(\xi^2 - i\xi)}.$$

A continuación mostramos gráficamente la densidad marginal del modelo de Heston dada por (1.3) para diferentes valores de ρ , con parámetros:

κ	θ	σ	ρ	x
2	0.025	0.225	-0.5, 0, +0.5	(-0.5, 0.5)

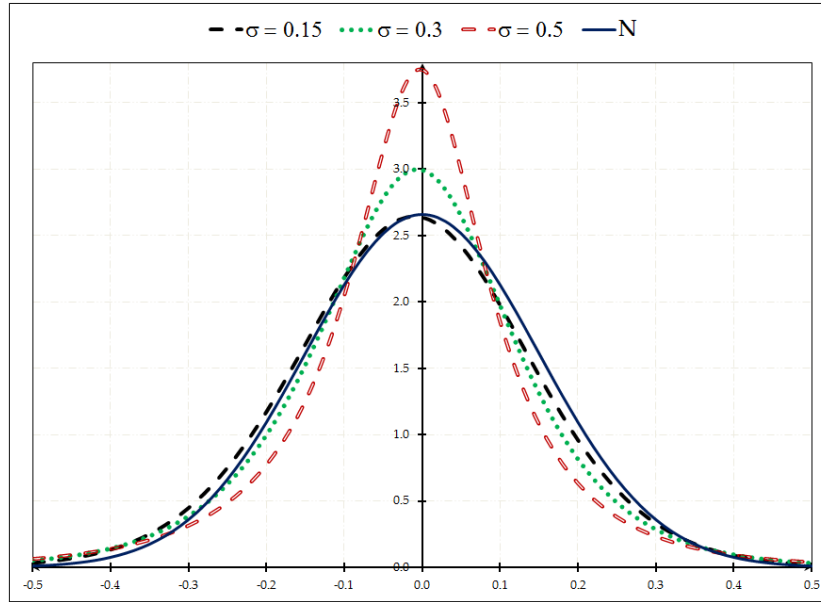
La Gráfica 1.1 muestra la función de densidad marginal del modelo de Heston para diferentes valores de ρ y la comparamos con la densidad de una distribución normal $\mathcal{N} \sim (0, 0.15)$. Observamos que el parámetro de correlación efectivamente controla el sesgo de la densidad. Si $\rho = -0.5$ la densidad muestra sesgo negativo, mientras que si $\rho = +0.5$, la densidad muestra sesgo positivo, en el caso de $\rho = 0$, la dens



Gráfica 1.1: Función de densidad marginal del modelo de Heston para diferentes valores de ρ . Fuente: elaboración propia.

De la misma manera mostramos la densidad del modelo de Heston para diferentes valores del parámetro de volatilidad de la varianza σ , con parámetros dados por:

κ	θ	σ	ρ	x
2	0.025	0.15, 0.3, 0.5	0	(-0.5, 0.5)



Gráfica 1.2: Función de densidad marginal del modelo de Heston para diferentes valores de σ . Fuente: elaboración propia.

La Gráfica 1.2 muestra la función de densidad marginal del modelo de Heston para diferentes valores de σ y la comparamos con la densidad de una distribución normal $\mathcal{N} \sim (0, 0.15)$. Observamos que el parámetro de volatilidad de la volatilidad controla la curtosis de la densidad, por lo tanto, si σ es relativamente alta entonces la densidad presenta exceso de curtosis con su correspondiente efecto en las colas.

1.4 La ecuación diferencial parcial del modelo de Heston (1993)

Para deducir la ecuación diferencial parcial (EDP) de este modelo seguiremos la misma metodología de Black y Scholes, es decir, con una posición corta en el derivado y posteriormente administrar el riesgo de mercado en la parte estocástica. Esto dará lugar a una representación en forma de EDP del precio de la opción. Recordemos que en el mundo de ByS con volatilidad constante, era suficiente el activo subyacente y la cuenta bancaria para la cobertura, pero ahora como tenemos una fuente más de riesgo, estos dos instrumentos no serán suficientes para eliminar el riesgo de volatilidad, para cubrir la posición corta en el derivado vamos a utilizar la cuenta bancaria, el activo subyacente y un derivado adicional.

Consideremos un derivado V , y denotemos su función de pago por $V = V(t, S, v)$. La forma funcional de V dependerá de las condiciones del contrato, tales como el vencimiento, la función de pago, la opcionalidad, etc. Por lo tanto, el proceso de este derivado estará dado por $V_t = V(t, S_t, v_t)$. De acuerdo con el argumento de ByS, si conocemos la forma funcional de V , podemos determinar la dinámica del precio del derivado mediante una aplicación del Lema de Itô:

$$dV_t = \left(\alpha_t^V + \mu S_t \frac{\partial V}{\partial S} \right) dt + \beta_t^{V,S} dW_{1,t} + \beta_t^{V,v} dW_{2,t} \quad (1.4)$$

donde:

$$\begin{aligned}\alpha_t^V &= \frac{\partial V}{\partial t} + \alpha(v_t) \frac{\partial V}{\partial v} + \frac{1}{2} v_t S_t^2 \frac{\partial^2 V}{\partial S^2} + \frac{1}{2} \beta_t^2(v_t) \frac{\partial^2 V}{\partial v^2} + \rho \sqrt{v_t} S_t \beta(v_t) \frac{\partial^2 V}{\partial S \partial v}, \\ \beta_t^{V,S} &= \sigma S_t \frac{\partial V}{\partial S}, \\ \beta_t^{V,v} &= \beta(v_t) \frac{\partial V}{\partial v}.\end{aligned}$$

De la expresión observemos que si formamos un portafolio utilizando sólo el subyacente y la cuenta bancaria, no será posible replicar el proceso V_t , ya que la fuente de riesgo $W_{2,t}^v$ no se puede replicar. El mercado que se basa sólo en estos instrumentos es incompleto, ya que el derivado no puede replicarse. Pero se puede completar el mercado dinámicamente con otro derivado U , con función de pago $U(t, S, v)$, lo cual funciona, por supuesto, si U depende en realidad de $W_{2,t}^v$, que suele ser el caso.

Por conveniencia en la notación prescindimos de los subíndices en t : $S = S_t, v = v_t, W_1 = W_{1,t}, W_2 = W_{2,t}^v$. Sea un portafolio Π_t compuesto por: una opción $V = V(t, S, v)$, Δ unidades de un subyacente S y ϕ unidades de otra opción $U = U(t, S, v)$ para cubrir la volatilidad:

$$\Pi = V + \Delta S + \phi U, \quad (1.5)$$

con $\Pi = \Pi_t$. Si el portafolio es autofinanciable, el cambio en el valor del portafolio debido a cambios en sus componentes es:

$$d\Pi = dV + \Delta dS + \phi dU, \quad (1.6)$$

Al aplicar el Lema de Itô a V , derivamos con respecto a las variables t, S y v como sigue:

$$dV = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S} dS + \frac{\partial V}{\partial v} dv + \frac{1}{2} v S^2 \frac{\partial^2 V}{\partial S^2} dt + \frac{1}{2} \sigma^2 v \frac{\partial^2 V}{\partial v^2} dt + \rho \sigma v S \frac{\partial^2 V}{\partial v \partial S} dt,$$

dado que $(dS)^2 = v S^2 (dW_1)^2 = v S^2 dt$, $(dv)^2 = \sigma v dt$, y $dS dv = v S \sigma dW_1 dW_2 = v S \rho dt$, además de que $(dt)^2 = 0$ y $dW_1 dt = dW_2 dt = 0$. De igual forma al aplicar el Lema de Itô a U tenemos que:

$$dU = \frac{\partial U}{\partial t} dt + \frac{\partial U}{\partial S} dS + \frac{\partial U}{\partial v} dv + \frac{1}{2} v S^2 \frac{\partial^2 U}{\partial S^2} dt + \frac{1}{2} \sigma^2 v \frac{\partial^2 U}{\partial v^2} dt + \rho \sigma v S \frac{\partial^2 U}{\partial v \partial S} dt.$$

Al sustituir las ecuaciones anteriores en el cambio en el valor del portafolio y reordenando términos:

$$\begin{aligned}d\Pi &= dV + \Delta dS + \phi dU \\ &= \left\{ \frac{\partial V}{\partial t} + \frac{1}{2} v S^2 \frac{\partial^2 V}{\partial S^2} + \rho \sigma v S \frac{\partial^2 V}{\partial v \partial S} + \frac{1}{2} \sigma^2 v \frac{\partial^2 V}{\partial v^2} \right\} dt + \\ &\quad \phi \left\{ \frac{\partial U}{\partial t} + \frac{1}{2} v S^2 \frac{\partial^2 U}{\partial S^2} + \rho \sigma v S \frac{\partial^2 U}{\partial v \partial S} + \frac{1}{2} \sigma^2 v \frac{\partial^2 U}{\partial v^2} \right\} dt + \\ &\quad \left\{ \frac{\partial V}{\partial S} + \phi \frac{\partial U}{\partial S} + \Delta \right\} dS + \left\{ \frac{\partial V}{\partial v} + \phi \frac{\partial U}{\partial v} \right\} dv.\end{aligned} \quad (1.7)$$

1.4.1 El portafolio libre de riesgo

Para cubrir el portafolio contra cambios en el subyacente y volatilidad, los dos últimos términos de la ecuación (1.7) dados por dS y dv deben ser cero. Esto implica que las ponderaciones adecuadas son:

$$\phi = -\frac{\frac{\partial V}{\partial v}}{\frac{\partial U}{\partial v}}; \quad \Delta = -\phi \frac{\partial U}{\partial S} - \frac{\partial V}{\partial S},$$

además, como hemos hecho el portafolio libre de riesgo, entonces debe ganar la tasa libre de riesgo r , por lo tanto:

$$d\Pi = r\Pi dt = r(V + \Delta S + \phi U)dt.$$

Ahora, con los valores de ϕ y Δ el cambio en el valor del portafolio libre de riesgo es:

$$d\Pi = \left\{ \frac{\partial V}{\partial t} + \frac{1}{2}vS^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma vS \frac{\partial^2 V}{\partial v \partial S} + \frac{1}{2}\sigma^2 v \frac{\partial^2 V}{\partial v^2} \right\} dt + \\ \phi \left\{ \frac{\partial U}{\partial t} + \frac{1}{2}vS^2 \frac{\partial^2 U}{\partial S^2} + \rho\sigma vS \frac{\partial^2 U}{\partial v \partial S} + \frac{1}{2}\sigma^2 v \frac{\partial^2 U}{\partial v^2} \right\} dt$$

reescribiendo la expresión anterior:

$$d\Pi = (A + \phi B)dt \quad (1.8)$$

entonces:

$$A + \phi B = r(V + \Delta S + \phi U) \\ A - \frac{\frac{\partial V}{\partial v}}{\frac{\partial U}{\partial v}} B = r \left[V + \left(\left(-\frac{\frac{\partial V}{\partial v}}{\frac{\partial U}{\partial v}} \right) \frac{\partial U}{\partial S} - \frac{\partial V}{\partial S} \right) S - \frac{\frac{\partial V}{\partial v}}{\frac{\partial U}{\partial v}} U \right],$$

simplificando términos se convierte en la igualdad:

$$\frac{A - rV + rS \frac{\partial V}{\partial S}}{\frac{\partial V}{\partial v}} = \frac{B - rU + rS \frac{\partial U}{\partial S}}{\frac{\partial U}{\partial v}}. \quad (1.9)$$

1.4.2 La EDP de Heston en función del precio

El lado izquierdo de la ecuación (1.9) es una función que depende sólo V , y el lado derecho es una función que también depende sólo de U , recordemos que V y U son las funciones de pago de dos derivados arbitrarios (opciones), lo que significa que el cociente anterior será análogo cualquier derivado. Esto implica que ambos lados se pueden escribir como una función $f(S, v, t)$. Según Heston, esta función es $f(S, v, t) = -k(\theta - v) + \lambda(S, v, t)$ con $\lambda(S, v, t)$ es la prima al riesgo por volatilidad. Escribimos el lado izquierdo de la ecuación (1.9) como sigue:

$$\frac{A - rV + rS \frac{\partial V}{\partial S}}{\frac{\partial V}{\partial v}} = -k(\theta - v) + \lambda(S, v, t).$$

Al sustituir A y simplificar términos obtenemos la EDP de Heston expresada en términos del subyacente S :

$$\frac{\partial V}{\partial t} + \frac{1}{2}vS^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma vS \frac{\partial^2 V}{\partial v \partial S} + \frac{1}{2}\sigma^2 v \frac{\partial^2 V}{\partial v^2} \\ - rV + rS \frac{\partial V}{\partial S} + [k(\theta - v) - \lambda(S, v, t)] \frac{\partial V}{\partial v} = 0 \quad (1.10)$$

Esta es la ecuación (6) en Heston (1993) p. 329. Alternativamente, la EDP en ecuación anterior se puede expresar como:

$$\frac{\partial V}{\partial t} + \mathcal{A}V - rV = 0$$

donde:

$$\begin{aligned} \mathcal{A} &= rS \frac{\partial}{\partial S} + \frac{1}{2}vS^2 \frac{\partial}{\partial S^2} \\ &+ [k(\theta - v) - \lambda(S, v, t)] \frac{\partial}{\partial v} + \frac{1}{2}\sigma^2 v \frac{\partial^2}{\partial v^2} + \rho\sigma v S \frac{\partial}{\partial v \partial S} \end{aligned} \quad (1.11)$$

es el generador del modelo de Heston. Según explica Lewis (2000), la primera línea en la ecuación (1.11) es el generador del modelo ByS, mientras que la segunda añade las correcciones por volatilidad estocástica.

1.4.3 La EDP de Heston en términos del logaritmo del precio spot

Si $x = \ln S$ y al reescribir la EDP en términos de x, t y v en lugar de S, t y v , nos conduce a una expresión más sencilla de la EDP, junto con las siguientes derivadas:

$$\begin{aligned} \frac{\partial V}{\partial S} &= \frac{\partial V}{\partial x} \frac{\partial x}{\partial S} = \frac{\partial V}{\partial x} \frac{1}{S}, \quad \text{por la regla de la cadena.} \\ \frac{\partial^2 V}{\partial v \partial S} &= \frac{\partial}{\partial v} \left(\frac{\partial V}{\partial S} \right) = \frac{\partial}{\partial v} \left(\frac{1}{S} \frac{\partial V}{\partial x} \right) = \frac{1}{S} \frac{\partial^2 V}{\partial v \partial x}, \\ \frac{\partial^2 V}{\partial S^2} &= \frac{\partial}{\partial S} \left(\frac{\partial V}{\partial S} \right) = \frac{\partial}{\partial S} \left(\frac{1}{S} \frac{\partial V}{\partial x} \right) \\ &= -\frac{1}{S^2} \frac{\partial V}{\partial x} + \frac{1}{S} \frac{\partial^2 V}{\partial S \partial x} = -\frac{1}{S^2} \frac{\partial V}{\partial x} + \frac{1}{S} \frac{\partial}{\partial x} \left(\frac{\partial V}{\partial S} \right) \\ &= -\frac{1}{S^2} \frac{\partial V}{\partial x} + \frac{1}{S^2} \frac{\partial^2 V}{\partial x^2}, \quad \text{por la regla del producto.} \end{aligned}$$

Al sustituir estas derivadas en la ecuación (1.10), todos los términos que con S se cancelan y obtenemos la EDP de Heston en términos de $x = \ln S$:

$$\begin{aligned} \frac{\partial V}{\partial t} + \frac{1}{2}v \frac{\partial^2 V}{\partial x^2} + \left(r - \frac{1}{2}v \right) \frac{\partial V}{\partial x} + \rho\sigma v \frac{\partial^2 V}{\partial v \partial x} + \\ \frac{1}{2}\sigma^2 v \frac{\partial^2 V}{\partial v^2} - rV + [k(\theta - v) - \lambda v] \frac{\partial V}{\partial v} = 0, \end{aligned} \quad (1.12)$$

donde al igual que en Heston el precio de mercado del riesgo es una función lineal de la volatilidad, de manera que $\lambda(S, v, t) = \lambda v$.

1.5 El precio de la opción de compra

El precio de la opción de compra es el valor presente de la esperanza del valor intrínseco:

$$\begin{aligned} C_T(K) &= e^{-r\tau} \mathbb{E}[\max(S_T - K, 0)] \\ &= e^{-r\tau} \mathbb{E}[S_T \mathbf{1}_{S_T > K}] - e^{r\tau} K \mathbb{E}[\mathbf{1}_{S_T > K}] \\ &= e^{x_t} P_1(x, v, \tau) - e^{-r\tau} K P_2(x, v, \tau). \end{aligned} \quad (1.13)$$

En esta expresión $P_j(x, v, \tau)$ cada una representan la probabilidad de que la opción de compra termine *In-The-Money*, condicional al valor $x_t = \ln S_t$ del subyacente y sobre la volatilidad v_t al tiempo t , donde τ es el plazo al vencimiento, por lo tanto:

$$P_j(x, v, \tau) = \mathbb{P}(\ln S_T > \ln K),$$

para $j = 1, 2$. Estas probabilidades se obtienen bajo diferentes medidas de probabilidad. En la ecuación (1.13), el valor esperado $\mathbb{E}[\mathbb{1}_{S_T > K}]$ es la probabilidad de que la opción de compra termine *In-The-Money* bajo la medida original \mathcal{Q} que hacen Brownianos a $W_{1,t}$ y $W_{2,t}^v$ en la ecuación (1.1), por lo tanto

$$\mathbb{E}^{\mathcal{Q}}[\mathbb{1}_{S_T > K}] = \mathcal{Q}(S_T > K) = \mathcal{Q}(\ln S_T > \ln K) = P_2(x, v, \tau).$$

Al evaluar $e^{-r\tau} \mathbb{E}[S_T \mathbb{1}_{S_T > K}]$ en la ecuación (1.13) es necesario cambiar la medida \mathcal{Q} original a otra medida \mathcal{P} . Definamos la derivada de Radon-Nikodym:

$$Z_t = \frac{S_t/S_T}{B_t/B_T},$$

con: $B_t = \exp\left(\int_0^t r du\right) = e^{rt}$, entonces la segunda esperanza es:

$$\begin{aligned} e^{-r\tau} \mathbb{E}^{\mathcal{Q}}[S_T \mathbb{1}_{S_T > K}] &= \mathbb{E}^{\mathcal{Q}}\left[\frac{B_t}{B_T} S_T \mathbb{1}_{S_T > K}\right] \\ &= \mathbb{E}^{\mathcal{P}}\left[\frac{B_t}{B_T} S_T \mathbb{1}_{S_T > K} Z_t\right] \\ &= S_t \mathbb{E}^{\mathcal{P}}[\mathbb{1}_{S_T > K}] \\ &= e^{xt} \mathcal{P}(S_T > K) = e^{xt} P_1(x, v, \tau). \end{aligned}$$

Esto implica que el precio de la opción de compra en la ecuación (1.13) puede escribirse en términos de ambas medidas:

$$C_T(K) = S_t \mathcal{P}(S_T > K) - K e^{-r\tau} \mathcal{Q}(S_T > K).$$

Ahora obtengamos las derivadas de $C_T(K)$:

$$\begin{aligned} \frac{\partial C}{\partial t} &= e^x \frac{\partial P_1}{\partial t} - K r e^{-r\tau} P_2 - K e^{-r\tau} \frac{\partial P_2}{\partial t} \\ &= e^x \left[\frac{\partial P_1}{\partial t} \right] - K e^{-r\tau} \left[r P_2 + \frac{\partial P_2}{\partial t} \right], \end{aligned}$$

$$\begin{aligned} \frac{\partial C}{\partial x} &= e^x P_1 + e^x \frac{\partial P_1}{\partial x} - e^{-r\tau} K \frac{\partial P_2}{\partial x} \\ &= e^x \left[P_1 + \frac{\partial P_1}{\partial x} \right] - K e^{-r\tau} \left[\frac{\partial P_2}{\partial x} \right], \end{aligned} \tag{1.14}$$

$$\begin{aligned} \frac{\partial C}{\partial v} &= e^x \left[\frac{\partial P_1}{\partial v} \right] - K e^{-r\tau} \left[\frac{\partial P_2}{\partial v} \right], \\ \frac{\partial^2 C}{\partial v^2} &= e^x \left[\frac{\partial^2 P_1}{\partial v^2} \right] - K e^{-r\tau} \left[\frac{\partial^2 P_2}{\partial v^2} \right], \end{aligned}$$

y

$$\begin{aligned} \frac{\partial^2 C}{\partial x \partial v} &= e^x \frac{\partial P_1}{\partial v} + e^x \frac{\partial^2 P_1}{\partial v \partial x} - e^{-r\tau} K \frac{\partial^2 P_2}{\partial v \partial x} \\ &= e^x \left[\frac{\partial P_1}{\partial v} + \frac{\partial^2 P_1}{\partial v \partial x} \right] - K e^{-r\tau} \left[\frac{\partial^2 P_2}{\partial v \partial x} \right] \end{aligned}$$

En la primera línea de la ecuación anterior tenemos la derivada con respecto a t pero realmente necesitamos la derivada con respecto al vencimiento $\tau = T - t$, es decir: $\frac{\partial C}{\partial \tau} = \frac{\partial C}{\partial t} \frac{\partial t}{\partial \tau} = -\frac{\partial C}{\partial t}$. Estas derivadas las utilizaremos en la siguiente sección.

1.5.1 La ecuación diferencial parcial de P_1 y P_2

Como el precio de la opción de compra C en la ecuación (1.13) es también una opción, también satisface la EDP en (1.12), que escribimos en seguida en términos de C pero con la derivada respecto al tiempo en τ en vez de respecto a t :

$$\begin{aligned} -\frac{\partial C}{\partial \tau} + \frac{1}{2}v \frac{\partial^2 C}{\partial x^2} + \left(r - \frac{1}{2}v\right) \frac{\partial C}{\partial x} + \rho\sigma v \frac{\partial^2 C}{\partial v \partial x} + \\ \frac{1}{2}\sigma^2 v \frac{\partial^2 C}{\partial v^2} - rC + [k(\theta - v) - \lambda v] \frac{\partial C}{\partial v} = 0 \end{aligned} \quad (1.15)$$

Para obtener la EDP de Heston para P_1 y P_2 , observamos que la EDP en la ecuación (1.15) es independiente de las condiciones de la opción de compra. Por lo tanto, haciendo $K = 0$ y $S = 1$ en (1.13) obtenemos una opción cuyo precio es simplemente P_1 que también seguirá la EDP. De la misma manera, mediante haciendo $S = 0, K = 1$ y $r = 0$ en la ecuación (1.13) obtenemos una opción cuyo precio es $-P_2$, y ya que $-P_2$ sigue la EDP, entonces también la sigue P_2 . Esto implica que a partir de las derivadas en (1.14) podemos reagrupar los términos comunes para P_1 y sustituir en la EDP (1.15) resultando en:

$$\begin{aligned} -\frac{\partial P_1}{\partial \tau} + \frac{1}{2}v \left[P_1 + 2\frac{\partial P_1}{\partial x} + \frac{\partial^2 P_1}{\partial x^2} \right] + \left(r - \frac{1}{2}v\right) \left[P_1 + \frac{\partial P_1}{\partial x} \right] \\ + \rho\sigma v \left[\frac{\partial P_1}{\partial v} + \frac{\partial^2 P_1}{\partial x \partial v} \right] + \frac{1}{2}\sigma^2 v \frac{\partial^2 P_1}{\partial v^2} \\ - r \left[e^x P_1 - e^{-r\tau} K P_2 \right] + [k(\theta - v) - \lambda v] \frac{\partial P_1}{\partial v} = 0. \end{aligned} \quad (1.16)$$

Si $K = 0$ y $S = 1$ de modo que la EDP de P_1 en (1.16) se convierte en:

$$\begin{aligned} -\frac{\partial P_1}{\partial \tau} + \left(r + \frac{1}{2}v\right) \frac{\partial P_1}{\partial x} + \frac{1}{2}v \frac{\partial^2 P_1}{\partial x^2} + \rho\sigma v \frac{\partial^2 P_1}{\partial x \partial v} + \\ [\rho\sigma v + k(\theta - v) - \lambda v] \frac{\partial P_1}{\partial v} + \frac{1}{2}\sigma^2 v \frac{\partial^2 P_1}{\partial x \partial v} = 0 \end{aligned} \quad (1.17)$$

Análogamente, de las derivadas en (1.14), agrupamos los términos comunes de P_2 y sustituimos en la EDP para P_2 en la ecuación (1.15):

$$\begin{aligned} \left[rP_2 - \frac{\partial P_2}{\partial \tau} \right] + \frac{1}{2}v \frac{\partial^2 P_2}{\partial x^2} + \left(r - \frac{1}{2}v\right) \frac{\partial P_2}{\partial x} + \rho\sigma v \frac{\partial^2 P_2}{\partial v \partial x} + \\ \frac{1}{2}\sigma^2 v \frac{\partial^2 P_2}{\partial v^2} - r \left[e^x P_1 - e^{-r\tau} K P_2 \right] + [k(\theta - v) - \lambda v] \frac{\partial P_2}{\partial v} = 0 \end{aligned} \quad (1.18)$$

Si $S = 0$ y $K = -1$ y $r = 0$ tal que la EDP para P_2 en la ecuación (1.18) se transforma en

$$\begin{aligned} -\frac{\partial P_2}{\partial \tau} + \frac{1}{2}v \frac{\partial^2 P_2}{\partial x^2} + \left(r - \frac{1}{2}v\right) \frac{\partial P_2}{\partial x} + \rho\sigma v \frac{\partial^2 P_2}{\partial v \partial x} \\ + \frac{1}{2}\sigma^2 v \frac{\partial^2 P_2}{\partial v^2} + [k(\theta - v) - \lambda v] \frac{\partial P_2}{\partial v} = 0 \end{aligned} \quad (1.19)$$

Por convenirnos en la notación combinamos las ecuaciones (1.17) y (1.19) en una única expresión:

$$\begin{aligned} & -\frac{\partial P_j}{\partial \tau} + \rho \sigma v \frac{\partial^2 P_j}{\partial x \partial v} + \frac{1}{2} v \frac{\partial^2 P_j}{\partial x^2} + \frac{1}{2} v \sigma^2 \frac{\partial^2 P_j}{\partial v^2} \\ & + (r + u_j v) \frac{\partial P_j}{\partial x} + (a - b_j v) \frac{\partial P_j}{\partial v} = 0 \end{aligned} \quad (1.20)$$

para $j = 1, 2$ y con: $u_1 = \frac{1}{2}$, $u_2 = -\frac{1}{2}$, $a = \kappa \theta$, $b_1 = \kappa + \lambda - \rho \sigma$, y $b_2 = \kappa + \lambda$. Esta es la ecuación (12) de Heston (1993) p. 330, pero en términos τ en lugar de t , lo que explica el signo negativo en el primer término de (1.20).

1.5.2 Determinación de las funciones características

Cuando las funciones características $f_1(\phi; x, v)$ y $f_2(\phi; x, v)$ correspondientes a las probabilidades *In-The-Money* son conocidas, cada probabilidad se puede recuperar de su función característica a través del Teorema de inversión³ (véase Waller, *et al* (1995) para una revisión de los métodos de inversión):

$$P_j = \mathbb{P}(\ln S_T > \ln K) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[\frac{e^{-i\phi \ln K} f_j(\phi; x, v)}{i\phi} \right] d\phi. \quad (1.21)$$

Al vencimiento cuando $t = T$ de modo que $\tau = 0$, las probabilidades están sujetas a la condición terminal

$$P_j(x, v, 0) = \mathbf{1}_{x > \ln K},$$

lo cual simplemente significa que al vencimiento, cuando $S_T > K$ la probabilidad de que la opción de compra este ITM es la unidad. Heston supone que las funciones características son de la forma log-lineal:

$$f_j(\phi; x, v) = \exp(C_j(\tau, \phi) + D_j(\tau, \phi)v + i\phi x) \quad (1.22)$$

Donde C_j y D_j son coeficientes y $\tau = T - t$ el plazo al vencimiento. Las funciones características f_j seguirán la EDP en la ecuación (1.20). Esto es una consecuencia del Teorema de Feynman-Kač⁴, que establece que si una función $f(x_t, T)$ de la ecuación diferencial estocástica x_t satisface la EDP $\frac{\partial f}{\partial t} - rf + \mathcal{A}f = 0$ con condición terminal $f(x_T, T)$ donde $x_t = (\ln S_t, v_t)$ es la difusión bivariada de Heston y \mathcal{A} es el generador de Heston de la ecuación (1.11):

$$\begin{aligned} \mathcal{A} &= rS \frac{\partial}{\partial S} + \frac{1}{2} v S^2 \frac{\partial}{\partial S^2} \\ &+ [k(\theta - v) - \lambda(S, v, t)] \frac{\partial}{\partial v} + \frac{1}{2} \sigma^2 v \frac{\partial^2}{\partial v^2} + \rho \sigma v S \frac{\partial}{\partial v \partial S} \end{aligned}$$

entonces la solución de f es:

$$f(x_t, t) = \mathbb{E}[f(x_T, T) | x_t = x, v_t = v].$$

Al utilizar $f(x_T, T) = e^{i\phi x_T} = e^{i\phi \ln S_T}$ produce la solución:

$$f(x_t, t) = \mathbb{E}[e^{i\phi \ln S_T} | \ln S_t = S, v_t = v],$$

³ Una propiedad básica de la función característica es la relación uno a uno con la función de distribución. Cada variable aleatoria posee una función característica única y de hecho la función característica caracteriza la distribución exclusivamente. El Teorema de inversión es el teorema fundamental de la teoría de la función característica porque vincula la función característica con la función de distribución por medio de una Transformada inversa de Fourier.

⁴ Véase por ejemplo el capítulo 20 de Venegas-Martínez (2008) y las referencias respectivas.

que es la función característica para $\ln S_T$, por lo tanto de la ecuación (1.20) la EDP de la función característica f_j es:

$$\begin{aligned} -\frac{\partial f_j}{\partial \tau} + \rho\sigma v \frac{\partial^2 f_j}{\partial x \partial v} + \frac{1}{2}v \frac{\partial^2 f_j}{\partial x^2} + \frac{1}{2}v\sigma^2 \frac{\partial^2 f_j}{\partial v^2} \\ + (r + u_j v) \frac{\partial f_j}{\partial x} + (a - b_j v) \frac{\partial f_j}{\partial v} = 0. \end{aligned} \quad (1.23)$$

Para resolver esta EDP necesitamos las siguientes derivadas:

$$\begin{aligned} \frac{\partial f_j}{\partial \tau} &= - \left(\frac{\partial C_j}{\partial \tau} + \frac{\partial D_j}{\partial \tau} v \right) f_j, \\ \frac{\partial f_j}{\partial x} &= i\phi f_j, \\ \frac{\partial^2 f_j}{\partial x^2} &= -\phi^2 f_j, \\ \frac{\partial f_j}{\partial v} &= D_j f_j, \\ \frac{\partial^2 f_j}{\partial v^2} &= D_j^2 f_j, \\ \frac{\partial^2 f_j}{\partial v \partial x} &= i\phi D_j f_j. \end{aligned}$$

Al sustituir estas derivadas en la ecuación (1.23) y al eliminar los términos f_j obtenemos:

$$\begin{aligned} - \left(\frac{\partial C_j}{\partial \tau} + v \frac{\partial D_j}{\partial \tau} \right) + \rho\sigma v i\phi D_j - \frac{1}{2}v\phi^2 + \frac{1}{2}v\sigma^2 D_j^2 \\ + (r + u_j v) i\phi + (a - b_j v) D_j = 0, \end{aligned} \quad (1.24)$$

o de manera equivalente:

$$\begin{aligned} v \left(-\frac{\partial D_j}{\partial \tau} + \rho\sigma i\phi D_j - \frac{1}{2}\phi^2 + \frac{1}{2}\sigma^2 D_j^2 + u_j i\phi - b_j D_j \right) \\ - \frac{\partial C_j}{\partial \tau} + r i\phi + a D_j = 0. \end{aligned} \quad (1.25)$$

Esto produce dos de ecuaciones diferenciales ordinarias:

$$\frac{\partial D_j}{\partial \tau} = \rho\sigma i\phi D_j - \frac{1}{2}\phi^2 + \frac{1}{2}\sigma^2 D_j^2 + u_j i\phi - b_j D_j \quad (1.26a)$$

$$\frac{\partial C_j}{\partial \tau} = r i\phi + a D_j. \quad (1.26b)$$

Estas son las ecuaciones (A7) en el Apéndice en Heston (1993) p. 341. S. Heston especifica las condiciones iniciales como $D_j(0, \phi) = 0$ y $C_j(0, \phi) = 0$. La ecuación (1.26a) es una ecuación de Riccati⁵ (Boyce and DiPrima (2008), p. 133) en D_j , mientras que la segunda es una ecuación diferencial ordinaria (EDO) en C_j que se puede resolver fácilmente una vez obtenida D_j .

⁵ Las ecuaciones de Riccati llevan el nombre en honor de un noble veneciano Jacopo Francesco Riccati (1676-1754), quién declinó dar cátedras en Italia, Austria y Rusia por continuar sus estudios de matemáticas en casa. Riccati estudió ampliamente estas ecuaciones, sin embargo, fue Euler en 1760 quién resolvió la ecuación.

1.6 Solución a la ecuación de Heston–Riccati

La ecuación de Riccati para $y(t)$ con coeficientes $P(t)$, $Q(t)$ y $R(t)$ de la forma:

$$\frac{dy(t)}{dt} = P(t) + Q(t)y(t) + R(t)y(t)^2, \quad (1.27)$$

se puede resolver al considerar la EDO de segundo orden en $w(t)$:

$$w'' - \left[\frac{P'}{P} + Q \right] w' + PRw = 0 \quad (1.28)$$

que reescribimos como $w'' + bw' + cw = 0$, entonces la solución a la ecuación (1.27) es:

$$y(t) = -\frac{w'(t)}{w(t)} \frac{1}{R(t)}.$$

Para resolver la ecuación (1.28) recurrimos a la ecuación auxiliar $r^2 + br + c = 0$ que tiene dos soluciones dadas por:

$$\alpha = \frac{-b + \sqrt{b^2 - 4ac}}{2},$$

$$\beta = \frac{-b - \sqrt{b^2 - 4ac}}{2}.$$

La solución a la EDO de segundo orden (1.28) es: $w(t) = Ae^{\alpha t} + Be^{\beta t}$, con A y B constantes. Por lo tanto, la solución a la ecuación de Riccati es:

$$y(t) = -\frac{A\alpha e^{\alpha t} + B\beta e^{\beta t}}{Ae^{\alpha t} + Be^{\beta t}} \frac{1}{R(t)}.$$

A partir de las ecuaciones (1.26a–1.26b) la ecuación Heston–Riccati es:

$$\frac{\partial D_j}{\partial \tau} = P_j - Q_j D_j + R D_j^2, \quad (1.29)$$

donde:

$$P_j = u_j i \phi - \frac{1}{2} \phi^2,$$

$$Q_j = b_j - \rho \sigma i \phi,$$

$$R = \frac{1}{2} \sigma^2. \quad (1.30)$$

En este caso la correspondiente EDO es:

$$w'' + Q_j w' + P_j R = 0 \quad (1.31)$$

de modo que: $D_j = -\frac{1}{R} \frac{w'}{w}$. La función auxiliar es: $r'' + Q_j r' + P_j R = 0$ con raíces:

$$\alpha = \frac{-Q_j + \sqrt{Q_j^2 - 4P_j R}}{2} = \frac{-Q_j + d_j}{2}$$

y

$$\beta = \frac{-Q_j - \sqrt{Q_j^2 - 4P_j R}}{2} = \frac{-Q_j - d_j}{2},$$

donde:

$$\begin{aligned}
 d_j &= \alpha - \beta \\
 &= \sqrt{Q_j^2 - 4P_jR} \\
 &= \sqrt{(\rho\sigma i\phi - b_j)^2 - \sigma^2(2u_j i\phi - \phi^2)}.
 \end{aligned} \tag{1.32}$$

Por simplicidad en la notación a veces se omite el subíndice j en algunas variables. De esta manera la solución a la ecuación de Heston–Riccati (1.29) es:

$$D_j = -\frac{1}{R} \frac{w'}{w} = -\frac{1}{R} \left(\frac{A\alpha e^{\alpha t} + B\beta e^{\beta t}}{Ae^{\alpha t} + Be^{\beta t}} \right) = -\frac{1}{R} \left(\frac{K\alpha e^{\alpha t} + \beta e^{\beta t}}{Ke^{\alpha t} + e^{\beta t}} \right), \tag{1.33}$$

donde $K = A/B$. La condición inicial $D_j(0, \phi) = 0$ implica que se hace $t = 0$ en la ecuación (1.33), de manera que el numerador es $K\alpha + \beta = 0$, de donde $K = -\frac{\beta}{\alpha}$ por lo tanto, la solución de D_j se convierte en:

$$\begin{aligned}
 D_j &= -\frac{\beta}{R} \left(\frac{-e^{\alpha\tau} + e^{\beta\tau}}{-g_j e^{\alpha\tau} + e^{\beta\tau}} \right) \\
 &= -\frac{\beta}{R} \left(\frac{1 - e^{d_j\tau}}{1 - g_j e^{d_j\tau}} \right) \\
 &= \frac{Q_j + d_j}{2R} \left(\frac{1 - e^{d_j\tau}}{1 - g_j e^{d_j\tau}} \right),
 \end{aligned} \tag{1.34}$$

con:

$$\begin{aligned}
 g_j &= -K \\
 &= \frac{b_j - \rho\sigma i\phi + d_j}{b_j - \rho\sigma i\phi - d_j}.
 \end{aligned}$$

Podemos escribir la solución de D_j como:

$$D_j = \frac{b_j - \rho\sigma i\phi + d_j}{\sigma^2} \left(\frac{1 - e^{d_j\tau}}{1 - g_j e^{d_j\tau}} \right). \tag{1.35}$$

La solución para D_j se encuentra al integrar la segunda ecuación en (1.26a)–(1.26b), por lo tanto:

$$C_j = \int_0^\tau ri\phi dy + \alpha \left(\frac{Q_j + d_j}{2R} \right) \int_0^\tau \left(\frac{1 - e^{d_j y}}{1 - g_j e^{d_j y}} \right) dy.$$

La primera integral es $ri\phi\tau$ y la segunda se resuelve al sustituir $x = e^{d_j y}$, de donde $dx = d e^{d_j y} dy$ y $y = \frac{1}{d_j} \ln x$, entonces:

$$C_j = ri\phi\tau + \frac{\alpha}{d_j} \left(\frac{Q_j + d_j}{2R} \right) \int_1^{e^{d_j\tau}} \left(\frac{1 - x}{1 - g_j x} \right) \frac{1}{x} dx. \tag{1.36}$$

Al resolver la integral en la ecuación anterior por fracciones parciales:

$$\begin{aligned}
 \int_1^{e^{d_j\tau}} \frac{1 - x}{x(1 - g_j x)} dx &= \int_1^{e^{d_j\tau}} \left[\frac{1}{x} - \frac{1 - g_j}{1 - g_j x} \right] dx \\
 &= \left[\ln x + \frac{1 - g_j}{g_j} \ln(1 - g_j x) \right]_{x=1}^{x=e^{d_j\tau}} \\
 &= \left[d_j\tau + \frac{1 - g_j}{g_j} \ln \left(\frac{1 - g_j e^{d_j\tau}}{1 - g_j} \right) \right].
 \end{aligned} \tag{1.37}$$

Al sustituir el resultado anterior en (1.36), y al sustituir obtenemos la solución para C_j :

$$C_j = ri\phi\tau + \frac{a}{\sigma^2} \left[(b_j - \rho\sigma i\phi + d_j)\tau - 2 \ln \left(\frac{1 - g_j e^{d_j\tau}}{1 - g_j} \right) \right]. \quad (1.38)$$

De esta manera hemos completado el desarrollo original de la EDP del modelo de Heston.

1.7 Resumen

El precio de la opción de compra es:

$$C_T(K) = e^{x_t} P_1(x, v, \tau) - e^{-r\tau} K P_2(x, v, \tau), \quad (1.39)$$

con probabilidades P_1 y P_2 para una opción In-The-Money:

$$P_j = \mathbb{P}(\ln S_T > \ln K) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[\frac{e^{-i\phi \ln K} f_j(\phi; x, v)}{i\phi} \right] d\phi, \quad (1.40)$$

y con las funciones características f_1 y f_2 dadas por:

$$f_j(\phi; x, v) = \exp(C_j(\tau, \phi) + D_j(\tau, \phi)v + i\phi x),$$

donde:

$$\begin{aligned} C_j &= ri\phi\tau + \frac{\kappa\theta}{\sigma^2} \left[(b_j - \rho\sigma i\phi + d_j)\tau - 2 \ln \left(\frac{1 - g_j e^{d_j\tau}}{1 - g_j} \right) \right]; \\ D_j &= \frac{b_j - \rho\sigma i\phi + d_j}{\sigma^2} \left(\frac{1 - e^{d_j\tau}}{1 - g_j e^{d_j\tau}} \right), \\ g_j &= \frac{b_j - \rho\sigma i\phi + d_j}{b_j - \rho\sigma i\phi - d_j}, \\ d_j &= \sqrt{(\rho\sigma i\phi - b_j)^2 - \sigma^2 (2u_j i\phi - \phi^2)}, \end{aligned}$$

es importante señalar que de (1.40) requerimos calcular las integrales para P_1 y P_2 mediante un algoritmo de integración numérica, una vez obtenido el precio de la opción de compra, el precio de la opción de venta la obtenemos mediante la paridad *put-call*:

$$P_T(K) = C_T(K) + K e^{-r\tau} - S_t.$$

A partir de (1.39) podemos determinar las sensibilidades del precio de la opción ante cambios en los diferentes parámetros, lo que se denominan comúnmente como las “griegas”, al respecto las “griegas” de un modelo de difusión con saltos y que incluyen el modelo de Heston como un caso particular se encuentran en el artículo de Bakshi, Cao, y Chen (1997); un análisis comparativo con las del modelo de ByS está en Ortiz Ramírez (2012).

1.8 Conclusiones

En este capítulo discutimos las características teóricas del modelo de Heston (1993). Este modelo extiende el modelo Black y Scholes al suponer que la volatilidad del precio del subyacente es conducida por un proceso de difusión, además de incluir una correlación arbitraria entre la volatilidad y los rendimientos del subyacente. A partir de la función de densidad marginal obtenida por Dragulescu y Yakovenko (2002) y con parámetros dados, graficamos dicha densidad. Observamos que el parámetro de correlación modela el sesgo de la densidad. La correlación positiva proviene

de una alta volatilidad al aumentar el precio del subyacente, lo cual se refleja en la cola derecha de la densidad. Por su parte una correlación negativa esta asociada con un baja volatilidad. Mientras que el parámetro de volatilidad de la varianza modela la curtosis de la densidad, lo cual produce colas pesadas. Estas características impactan en los precios de opciones según sea la relación entre el precio del subyacente y el precio de ejercicio.

Planteamos la EDP de Heston mediante un portafolio compuesto por un subyacente, una opción para cubrir el subyacente y otra opción para cubrir la volatilidad del subyacente, esto con el propósito de expresar la EDP en términos del precio del subyacente. Posteriormente con el logaritmo del precio del subyacente deducimos la EDP de una opción de compra Dentro del dinero (In The Money) bajo estos supuestos. Mediante un resultado del Teorema de inversión recuperamos las probabilidades de las funciones características con el supuesto de que son de una forma log-lineal. Una vez determinadas las funciones características y por el Teorema de Feynman-Kač junto con cambios de variable se resuelve la ecuación diferencial parcial. La fórmula del precio de la opción obtenida es similar a la fórmula de Black y Scholes, pero las probabilidades implican una integral con integrandos complejos, que se puede resolver por algún método de integración numérica.

Por último es importante señalar que Bakshi y Madan (2000) proponen un marco general en el que la función característica de la densidad de precio-estado se utiliza para valuar opciones analíticamente de cualquier transformación arbitraria del precio del subyacente, específicamente el modelo de Heston es un caso particular y obtiene la misma EDP. Gatheral (2000) presenta una derivación alternativa mediante la transformada de Fourier.

1.9 Bibliografía

- Bakshi, G., C. Cao y Z. Chen. (1997). “Empirical Performance of Alternative Option Pricing Models”. *Journal of Finance*, Vol. 52, No. 5, pp. 2003-2049.
- Bakshi, G. y Madan D. (2000). “Spanning and Derivative Security Valuation”. *Journal of Financial Economics*, Vol. 55, pp. 205-238.
- Bates, D. (1996). “Jumps and Stochastic Volatility: Exchange Rate Processes in Deutschemark Options”, *Review of Financial Studies* Vol.9, No. 1, pp. 69108.
- Black, F. y M. Scholes (1973). “The Pricing of Options and Corporate Liabilities”. *The Journal of Political Economy*, Vol. 81, No. 3, pp. 637-654.
- Boyce, William E. y DiPrima, Richard C., (2008). *Elementary Differential Equations and Boundary Value Problems*, John Wiley & Sons, Inc., Hoboken, New Jersey.
- Čížek, Pavel; Härdle, Wolfgang Karl; Weron, Rafal (Eds.) (2011). *Statistical Tools for Finance and Insurance*, Springer-Verlag, Berlin Heidelberg, 2nd. edition.
- Cox, J. C., J. E. Ingersoll, y S. A. Ross (1985). “A Theory of the Term Structure of Interest Rates”. *Econometrica*, Vol. 53, No. 2, pp. 385-407.
- Dragulescu, A. A. y Yakovenko, V. M. (2002). “Probability distribution of returns in the Heston model with stochastic volatility”, *Quantitative Finance*, Vol. 2, pp. 443-453.
- Duffie, D., Pan, J. y Singleton, K. (2000). “Transform analysis and asset pricing for affine jumpdiffusions”. *Econometrica*, Vol. 68, No. 6, pp. 1343-1376.
- Gatheral, Jim (2006). *The volatility surface: a practitioner’s guide*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Harvey, A., Ruiz, E. y Shephard, N.(1994). “Multivariate stochastic variance models”. *Review of Economic Studies*, Vol. 61, No. 2, pp. 247-264.
- Heston, S. I. (1993). “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options”. *Review of Financial Studies*, Vol. 6, No. 2, pp. 327-343.

- Hull, J. y A. White (1987). "The Pricing of Options on Assets with Stochastic Volatility". *Journal of Finance*, Vol. 42, pp. 281-300.
- Hull, J. y A. White (1988). "An Analysis of the Bias in Option Pricing Caused by a Stochastic Volatility". *Advances in Futures and Options Research*, Vol. 3, pp. 29-61.
- Lewis, A. L. (2000). *Option Valuation Under Stochastic Volatility: With Mathematica Code*. Newport Beach, Calif.: Finance Press.
- Merton, C. R. (1973). "Theory of Rational Option Pricing". *Bell Journal of Economics and Management Science*, Vol. 4, No. 1, pp. 141-183.
- Merton, R. C., (1976), "Option Pricing when Underlying Stock Returns are Discontinuous". *Journal of Financial Economics*, Vol. 3, No. 1-2, pp. 125-144.
- Ortiz Ramírez, A., Venegas Martínez, F., López Herrera, F. (2011). "Una nota sobre la sensibilidad de los parámetros del modelo de volatilidad estocástica de Heston". *Quantitativa, Revista de Economía*, julio-diciembre 2011, pp. 81-103.
- Ortiz Ramírez, A., "Griegas del modelo de volatilidad estocástica de Heston", en *Avances recientes en valuación de activos y administración de riesgos*, Volumen III, Capítulo 3, Centros Culturales de México A.C., México 2012. ISBN 978-607-7905-05-9.
- Rouah, D. F., (2011). "Derivation of the Heston Model". Sapient Global Markets.
- Shephard, Neil, (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press, New York.
- Schöbel, R. y Zhu J.W. (1999). "Stochastic volatility with an Ornstein-Uhlenbeck process: an extension", *European Finance Review*, Vol. 3, No. 1, pp. 23-46.
- Venegas Martínez, F. (2008). *Riesgos financieros y económicos, productos derivados y decisiones económicas bajo incertidumbre*. 2a. ed., Cengage Learning México.
- Waller, L. A., Turnbull, B. W. y Hardin, J. M. (1995). "Obtaining Distribution Functions by Numerical Inversion of Characteristic Functions with Applications", *The American Statistician*, Vol. 49, No. 4, pp. 346-350.
- Zhu, Jianwei (2010). *Applications of Fourier Transform to Smile Modeling: Theory and Implementation*, Springer Finance Series, 2nd. edition, Springer-Verlag Berlin Heidelberg.

Capítulo 2

Riesgos estratégicos y adquisiciones internacionales: valuación con opciones simultáneas compuestas del crecimiento global de CEMEX

Alejandra Cabello Rosales

Universidad Nacional Autónoma de México

Christian Bucio Pacheco

Universidad Nacional Autónoma de México

2.1 Introducción

La globalización económica ha abierto nuevos mercados tanto para las empresas de países desarrollados como a aquellas de economías emergentes. De ahí que buscando incrementar su valor, limitado en sus propios mercados, las empresas de estas últimas economías han buscado ampliar sus operaciones y penetrar los mercados internacionales. Sin embargo, estas son decisiones estratégicas cuyos riesgos deben ser valorados cuidadosamente, no solo por la gran competitividad en los mercados mundiales, sino en particular porque una decisión estratégica mal fundamentada puede conllevar a grandes pérdidas e incluso la quiebra corporativa. En este

respecto, una estrategia generalizada de crecimiento en las últimas décadas han sido las fusiones y adquisiciones cuya práctica también se ha extendido a empresas de países en vías de desarrollo. Su éxito, sin embargo no está garantizado. Los fracasos pueden ser evitados con una valuación que permita explotar los riesgos de impacto positivo (*upside risks*) y minimizar los riesgos de impacto negativo (*downside risks*). Esto es, una empresa adversa al riesgo desarrollará pocas o nulas oportunidades de crecimiento mediante fusiones y adquisiciones, en tanto que empresas propensas al riesgo podrían caer en trampas de liquidez en cuanto a sus obligaciones y podrían incluso quebrar. El intermedio óptimo y máxima medida de éxito es por tanto el valor de la empresa que se incrementa como consecuencia del riesgo asumido, como consecuencia de una generación de flujos de efectivo positivos a largo plazo que sobreponen los riesgos de las adquisiciones llevadas a cabo. Esto es, las buenas decisiones estratégicas generan riesgos con altos impactos positivos y riesgos con limitados impactos negativos.

Tomando en cuenta esta visión sobre la asunción de riesgos, la valuación de empresas, la valuación de proyectos de inversión, la valuación de fusiones y adquisiciones empresariales es un problema de opciones reales; concretamente la toma de riesgos por parte de una empresa puede verse como el caso de una opción de venta. Así, una inadecuada valuación de las estrategias de una empresa se refleja inmediatamente en su valor de mercado. Este es el caso de CEMEX, empresa mexicana que en las últimas décadas siguió agresivas estrategias de adquisiciones que al presente han colocado el valor de sus acciones a niveles extraordinariamente bajos, en relación a los picos obtenidos de más de \$37 pesos mexicanos, en enero del 2007, a precios por acción de menos de \$4.00 pesos hacia junio del 2012.

Tomando en cuenta las consideraciones anteriores, este trabajo tiene por objetivo aplicar el análisis de opciones reales para determinar la viabilidad de la adquisición de la empresa inglesa RMC y la australiana Rinker que representan dos corporaciones globales desarrolladas y que fueron adquiridas por Cemex en 2005 y 2007.

Cemex realizó durante las últimas dos décadas una agresiva estrategia de crecimiento basada en adquisiciones internacionales de la competencia; sus adquisiciones las realizó con la contratación de un alto endeudamiento solicitado en los mercados internacionales. Actualmente Cemex tiene severos problemas de liquidez para hacer frente a los compromisos adquiridos por

su alto endeudamiento. Es un hecho que debido a sus dos últimas grandes adquisiciones Cemex rompió totalmente con su otrora tendencia para hacer buen frente a sus compromisos adquiridos de su deuda y control de sus flujos de efectivo. Así, analizar la viabilidad de dichas adquisiciones es el motivo central de este trabajo empleando el enfoque de opciones reales como una herramienta de análisis financiero complementaria a la tradicional de valor presente neto, a fin de ejemplificar y procurar contestar si métodos de análisis más sofisticados hubieran coadyuvado a anticipar que no se comprara RMC ni Rinker. Igualmente en este trabajo se subraya la necesidad de presentar el resultado del desempeño de una gran empresa mexicana para mostrar el camino que otras empresas grandes mexicanas deben evitar o emprender pero como resultado de una decisión derivada de estudios más sofisticados, al atreverse a ser cada vez más globales.

La presente investigación está estructurada en tres partes. La primera reseña como de una posición financiera pequeña, Cemex se movió rápidamente a una situación novedosa de alto crecimiento, subrayando las estrategias de mediados de la última década con sus dos adquisiciones más importantes: RMC y Rinker. La segunda parte muestra con un análisis financiero general de tendencias varios aspectos de deterioro financiero de la empresa; destaca la pérdida de su posición cómoda que mantuvo hasta antes de sus compras más importantes. Estas grandes compras fueron acompañadas por una visión futura muy optimista y (seguramente) carentes de análisis más profundos que incorporen otros escenarios económicos lo que (aparentemente) puede considerarse como causantes de colocar a la empresa en una posición de *downside risk*. La tercera parte toma como punto de partida, el qué hubiera pasado si se recurre al enfoque de opciones reales, como una herramienta más flexible a la tradicional de valor presente para el análisis de una estrategia de adquisiciones de Cemex, por lo que se presentan los resultados de este enfoque para el caso de opciones compuestas secuenciales utilizando los datos financieros de Cemex históricos, y de flujos simulados con Monte Carlo, ofreciendo además en esta sección los elementos centrales metodológicos de su cálculo. El trabajo finalmente concluye resumiendo los resultados y la importancia de la metodología de opciones reales para una toma de decisiones estratégicas correctas.

Cabe recalcar que la técnica de opciones reales (OR) no se propone para desplazar técnicas de evaluación que descuentan los flujos de efectivo porque de hecho los complementa; el enfoque de OR aporta el cálculo de valores adicionales en la evaluación de proyectos que en el enfoque

tradicional pasan desapercibidos (Bailey *et al* 2012). Igualmente, de acuerdo con Trigeorgis (1999) este enfoque permite flexibilidad para la adaptación y revisión de decisiones futuras en respuesta a las cambiantes circunstancias; en este caso es como concientizar al tomador de decisión que el futuro está lleno de alternativas y opciones mismas que son objeto de agregación de valor.

2.2 Crecimiento y presencia global

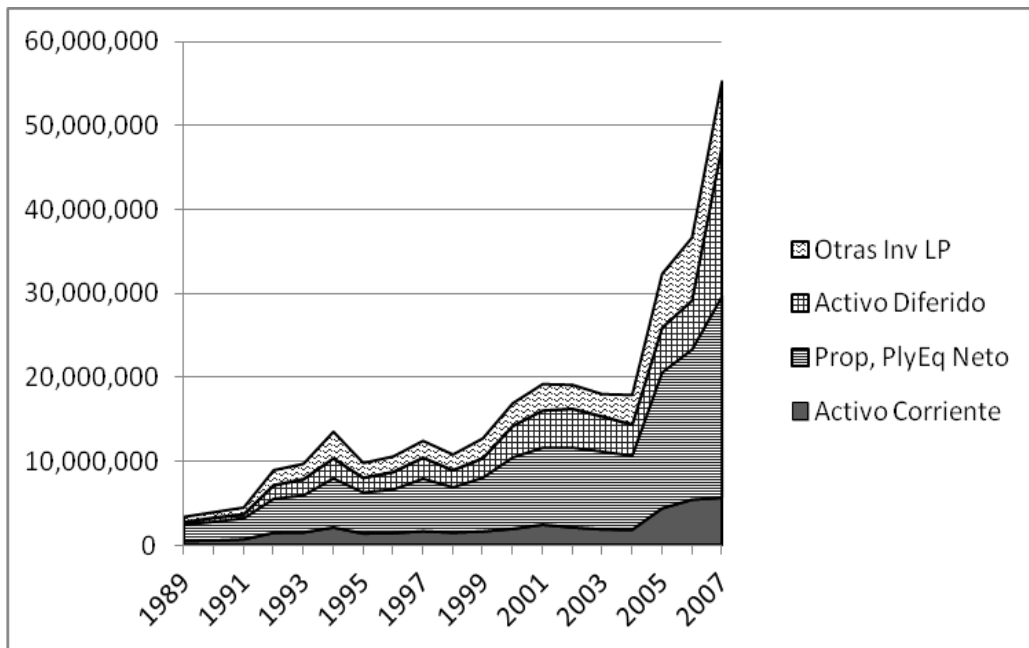
Cemex tiene más de 100 años de existencia; inició sus operaciones en el año 1906.¹ Su periodo de expansión comenzó en las décadas de los 1970's y 1980's, y durante las dos últimas décadas su estrategia de expansión se enfocó a la compra de compañías grandes e internacionales de su sector. Cemex evolucionó de una compañía pequeña y local a ser una compañía global, esto gracias al liderazgo de Don Lorenzo Zambrano, su principal propietario y ejecutivo. Cemex llegó a ubicarse como la tercera compañía cementera en el mundo, después de la francesa Lafarge y la suiza Holcim.

A principios de 2007, y debido a sus grandes adquisiciones Cemex logró cantidades anuales de ventas espectaculares de más de US \$15 miles de millones y logró una presencia en más de 50 países en todos los continentes. A inicios de dicho año ya se había ubicado como una de las más importantes empresas cementeras a nivel mundial.

La Figura 2.1 muestra su crecimiento en activos totales para el período 1988-2007, los cuales crecieron de un nivel de US \$1,710.2 millones de dólares de fines de diciembre de 1988 a US\$11,414.2 millones a fines de 1994, año de inicio de la crisis del peso mexicano; posteriormente estos activos crecieron a la impresionante cantidad de US\$58,735.6 millones en el tercer trimestre del 2007.

¹ Para una revisión resumida de la historia de Cemex y su crecimiento, ver Cabello (2009).

Figura 2.1 Activos Totales de Cemex (Miles de US dólares)



La presencia global de Cemex realmente se observa con el impresionante crecimiento de sus activos totales, pero igualmente con su presencia entre los países más desarrollados. Entre los países de Europa (incluidos los países de la Europa del este) destaca su presencia en Inglaterra, España, Austria, Croacia, República Checa, Finlandia, Francia, Alemania, Hungría, y Latvia, Noruega, Polonia y Suecia. En América: Estados Unidos, Argentina, Colombia, Costa Rica, República Dominicana, Nicaragua, Panamá, Puerto Rico, y Venezuela, así como también tiene varias operaciones en la región caribeña. Del medio oriente destacan: Egipto, Israel, y los Emiratos Árabes Unidos. En Asia destacan: Bangladesh, China, Malasia, las Filipinas, Taiwan y Tailandia. Realmente, Cemex se estaba consolidando con una gran presencia internacional. Sin embargo, sus propietarios deseaban una mayor penetración en Estados Unidos por medio de importantes adquisiciones con mercado en este país americano.

2.2.1 Adquisiciones internacionales más importantes

Las dos grandes adquisiciones de RMC y Rinker ocurrieron desde mediados de la última década. El 1 de marzo del 2005, se anunció que Cemex había concluido la adquisición de RMC Group,

de Inglaterra, que era uno de los más grandes productores de cemento y agregados de Europa, así como el mayor abastecedor del mundo de cemento premezclado (*ready mix cement*). El valor de la empresa fue de US\$4.1 miles de millones.² Hasta ese momento esta correspondió a la mayor compra que Cemex había realizado en toda su historia, pero vendría otra aún más importante.

En julio del 2007 Cemex adquirió a la australiana Rinker. Cemex adquirió el 95.62% de las acciones de Rinker Group Limited al precio de US\$ 15.85 por acción. El valor total de la transacción fue de aproximadamente US\$15.1 miles de millones.³ Esta adquisición correspondió a la compra más importante jamás realizada por Cemex, cumpliendo con su estrategia de penetrar al mercado americano con más fuerza. Cabe destacar que las expectativas de la nueva combinación eran realmente optimistas ya que se esperaba crear a la compañía no sólo más grande, sino también como la más redituable del sector de materiales de construcción en todo el mundo. Debido a esta adquisición se esperaron ventas por más de US\$21.67 miles de millones y con más de 67,000 empleados en más de 50 países alrededor del mundo.⁴

Si bien Cemex nunca había experimentado otra cosa que no hubieran sido éxitos, es igualmente cierto que nunca había tenido adquisiciones de otras empresas con montos tan importantes. Idealmente, desde un análisis *ex-post*, cabe reiterar la necesidad de incorporar en los análisis correspondientes diferentes escenarios futuros y del análisis de opciones reales por tratarse de decisiones estratégicas de alto riesgo. Este argumento sugerido anteriormente se afirma por el hecho de que la adquisición de Rinker sucedió en un lapso poco mayor a dos años después de haber adquirido a RMC, de hecho escaso para consolidar la primera gran compra que en sí misma ya resultaba arriesgada. A pesar de que para entonces ya se manifestaban claros síntomas de la crisis *subprime* seguramente muchos tomadores de decisiones empresariales no reconocieron dicho problema debido a la euforia financiera que predominaba en los mercados reales y financieros internacionales de los países desarrollados. La excesiva ambición conllevó a la crisis subprime y sus secuelas de 2008 y 2009, afectando severamente al sector de la construcción. Con la mundialización de esta crisis e incluyendo su manifestación en las economías emergentes las perspectivas económicas de Cemex se desplomaron.

² Cemex, Informe Anual, 2005.

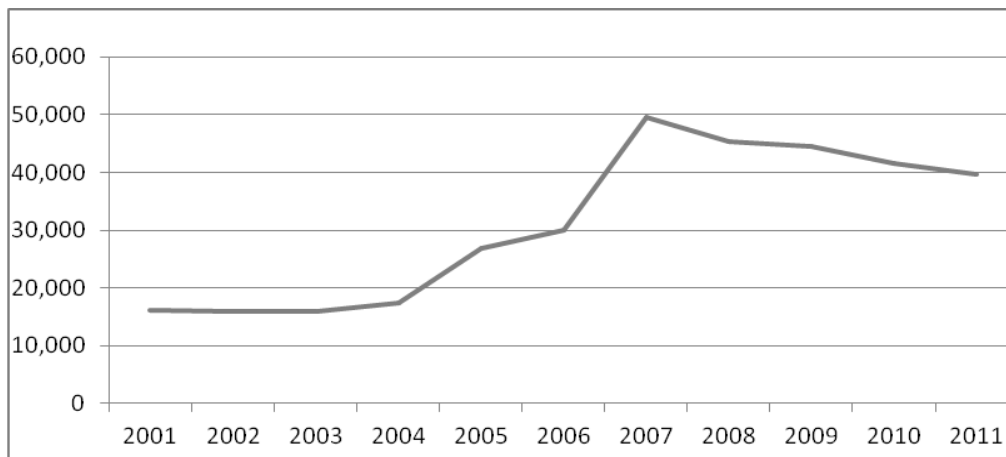
³ Cemex, Informe Anual, 2007.

⁴ Ibid.

2.3 Deterioro financiero de CEMEX

Como se mencionó anteriormente, después de las compras más importantes de toda su historia los ejecutivos de Cemex esperaban obtener un aumento importante no sólo del crecimiento de la empresa medido por el tamaño de sus activos totales, sino también de su rentabilidad; sin embargo, lo que realmente sucedió fue otra historia. Respecto al crecimiento de sus activos se puede apreciar en la Figura 2.2 que después de las compras de RMC y Rinker, los activos totales ascendieron significativamente alcanzando un pico de alrededor de US\$50,000 millones, descendiendo después y continuamente hasta registrar US\$40,000 millones a fines del 2011. Este declive tuvo lugar debido a la venta de algunos activos a fin de pagar las deudas asociadas con sus adquisiciones.

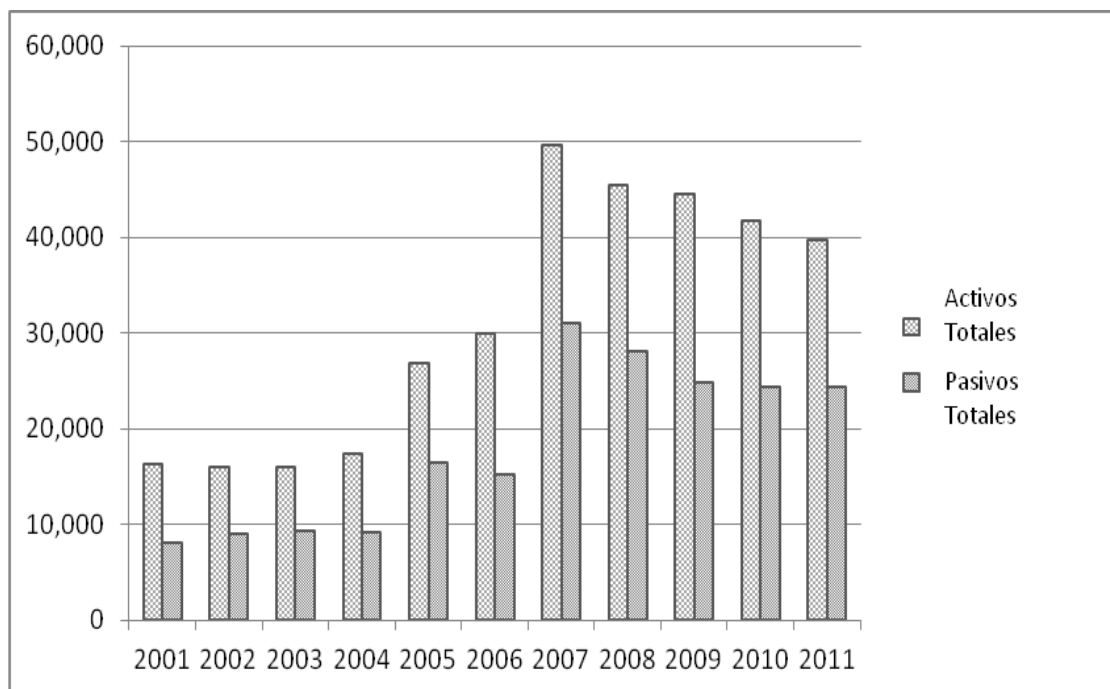
Figura 2.2. Activos Totales de Cemex (Millones de US dólares)



La Figura 2.2 confirma esta situación; se puede observar el tamaño de los activos totales y su comportamiento a través de los años de operación de Cemex para el periodo 2001-2011. También se observa en la Figura 3 el comportamiento de su endeudamiento total. Aunque en los últimos años ha tendido a decrecer, se puede observar un incremento espectacular de la deuda total de Cemex en 2007, esto es, el endeudamiento crece de una manera más que proporcional en relación al crecimiento de los activos totales precisamente después de sus compras de RMC y Rinker. Cabe destacar nuevamente que si bien el tamaño de la deuda total es menor de 2009 a

2011, en relación al 2007 y 2008, el declive de sus activos totales claramente indica que la proporción de la deuda tendió a crecer hacia finales del año 2011.

Figura 2.3 Activos Totales y Deuda Total de Cemex

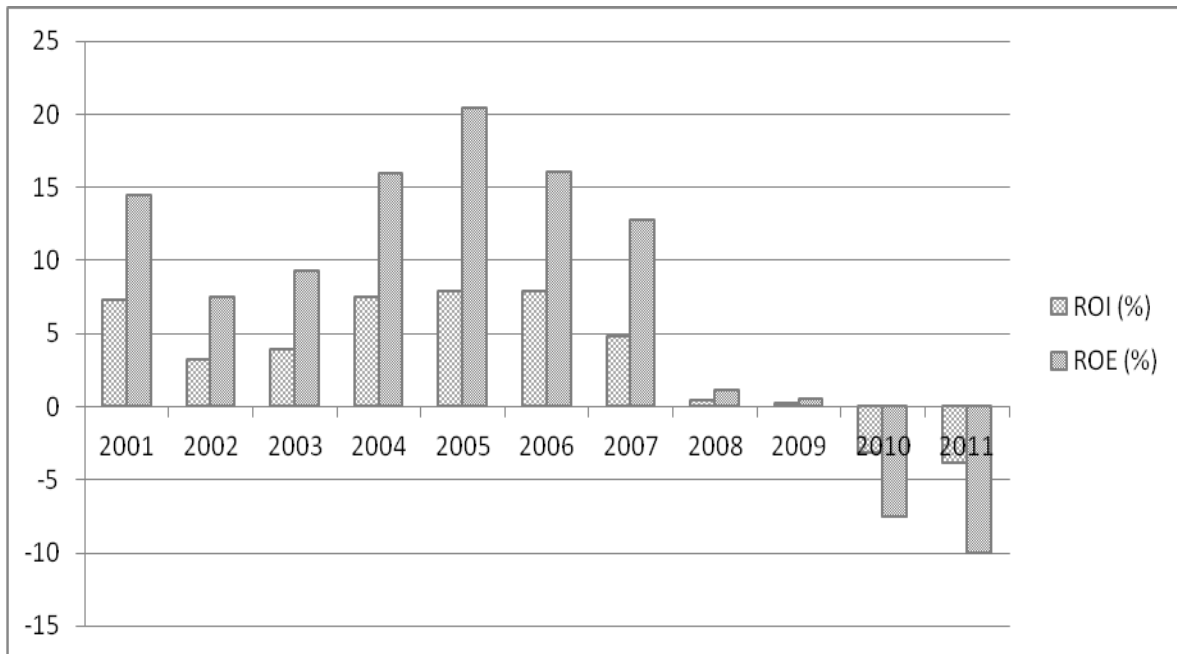


La Figura 2.4 muestra los resultados de Cemex en relación al Rendimiento a su Inversión, y al Rendimiento al Capital Contable, ROI y ROE (por sus siglas en inglés); se observa un comportamiento muy irregular tanto en los años de antes de las compras y durante los años en que estas ocurrieron, pero posteriormente a los años de las compras la tendencia muestra un grave deterioro de ambos componentes del ROI y ROE, resultando ambos negativos para el año 2010 y 2011. Un hecho nunca antes manifestado en toda la vida de Cemex.

Los resultados negativos del Rendimiento a la Inversión (ROI) indican además que tanto el rendimiento como la productividad de Cemex se deterioraron, y el irregular comportamiento del ROE indica la misma situación pero además indica que a pesar del alto endeudamiento los rendimientos al capital accionario de Cemex no lograron mostrarse positivos. Es decir, el componente de la deuda incluido en el cálculo del ROE, que en otros años fue un multiplicador

del rendimiento al capital contable, en 2010 y 2011, su aportación se revirtió; el rendimiento al capital de Cemex se volvió negativo debido al deterioro tanto de la rentabilidad como de la productividad de Cemex inobjetablemente derivadas de las adquisiciones de RMC y Rinker.

Figura 2.4 Rendimiento a la Inversión (ROI, %) y Rendimiento al Capital Contable (ROE, %) de Cemex

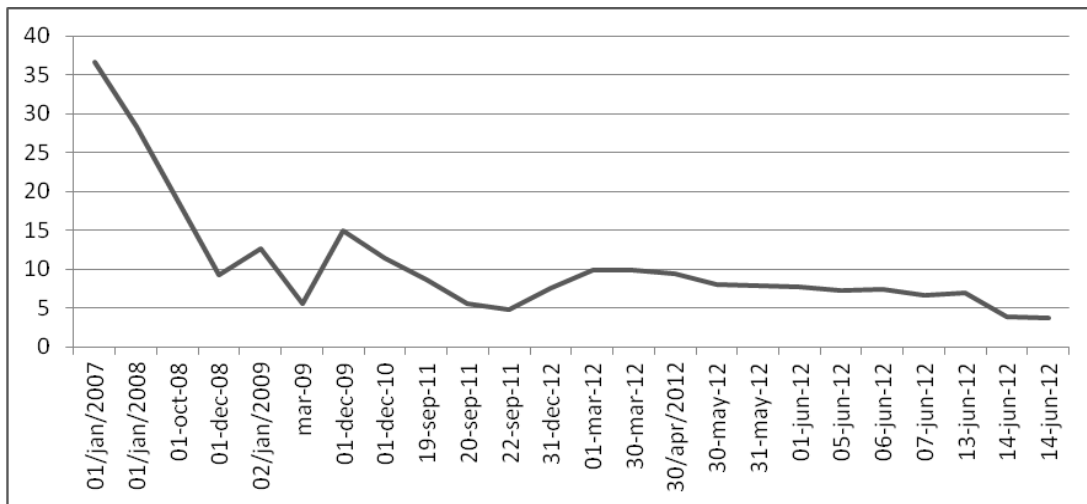


La empresa que apenas no más de media década atrás demostraba un excelente desempeño comienza realmente a mostrar verdaderos y graves deterioros con graves consecuencias en el precio de su acción.

Por otro lado la Figura 2.5 muestra como de un precio por acción de más de \$37 pesos mexicanos, en enero del 2007, se abarató a menos de \$4.00 pesos hacia junio del 2012; un hecho muy lamentable.

Una importante lección del deterioro de Cemex para otras empresas multinacionales es que deben manejar con mucho cuidado su deuda; esta conclusión es particularmente válida para el caso de grandes empresas nacionales tradicionales en cuanto a su limitado gobierno corporativo.

Figura 2.5 Comportamiento del Precio de la Acción de Cemex en el Mercado de Valores de México



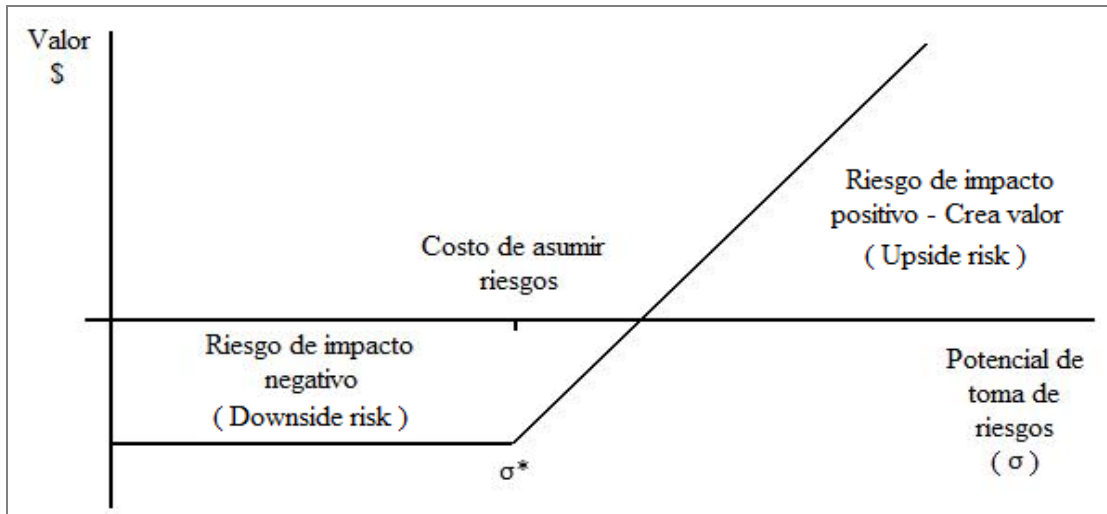
Fuente:<http://mx.finance.yahoo.com/q/hp?s=CEMEXCPO.MX&a=00&b=1&c=2003&d=05&e=14&f=2012g=m&z=66&y=66>

2.4 El riesgo ante decisiones de inversión

El riesgo es un hecho inmerso en cualquier decisión financiera. En teoría a mayor riesgo se debe esperar la obtención de un mayor rendimiento. Igualmente, una buena administración del riesgo implica, como se señaló anteriormente, asumir riesgos de impacto positivo que sobrepongan los riesgos de impacto negativo. Sin embargo, el apego aun existente a las metodologías tradicionales de valuación de proyectos ha impedido que el riesgo y potencial flexibilidad que caracteriza a todo proyecto de inversión no sean considerados adecuadamente. Con la metodología de opciones reales se sobrepone estas limitaciones. Concretamente la toma de riesgos por parte de una empresa puede verse como el caso de una opción de venta, como se resume en la Figura 2.6. En el caso de adquisiciones, el precio de ejercicio, en términos de riesgo, esta dado por el nivel de riesgo aceptable para la empresa adquiriente, esto es, la desviación estándar desde la cual se generan ganancias. Las pérdidas limitadas anteriores a este nivel de

riesgo están dadas por riesgos de impacto negativo cuyo efecto es sobrepuesto gracias a los impactos de riesgo positivo.

Figura 2.6 Riesgo y Valuación con Opciones Reales



En términos estadísticos, la media de los flujos esperados sirve de ancla para medir su dispersión:

$$VE = \sum_{i=1}^n x_i p_i \quad (2.1)$$

Donde,

VE= valor esperado

x_i = variable aleatoria

p_i = probabilidad de ocurrencia de cada variable aleatoria

La variabilidad se mide con el cálculo de la desviación estándar del valor esperado:

$$\sigma = \sqrt{\sum (D - VE)^2 * p} \quad (2.2)$$

D= cada valor esperado

Conforme la desviación estándar es más alta, el riesgo del proyecto o de la inversión es más alto.⁵ El valor resultante de la variabilidad se debe confrontar contra el rendimiento esperado del proyecto y ubicarlo en una línea o frontera que indique si la decisión está siendo maximizada en función de la meta de la empresa. Una mala decisión financiera conllevaría un alto riesgo y por tanto los impactos negativos (downside risk) tenderían a prevalecer. Al contrario una buena decisión financiera conllevaría a la generación de flujos de impacto positivo (upside risk).

Con la modelación de opciones reales, la desviación estándar de los flujos de efectivo puede incorporarse tanto en su aplicación de la fórmula de Black y Scholes, como en el método binomial, de mayor efectividad para la valuación de proyectos. Además, cabe recalcar que el riesgo medido con la desviación estándar puede también ser aplicado en Opciones Reales con métodos más sofisticados como el de simulación Monte Carlo (MC); esto es especialmente útil para proyectos de alto riesgo. El método del árbol binomial permite construir una *lattice* o reticulado (ramal) del comportamiento de los beneficios esperados del proyecto, a través del tiempo, a partir de los cuales y con el método backward o hacia atrás permite construir otra *lattice* del valor de la opción con probabilidades que son calculadas derivadas del método binomial aplicado al subyacente. Esto permite construir un comportamiento dinámico del valor subyacente que es la principal variable de entrada, lo que no hace el valor presente neto, metodología esencialmente estática.

Más allá de las variables de entrada como la volatilidad, los teóricos de las opciones reales también proponen que además del método binomial se empleen métodos de simulación. Pueden ser utilizados tanto para simular valores residuales como para obtener cada uno de los renglones específicos por ejemplo para simular a futuro tanto los renglones de ventas como de los costos y gastos del estado de resultado de varios años.

El método MC es un modelo estocástico, implicando que incluye una parte determinista y una aleatoria. En una decisión financiera es muy empleado el método con movimiento browniano geométrico:

$$dS_t = \mu_t S_t dt + \sigma_t S_t dz_t \quad (2.3)$$

⁵ Una medida más exacta del riesgo es ajustando la desviación estándar a un coeficiente de variación para incorporar el tamaño de los flujos esperados.

Esta ecuación señala cómo calcular el precio del activo subyacente, donde μ y σ son la media y la desviación estándar de los resultados esperados y dz_t es una variable que se distribuye de manera normal con media igual a cero y varianza dt .⁶ Con este modelo se está implicando un cálculo de un valor central a partir de un modelo que plantea un comportamiento determinista con agregación de un valor aleatorio de los residuales. A diferencia de una situación de resultados esperados de manera determinista que podría implicar sólo tres tiradas, o tres resultados esperados, el método Monte Carlo realiza muchas tiradas como pueden ser más de 10,000 para obtener el valor central, que es su media, y la variabilidad que es la desviación estándar.

Concretamente en este trabajo se utilizan tanto el método binomial para obtener la *lattice* del valor subyacente que en este caso corresponde al flujo de efectivo, como el método monte carlo para simular en un escenario los flujos esperados después de la compras de RMC y Rinker; esto permite incorporar explícitamente al riesgo en la decisión de realizar las compras de RMC o Rinker.

2.4.1 Enfoque de opciones reales para el análisis de las adquisiciones de rmc y rinker por parte de CEMEX

El enfoque de opciones reales es una metodología que permite al administrador de una corporación obtener importante información, más allá de la que aporta el enfoque tradicional de valor presente neto, hecho importante en relación a la valuación de una estrategia de adquisiciones. Dentro de este enfoque se encuentra el de una opción secuencial compuesta. Esta es simplemente una opción sobre otra opción, el cual puede ser una técnica muy útil para estudiar una estrategia para adquirir dos importantes compañías en dos fases, donde la compra más tardía depende del éxito de la compra previa o más temprana. Para los propósitos que siguen en esta sección, la primera adquisición ocurre en el año 1, la segunda adquisición ocurre dos años después, como realmente Cemex lo hizo con RMC y Rinker. Los resultados calculados ofrecen un importante ejemplo metodológico y muy ilustrativo para la toma de decisiones estratégicas de expansión de una empresa; dicha metodología incorpora flexibilidad y claridad para facilitar la

⁶ Para un mayor detalle de la explicación de las variables de este método y de otros de simulación ver: Bucio Pacheco (2012).

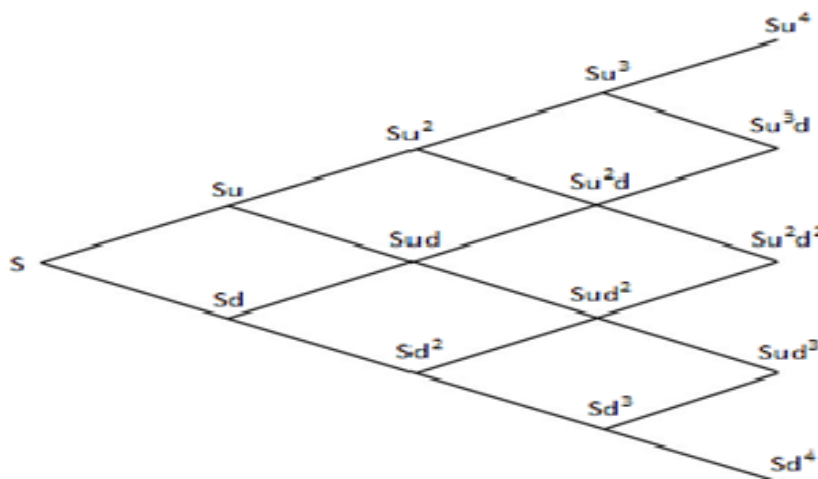
toma de decisiones de forma anticipada y analizar con más conciencia la conveniencia o no en las adquisiciones; más aun tratándose de adquisiciones con valores sumamente cuantiosos.⁷

2.4.2 Cálculo de la Opción Secuencial Compuesta

La metodología del cálculo de opciones secuenciales compuestas involucra tres frases que se explican en forma general a continuación:

Fase I.- En esta fase se calcula la *lattice* (árbol) del activo subyacente inicial (los flujos de efectivo esperados del proyecto traídos o descontados a valor presente), ver Figura 2.7.⁸ Esta *lattice* calcula la evolución del valor presente de los flujos de efectivo futuros.⁹

Figura 2.7 Lattice del Activo Subyacente



Aquí, S=activo subyacente, σ =volatilidad, X1= primera fase (un año de expiración), X2=segunda fase de expiración (3 años).

⁷ Queda fuera del alcance de este trabajo el ejercicio matemático de este tipo de opción compuesta ya que existe abundante literatura al respecto. Sin embargo para el lector interesado se recomienda: Kwok, Y. K. en: <http://www.math.ust.hk/~maykwok/publications/Articles/comp%20option.pdf> (consultado en julio del 2012); y Geske, Robert (1979): 63-81.

⁸ La *lattice* se construye siguiendo el método binomial que se usa para valorar una opción. Este método fue desarrollado por Cox, Ross y Rubinstein. Para un buen resumen de este método ver: Vedovoto y Prior (2010): 18-21.

⁹ En Bailey, et. al. (2012) a la palabra *lattice* se le denomina reticulado, en este trabajo se emplea la palabra *lattice*. En dicha fuente se realiza una excelente ejemplificación y explicación de los conceptos fundamentales de opciones reales y sus diferencias con los conceptos de opciones financieras, así como de las probabilidades utilizadas por el método binomial, varias de las cuales son citadas en el presente trabajo.

La *lattice* corresponde a la construcción de un árbol binomial que muestra el valor de los flujos de efectivo esperado en cada nodo o rama, y para cada año. El ejemplo de la Figura 7 corresponde a 4 años.

El Cálculo de cada nodo corresponde a calcular las fórmulas de la binomial, tanto para valores al alza (u) como los valores que bajan (d):

$$u = e^{\sigma\sqrt{t}} \quad \text{y} \quad d = e^{-\sigma\sqrt{t}} = \frac{1}{u} \quad (2.4)$$

Aquí, el alza (u) y la baja (d) tienen una probabilidad determinada: p, para el alza y (1-p) para la baja.

Fase II.- En esta fase se requiere calcular la opción más tardía (*lattice* de la segunda opción, Figura 2.8), también se puede decir que es la segunda vuelta.

Paso 1: Cálculo de la Lattice de la Opción Segunda o Tardía

Los cálculos de la segunda opción representan el máximo entre Ejecutar (valor presente neto), o Mantener la Opción Abierta (calculado con el método backward o reversible a partir del segundo árbol del método binomial—excepto en el tiempo del vencimiento).

El método backward o reversible utiliza la siguiente fórmula, que es básicamente el empleo de una probabilidad neutral al riesgo:

$$p = \frac{e^{rf(\delta)} - d}{u - d} \quad (2.5)$$

El cálculo de la segunda opción referente al máximo entre ejecutar y mantener la opción abierta se plantea de la siguiente manera:

$$\text{Máximo}[Ejecutar, \text{Mantener la Opción Abierta}]$$

ó

$$\text{Max}[X, \text{Resultado Backward}] \quad (2.6)$$

excepto al vencimiento,

$$\text{Máximo [Ejecutar, 0]}. \quad (2.7)$$

Se aclara que el resultado de ejecutar se realiza utilizando el costo de la segunda opción o tardía, X2 como se indica a continuación:

$$\text{Ejecutar: } X = [\text{Evolución del Subyacente} - X2] \quad (2.8)$$

X2: Costo de adquisición de Rinker

Mantener la Opción Abierta = resultado del método backward el cuál se calcula tomando en cuenta las probabilidades neutrales al riesgo (se calcula como el segundo árbol con cálculo reversible a partir del primer árbol de la lattice del comportamiento del subyacente, y que corresponde a la lattice de la Fase I de esta metodología de opciones secuenciales compuesta).

$$\text{Mantener la Opción Abierta} = [p(\text{nodo}_{\text{alza}}) + (1-p)(\text{nodo}_{\text{baja}})] \exp(-rf \, dt) \quad (2.9)$$

Paso 2: Cálculo de la Lattice de la Opción Primera o Temprana

Se refiere al cálculo de la opción más temprana (*lattice* de la primera opción, Figura 2.8, que corresponde a la primera vuelta.

Los resultados de la lattice de la Opción Segunda alimentan las estimaciones de la Primera Opción, donde para ejecutar se utiliza el costo de la primera opción, que es la opción más temprana, X1, y que permite completar la decisión de invertir o no invertir de cada punto de decisión.

Una aclaración importante: la opción 1 tiene un nuevo vencimiento diferente al de la opción 2.

Por tanto la decisión de invertir o no para la opción 1 tiene en el año 1 su vencimiento. De esta manera también las probabilidades para el método backward o reversible son diferentes ya que se toma un nuevo punto de partida (en este caso un año) para calcular de manera reversible el punto inicial.

Se plantea así el cálculo de la primera opción:

Máximo [Ejecutar, Mantener la Opción Abierta]

ó

$$\text{Max}[X', \text{Resultado Backward}], \quad (2.10)$$

excepto al vencimiento,

$$\text{Máximo [Ejecutar, 0]}. \quad (2.11)$$

Se aclara que el resultado de ejecutar se realiza utilizando el costo de la primera opción o temprana, XI como se indica a continuación:

$$\text{Ejecutar: } X' = [\text{Resultado Máximo de Segunda Opción} - XI] \quad (2.12)$$

XI : Costo de adquisición de RMC

Mantener la Opción Abierta = resultado del método backward.

$$\text{Mantener la Opción Abierta} = \left[p(\text{nodo}_{\text{alza}}) + (1-p)(\text{nodo}_{\text{baja}}) \right] \exp(-rf dt). \quad (2.13)$$

En la segunda vuelta y primera vuelta al vencimiento, la decisión de invertir es como sigue:

Invertir = máximo entre ejecutar la opción – el costo del Año 1 > 0

No invertir = máximo entre ejecutar – el costo del Año 1 < 0

Los nodos intermedios en el segundo round deciden:

Dejar la opción de inversión abierta para el futuro si :

Ejecutar < resultado backward por tanto no invertir ahora y esperar

Invertir si: ejecutar > resultado backward o abierto.

Fase III. Presentación Combinada del Cálculo

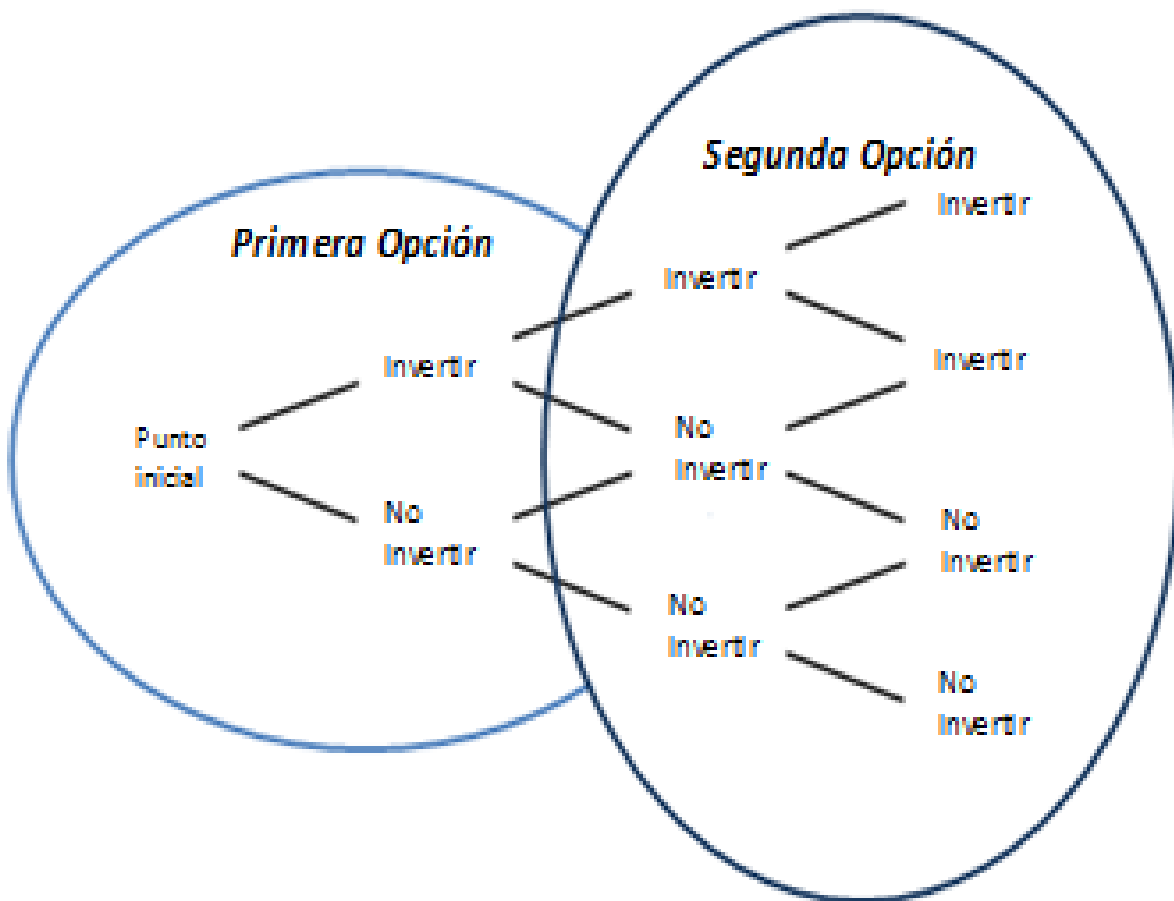
Los cálculos que se hacen se refieren a la valuación de la *lattice* de la opción combinada y resultan puntos de decisión de cuando invertir en la primera vuelta y segunda vuelta, o mantener

la opción de invertir (i.e. mantener la opción abierta de la adquisición de dos grandes compañías). La Figura 2.8 ejemplifica una posible situación de inversión o no que pudiera haber resultado.

En el caso de la Figura 2.8 la decisión resultante indica que se debe invertir en la opción 1 sólo si las condiciones económicas son favorables, pero no invertir en la segunda opción en ningún caso.

El punto inicial de esta *lattice* corresponde al cálculo del valor de la estrategia combinada de la opción de adquisiciones. Representa el punto más importante para conocer el valor que se le debe añadir al valor resultante con el método tradicional de valor presente.

Figura 2.8 Lattice de la Opción Combinada



La metodología debe resultar de fácil comprensión para el administrador o tomador de la decisión de adquisiciones. El uso gráfico de las *lattices* desarrollado con la ayuda del método binomial tiene este propósito.

2.4.3 Cálculo del valor de la opción secuencial compuesta de la estrategia de adquisiciones de CEMEX para el caso de RMC y Rinker

Siguiendo la metodología de las tres fases presentadas en la sección anterior para el cálculo del valor de la Opción Secuencial Compuesta, a continuación se presentan los resultados del cálculo del valor de la estrategia de opción de las adquisiciones de RMC y de Rinker por parte de Cemex.¹⁰

Las variables de entrada corresponden al caso de sinergias que significa utilizar los datos de los flujos de efectivo de Cemex de antes de la compra más los datos de los flujos de efectivo de RMC. Un estudio previo provee los datos iniciales que se usan como inputs (Cabello & Ortiz, 2011), y que son los siguientes.

Los datos de las variables de entrada que resultan para este caso son:

Tiempo subyacente: 3 años ←corresponde al año en que se realiza la segunda compra

Tiempo de la Opción: 1 año ←corresponde al año en que se realiza la primera compra

Volatilidad: 122.31% ←---corresponde al periodo trimestral 1988.4-2007.1

Tasa Libre de Riesgo: 3.85%

En billones de dólares:

X1: Costo de la Opción –Primera Opción--(compra temprana, RMC): US\$4.3

X2: Costo Subyacente –Segunda Opción--(compra tardía, Rinker): US\$15.30

¹⁰ Cálculos realizados en Excel, y con el paquete Real Option Toolkit Software.

Valores para el cálculo de la binomial y método backward o reversible:

$$u = 3.3977 \quad p = 0.2400$$

$$d = 0.2943 \quad 1-p = 0.7600$$

El cálculo de las tres fases resultó como sigue:

Fase I. Cálculo de la Lattice de la Evolución del Subyacente con el método binomial

En el siguiente cuadro se puede apreciar el cálculo de la evolución del subyacente (primer árbol).

Situación Real: Flujos de Efectivo y Valor Presente Reales para volatilidad trimestral de 122.31%, y valor presente de \$8.3 miles de millones de USD									
Subyacente:	8.3	r	4.05%	X1	4.3	-11.3	11.3		
		vol	122.31%	X2	15.3				
u	3.3977	T1	1						
d	0.2943	T2	3						
p	0.2407	Lattice	3	3					
1-p	0.7593	dt	1.00						
		FASE 1				FASE 2			
		Lattice del Subyacente				Opción 1: compra tardía o de Rinker		Opción 2: compra temprana o de RMC	
			325.56			310.26			
		95.82				81.13			
3.40	28.20	28.20	28.20		20.93	12.90		16.63	
S	8.30	8.30		5.34	2.98		3.84		
0.29	2.44	2.44			0.69	0.00		0.00	
		0.72				0.0			
		0.21				0.00			
					4.00				

Fase II: Cálculo de la opción 1 y de la opción 2, en ese orden.

Igualmente el cuadro anterior presenta los resultados del cálculo de la Fase II, la cual consiste de calcular la Opción 1 o tardía o compra de Rinker, (árbol intermedio), y el cálculo de la Opción 2 temprana o compra de RMC (tercer árbol).

Opción 1. Cálculo de la Lattice de la Opción Segunda o Tardía, árbol intermedio en el Cuadro Anterior

Máximo [X, Mantener la Opción Abierta], excepto al vencimiento.

Ejecutar: X: [Evolución del Subyacente – X2]

X2: Costo de adquisición de Rinker

Mantener la Opción Abierta: resultado método backward.

Opción 2. Cálculo de la Lattice de la Opción Primera o Temprana (tercer árbol en el cuadro anterior)

Máximo [X', Mantener la Opción Abierta], excepto al vencimiento.

Ejecutar: X': (\$5.34 – X1)

\$5.34= resultado del Máximo [X, Mantener la Opción Abierta] ←del Paso Uno

X1: Costo de adquisición de RMC

Mantener la opción abierta: resultado método backward o reversible de la binomial, con los datos obtenidos del Paso 1 para el año 1.

Fase III. Cálculo Combinado de la Opción. En esta Fase se presentan los resultados de la combinación del Cálculo de Opción 1 y Opción 2.

FASE 3. OPCIÓN COMPUESTA SECUENCIAL RESULTADO COMBINADO DE OPCIÓN 1 Y 2			
			310.26 <-Invertir
		81.10 <-abierta	
	16.63 <-Invertir		12.90 <-Invertir
3.84 <-Abierta		2.98 <-abierta	
	0.00 <- No Inv		0.00 <- No Invertir
		0.00 <-abierta	
			0.00 <- No Invertir
	Opción Compuesta Secuencial:	3.8434	
	VPN:	15.1400	

El valor de adquirir a los competidores hoy, o el valor estático de valor presente neto (valor presente menos los dos costos de implementación): -US\$11.30

$$\$8.3 - \$15.3 - \$4.3 = - \$11.3$$

En este caso el resultado con el método tradicional VPN se rechaza la estrategia de adquisición de las dos empresas. Sin embargo el cálculo del valor combinado de la opción es de \$3.84. Debido a que se obtuvo de un máximo entre ejecutar las compras o mantener la opción abierta, el valor resultante indica que la compra no debe realizarse sino dejar la opción de invertir abierta y esperar a tener más información. Lo siguiente que indica los resultados combinados o de la fase III son muy interesantes: Sí se puede ejercer la compra en la primera vuelta (RMC) pero en condiciones económicas muy favorables. Curiosamente el año dos se debe esperar y finalmente el año 3 se indica que se puede invertir pero sólo en muy buenas, y buenas condiciones económicas, no así si la situación es mala o muy mala económicamente. Los resultados en este análisis coinciden con los resultados de un estudio previo (Cabello, 2009) para el caso de invertir en la adquisición de RMC usando un caso de opción de expansión.

2.4.4 Cálculo del valor de la opción secuencial compuesta de la estrategia de adquisiciones de CEMEX para el caso de RMC y Rinker utilizando simulación Monte Carlo

Una propuesta metodológica que se recomienda en la literatura de opciones reales¹¹ es la utilización complementaria de la técnica de simulación Monte Carlo para realizar proyecciones de variables de entrada de la opción real.¹² En esta sección se presenta la proyección de los flujos de efectivo necesarios para determinar el valor presente como variable de entrada para el cálculo de la opción secuencial compuesta.

¹¹ Copeland y Antikarov, 2001.

¹² Ver capítulo 7 de Lara Haro, 2009.

Se realizó un planteamiento de cuatro situaciones para el cálculo de los flujos de efectivo proyectados utilizando cifras anuales. En el siguiente cuadro se ejemplifica una sola situación de los resultados de la proyección de los flujos de efectivo con el método de simulación Monte Carlo, empleando la desviación estándar de los rendimientos de datos reales del periodo del 1988 al 2011. En la zona sin sombrear se muestran los flujos de efectivo históricos o reales que reportó Cemex en sus estados financieros, y la zona sombreada reporta los flujos de efectivo proyectados de Cemex que resultaron de la simulación Monte Carlo, de los años del 2005 al 2011.

Simulación Anual de Flujos de Efectivo de Cemex
Años simulados: 2005 al 2011
Con volatilidad histórica del 1988-2011
-Miles de Dólares-

	Flujos de Efectivo	Rendimiento de los Flujos
1988	166,771	
1989	161,691	-0.030934507
1990	376,068	0.844082872
1991	625,644	0.509011541
1992	549,581	-0.12962535
1993	1,225,407	0.801872143
1994	725,362	-0.524357472
1995	630,416	-0.140290922
1996	1,227,225	0.666130883
1997	1,240,616	0.010852507
1998	1,069,936	-0.148009197
1999	1,320,368	0.210311652
2000	1,418,319	0.071561882
2001	2,041,218	0.364074321
2002	1,205,372	-0.526758456
2003	1,296,109	0.072578466
2004	1,851,613	0.356690452
2005	2,116,205	0.133567463
2006	2,379,226	0.117150467
2007	2,595,563	0.08702819
2008	2,915,318	0.116175752
2009	3,280,102	0.117895353
2010	3,684,851	0.116355683
2011	4,115,643	0.110564994

El siguiente cuadro muestra el cálculo de la volatilidad y del valor presente obtenido de los flujos de efectivo proyectados con la simulación Monte Carlo.; se resumen las 4 situaciones empleadas. Se muestra la volatilidad resultante del rendimiento de los flujos de efectivo, así como el valor presente de los flujos proyectados. Ambos resultados constituyen dos de las principales variables de entrada para el cálculo de las opciones reales compuestas secuenciales.

**Simulación Montecarlo de la
Volatilidad del Rendimiento de los Flujos de Efectivo Proyectados
y del Valor Presente de los Flujos de Efectivo Proyectados de Cemex
-Miles de usd-**

Volatilidad Anual Real , y Valor Presente Anual Real de Cemex, 1988-2011:			
		Volatilidad	Valor Presente
		42.045%	\$ 11,704,171
	Periodo Real*	Años Proyectados de Flujos con Simulación MonteCarlo: 2005-2011	
		Volatilidad	Valor Presente de flujos de efectivo simulados
Situación 1	1988-2011	33.815%	\$ 11,705,369
Situación 2	1988-2004	33.774%	\$ 12,073,923
Situación 3	1988-2006	33.770%	\$ 12,486,051
	Periodo Real*	Años Proyectados de Flujos con Simulación MonteCarlo: 2007-2011	
		Volatilidad	Valor Presente de flujos de efectivo simulados
Situación 4	1988-2006	33.811%	\$ 12,246,335

* Periodo Real se refiere al empleo de datos reales de flujo de efectivo y del periodo indicado para el cálculo de la desviación estándar de los rendimientos de los flujos de efectivo de Cemex.

Las 4 situaciones de cálculo de flujos se refieren a lo siguiente:

Situación 1.- Se refiere a la proyección de los flujos de efectivo para los años del 2005 al 2011, empleando la desviación estándar de los rendimientos anuales reales del periodo 1988-2011 (después de las 2 compras).

Situación 2.- Se refiere a la proyección de los flujos de efectivo para los años del 2005 al 2011 (después de las 2 compras), empleando la desviación estándar de los rendimientos anuales reales del periodo 1988-2004 (antes de las dos compras).

Situación 3.- Se refiere a la proyección de los flujos de efectivo para los años del 2005 al 2011, empleando la desviación estándar de los rendimientos anuales reales del periodo 1988-2006 (después de la primera compra).

Situación 4.- Se refiere a la proyección de los flujos de efectivo para los años del 2007 al 2011, empleando la desviación estándar de los rendimientos anuales reales del periodo 1988-2006.

Las 3 primeras situaciones calculan el valor presente de los flujos de efectivo proyectados, al fin del año 2004 o antes de las dos compras. La situación 4 calcula el valor presente de los flujos de efectivo proyectados para el fin del año 2006 o antes de la segunda compra.

Con los datos calculados con la simulación Monte Carlo presentados en el cuadro precedente se procede a continuación a realizar el cálculo de la opción compuesta secuencial. Se aclara que sólo se realiza el cálculo sólo para las dos primeras situaciones.

Situación 1. Cálculo de la Opción Compuesta Secuencial proyección de los flujos de efectivo para los años del 2005 al 2011, empleando la desviación estándar de los rendimientos anuales reales del periodo 1988-2011 (después de las dos compras).

El siguiente cuadro muestra los resultados de la Opción Compuesta Secuencial que contempla el resultado del valor presente de los flujos al 2004.

**Situación 1. Fase I y Fase II con Simulación Montecarlo
para volatilidad: 33.81% y valor presente: \$ 11.7 miles de millones de USD**

Subyacente:	11.7	r	0.04	X1	4.3	-7.9			
		vol:	0.3382	X2	15.3				
u	1.4024	T1	1						
d	0.7131	T2	3						
p	0.4732	Lattice	3	3					
1-p	0.5268	dt	1.00						
		FASE 1			FASE 2				
		lattice del subyacente			Opción 1: compra tardía o de Rinker		Opción 2: compra temprana o de RMC		
			32.27			16.97			
		23.01			8.29				
1.40	16.41	16.41		4.03	1.11			0.00	
S	11.70	11.70		1.95	0.50			0.00	
0.71	8.34	8.34		0.23	0.00			0.00	
		5.95			0.00				
		4.24			0.00				
					0.00				

Opción 1= opción tardía (Rinker); Opción 2= opción temprana (RMC)

La interpretación cualitativa de la decisión queda como se muestra a continuación:

FASE 3. Cálculo Combinado de la Opción

FASE 3. OPCIÓN COMPUESTA SECUENCIAL	
RESULTADO COMBINADO DE OPCIÓN 1 Y 2	
	16.97 <-Invertir
	8.29 <-abierta
0.00 <-Abierta	0.00 <- No Invertir
	1.11 <-Invertir
	0.50 <-abierta
	0.00 <- No Invertir
	0.00 <-abierta
	0.00 <- No Invertir
Opcion Compuesta Secuencial:	0.0000
Valor Presente Neto:	0.00

El valor de adquirir a los competidores hoy, o el valor estático de Valor presente neto (valor presente menos los dos costos de implementación): -US\$ 7.9

$$\$11.7 - \$15.3 - \$4.3 = - \$7.9$$

En este caso el resultado con el método tradicional VPN rechaza la estrategia de adquisición de las dos empresas. El cálculo del valor combinado de la opción compuesta secuencial resulta de \$0.00¹³, debido a que se obtuvo de un máximo entre ejecutar las compras o mantener la opción abierta, el valor resultante indica que las compras no deben realizarse sino dejar la opción de invertir abierta y esperar a contar con más información. Los resultados intermedios de la combinación de la fase III son muy interesantes. No se ejecuta la compra en la primera vuelta (RMC) o primer año, ni aún en condiciones económicas muy favorables. En el año dos se debe esperar y finalmente el año 3 indica que se puede invertir en la compra de la empresa Rinker pero sólo en muy buenas, y buenas condiciones económicas; no así si la situación (esperada) es mala o muy mala económicamente. Este resultado coincide con el cálculo de la opción compuesta que no utilizó simulación, que se basó en resultados reales. Sin embargo la situación 1 incluye el uso de flujos reales de todo el periodo del 1988 al 2011 para el cálculo Monte Carlo por lo que hay puntos de coincidencia de las variables de entrada.

Situación 2. Cálculo de la Opción Compuesta Secuencial con proyección de los flujos de efectivo para los años del 2005 al 2011, empleando la desviación estándar de los rendimientos anuales reales del periodo 1988-2004 (antes de las dos compras).

Situación 1. Fase I y Fase II con Simulación Montecarlo
para volatilidad: 33.77% y valor presente: \$ 12.07 miles de millones de USD

Subyacente:	12.07	r	3.85%	X1	4.3	-7.53	7.53
		Vol:	33.77%	X2	15.3		
u	1.4018	T1	1				
d	0.7134	T2	3				
p	0.4734	Lattice	3	3			
1-p	0.5266	dt	1.00				
		FASE 1					
		lattice del subyacente					
			33.25				
			23.72				
1.40	16.92		16.92		9.00		
S	12.07	12.07		2.21	4.47	0.74	1.62
0.71	8.61	8.61			0.34		0.08
		6.14				0.00	0.00
		4.38				0.00	

¹³ La teoría indica que si el valor de la opción es bajo es mejor esperar; si el valor de la opción es muy alto, y generalmente mayor que el VPN, significa que el costo de esperar para contar con mayor información es excesivo y tampoco se debe ejercer. Los resultados del presente estudio indican que es mejor esperar.

FASE 3. Cálculo Combinado de la Opción

FASE 3. OPCIÓN COMPUESTA SECUENCIAL RESULTADO COMBINADO DE OPCIÓN 1 Y 2

			17.95 <-Invertir
		9.00 <-abierta	
0.08 <-Abierta	0.17 <-Invertir		1.62 <-Invertir
	0.00 <- No Invertir	0.74 <-abierta	0.00 <- No Invertir
		0.00 <-abierta	0.00 <- No Invertir
Opcion Compuesta			
Secuencial:	0.0779		
Valor Presente Neto:	-7.53		

El valor de adquirir a los competidores hoy, o el valor estático de Valor presente neto (valor presente menos los dos costos de implementación): -US\$ 7.53

$$\$11.7 - \$15.3 - \$4.3 = - \$7.53$$

En este caso el resultado con el método tradicional VPN está rechazando la estrategia de adquisición de las dos empresas. Sin embargo el cálculo del valor combinado de la opción resulta de \$0.0779 miles de millones de dólares de estados unidos, debido a que se obtuvo de un máximo entre ejecutar las compras o mantener la opción abierta. El valor resultante combinado indica que las compras no deben realizarse sino dejar la opción de invertir abierta y esperar a tener más información. Los resultados combinados o de la fase III en los nodos intermedios indican que a diferencia de la situación 1 en esta situación 2 sí se debe ejecutar la 1ª. compra en la primera vuelta (RMC) aunque en condiciones económicas muy favorables. El año dos es para esperar y finalmente el año 3 indica que se puede invertir en la compra de la segunda empresa (Rinker) pero sólo en muy buenas, y buenas condiciones económicas, y no si la situación está mala o muy mala económicamente. Este resultado es muy interesante puesto que la situación 2 se basa en el cálculo de la desviación de los flujos reales del periodo 1988 al 2004 o antes de las compras. Esta es la situación más idónea para la simulación suponiendo que estamos a fines del 2004 cuando se va a tomar la decisión de emprender primero la compra de RMC, y dos años después la compra de Rinker.

2.5 Resumen y Conclusiones

Desde la década pasada Cemex se ha convertido en una empresa global importante; entonces adquirió grandes corporaciones de su mismo sector, siendo las más significativas las que tuvieron lugar a mediados de la pasada década: la adquisición de compañía inglesa RMC en el año 2005 y la compra de la empresa australiana Rinker en el 2007. Realmente estas compras fueron parte de una agresiva estrategia de adquisiciones de empresas por parte de Cemex con el fin de ampliar estratégicamente sus negocios en mercados desarrollados, principalmente Estados Unidos y Europa. Sin embargo dichas adquisiciones fueron financiadas por el incremento de emisiones de deuda en los mercados financieros internacionales, bajo una creencia muy optimista del buen comportamiento de la economía global, por lo que sencillamente se esperaron buenos resultados en el rendimiento y los flujos de efectivo. Como consecuencia, en los años posteriores al año de adquisición de Rinker, Cemex ha mostrado tendencias financieras de alarmante deterioro. El análisis financiero muestra un importante declive de su rendimiento a la inversión implicado tanto en la pérdida de productividad como de rentabilidad, y sin una disminución importante de su endeudamiento. Esto ha conllevado a que el precio de la acción de Cemex igualmente se encuentre en niveles muy bajos, lo que implica que los riesgos de las compras suficientemente evaluados de tal manera que prevalecieron los riesgos de impacto negativo (*downside risk*) vis-a-vis los riesgos de impacto positivo (*upside risk*).

Considerando que este tipo de situación no es exclusivo para el caso de Cemex, este trabajo se propone y ejemplifica un análisis más sofisticado para utilizarlo para apoyar estrategias de adquisiciones recurrentes. Este es el enfoque de opciones reales compuestas secuenciales. En el presente trabajo se muestran los 3 principales pasos que involucra partiendo de la construcción de una *lattice* o reticulado en base al método binomial. Esto permite de manera inicial modelar un mundo cambiante, que en la práctica puede apoyar a los ejecutivos de una empresa a tomar decisiones correctas en cuanto a la evaluación de proyectos de inversión, no solo en cuanto a los riesgos asociados con dichos proyectos, sino también con relación a la oportuno (*timing*) del cuando aprovechar la oportunidad de inversión y del costo de obtener más información para posiblemente ejecutar el proyecto, i.e., el valor de la espera para no emprender la estrategia inmediatamente y esperar a tener más información.

Es así que primero se ejemplifica paso a paso el cálculo de 3 sendas metodológicas con información histórica de Cemex y posteriormente empleando la simulación Monte Carlo. La metodología de la opción real compuesta se refiere a calcular una opción sobre otra opción, traducida a realizar una compra sólo si la primera es conveniente. Los cálculos, prácticamente en todos los casos permiten concluir que Cemex no debió haber emprendido las compras, salvo en condiciones de buena y muy buena situación económica. Finalmente cabe subrayar que la ventaja de opciones reales consiste en crear una valoración agregada que no se avizora cuando se usa una herramienta tradicional como lo es valor presente neto que es un método estático en relación al enfoque de opciones reales. Las opciones compuestas coadyuvan a una mejor toma de decisiones

porque los proyectos de inversión realizados en cortos márgenes de tiempo no se deben ser vistos como proyectos individuales sino como proyectos interrelacionados.

2.5 Bibliografía

- Bailey, William, Benoit Couet, Ashish Bhandari, Soussan Faiz, Sundaram Srinivasan, & Helen Weeds, (2004) *Valoración de las Opciones Reales*, en Oilfield review (PRIMAVERA DEL 2004), en: http://www.slb.com/~media/Files/resources/oilfield_review/spanish04/spr04/p4_19.ashx, consultada el 23 de junio del 2012. (1-16 p.).
- Bucio Pacheco, Christian, (2012) *Dependencia y Valor en Riesgo: Un Estudio de Cópulas de los Mercados de Capitales de América*, Tesis de Doctorado, Facultad de Economía, Universidad Nacional Autónoma de México, México, D.F., 2012.
- Cabello Rosales, María Alejandra, (2009) *Enfoques de Valuación de la Empresa: Método Contable y Extendido y Método de Opciones Reales y Simulación con Aplicaciones al Caso Cemex*, Tesis de Doctorado, ITESM-CCM.
- Cemex, Informes Anuales, 2005, 2007, 2010, y 2011.
- Copeland, Tom, y Vladimir Antikarov, (2001) *Real Options. A practitioners Guide*, UK: London: Texere.
- De Lara Haro, (2009) *Medición y Control de Riesgos Financieros*. Incluye Riesgo de Mercado y de Crédito, México, D.F.: Limusa, 3ª. Ed..
- Geske, Robert, (1989) The Valuation of compound options, *Journal of Financial Economics*, 7: 63-81.
- Kulatilaka, N., (1995) *The Value of Flexibility: A general model or Real Options*. En Real Options in Capital Investments, Praeger .
- Kwok, Yue Kuen, Encyclopedía of Financial Engineering and Risk Management, Artículo sometido, en: <http://www.math.ust.hk/~maykwok/publications/Articles/comp%20option.pdf>, (consultado en Julio del 2012).
- Mun J., (2006) *“Real Options Analysis: Tools and Technique for Valuing Strategic Investment and Decisions.”* Nueva York, Nueva York, EUA: John Wiley & Sons.

Trigeorgis L., (1999) “*Real Options: A Primer*” en Alleman J. y E. Noam, eds. *The New Investment Theory of Real Options and Its Implications for Telecommunications Economics*. Boston, Mass. EUA.

Vedovoto, Graciela, y Diego Prior, (2010) “*Opciones: una propuesta para valorar proyectos de I+D en centros públicos de investigación.*” Universidad Autónoma de Barcelona: Trabajo de Investigación, Septiembre de 2010.

Yahoo: [http://: mx.finance.yahoo.com](http://mx.finance.yahoo.com) (consultado en Junio del 2012).

Capítulo 3

Concentración de la Exposición al Riesgo de los Retornos Accionarios en la Bolsa Mexicana de Valores en el Período 2004-2012, un Enfoque de Componentes Principales

Humberto Valencia-Herrera

Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Ciudad de México

Resumen

Durante el período de agravamiento de las consecuencias de la crisis crediticia en México, de junio de 2008 a febrero de 2009, se observa una mayor correlación en el movimiento de las activos en el mercado accionario mexicano. Una mayor proporción de la varianza es explicada por los componentes principales con mayores autovalores de la matriz de varianzas covarianzas. La tendencia iniciada en el período de junio de 2007 a mayo de 2008; aunque más débil a partir de marzo de 2009, aún no regresaba a los niveles previos a 2007 en el período de enero 2011 a junio de 2012,. Estas diferencias son estadísticamente significativas; el primer componente principal en cada uno de los sub-períodos de análisis es estadísticamente diferente al de otros sub-periodos o al de la muestra general, pero tienen correlación estadísticamente significativa entre sí. La hipótesis de una correlación igual a uno no se puede rechazar para el primer componente principal entre el período de análisis, de enero de 2004 a junio de 2012, y los sub-periodos de incertidumbre, de junio de 2007 a mayo de 2008, de crisis mexicana, de mayo de 2008 a febrero

de 2009 y de estabilización, de marzo de 2009 a diciembre de 2010, y entre los sub-periodos de crisis mexicana y de estabilización.

3.1 Introducción

El mercado accionario de México sufrió los efectos de la crisis crediticia en Estados Unidos de 2007-2009, y más recientemente, la crisis de deuda europea. En este artículo analiza usando componentes principales como se modificó la concentración de riesgo en el mercado mexicano dado estos cambios.

El aumento de los precios en el mercado inmobiliario de los Estados Unidos de 2001 a 2005 escondió la situación de las hipotecas basura, que tenían ya un deterioro de su calidad crediticia desde esa época (Demyanyk y Hemmert, 2007), sin embargo las entidades financieras no tomaron previsiones. Con el estallamiento de la burbuja inmobiliaria a mediados de 2006, la relevancia de la crisis de los bonos basura estructurados en los Estados Unidos tomo importancia para 2007. Sólo cuando las entidades calificadoras de riesgos como Standard and Poors y Moodys empezaron a ser investigadas por su responsabilidad y sólo cuando la crisis ya había iniciado, las calificadoras endurecieron sus criterios (Gorton, 2008), atrasos en ajustes de calificaciones que se han observado ya en otras instancias (Violi, 2004; BIS, 2005). Tres razones se han propuesto para explicar las altas calificaciones dadas a los bonos estructurados basura. La primera, que las calificadoras habían sido deliberadamente agresivas en su calificación de activos, asignando calificaciones muy altas. Esta explicación es coherente con la hipótesis de “búsqueda de calificación atractiva,” en que los emisores se quedaban con la agencia que les ofrecía la mejor calificación, lo que generó una inflación de calificaciones (Benmelech y Dlugosz, 2009). La segunda explicación es que había errores en los modelos de valuación, en particular en la correlación de las probabilidades de incumplimiento. La tercera razón tiene que ver con el gobierno corporativo y la gestión de riesgo en grandes instituciones financieras. Aún si el riesgo hubiera sido correctamente entendido en las instituciones, directivos de alto nivel y mesas de negociación pudieron neutralizar las alertas de los administradores de riesgo para mantener, o aun mejorar, sus paquetes de remuneración en el corto plazo, con un exceso de toma de riesgo (Violi, 2010). Para septiembre de 2008, Lehman

Brothers, el cuarto banco de inversión norteamericano, que gestionaba 46,000 millones de dólares en hipotecas, se declaró en quiebra. A partir de ese momento, las autoridades financieras norteamericanas toman medidas extremas para contrarlar la crisis crediticia. El gobierno de Estados Unidos anuncia un plan de rescate multimillonario, los dos bancos de inversión independientes, Goldman Sachs y Morgan Stanley, son convertidos en bancos comerciales, y la FED y JP Morgan rescataron a la mayor caja de ahorros del país, Washington Mutual (Perez, 2008). En lo internacional, un grupo de diez bancos internacionales, entre los que están el propio Bank of America, Barclays, Deutsche Bank y UBS, crean un fondo como garantía para el sistema financiero, en el que cada entidad aporta 7,000 millones de dólares. A partir de este momento, las bolsas internacionales, tienen tendencias bajistas, el 10 de octubre de 2008, hay una caída abrupta de las bolsas internacionales (Cacho, 2008), luego viene una recuperación, pero para el 25 de octubre de 2008, la caída de las bolsas es mundial, el peligro de una recesión mundial es declarada (Laffont, 2008).

Para enero de 2009, la economía mexicana, altamente dependiente de la economía estadounidense, estaba sufriendo las consecuencias de la crisis internacional. El gobierno ante la inminente caída de los ingresos petroleros se ve obligado a aumentar impuestos y tasas de interés de los bonos gubernamentales, para atraerse de recursos (Martínez, 2009) y el tipo de cambio sufre de una depreciación histórica, que para contrarestarla, se inician subastas diarias de venta de dólares y se tiene una histórica pérdida de reserva, la economía entra en recesión.

Para finales de 2009, el peligro de crisis de deuda soberana en países europeos se intensifica. El 9 de mayo de 2010, los ministros de finanzas de Europa aprueban un plan de rescate por 750,000 millones de euros para crear un Fondo de Estabilidad Financiera Europeo. En Octubre de 2011 y Febrero de 2012, los líderes de la Eurozona aprueban más medidas para prevenir el colapso de sus economías.

3.2 Metodología

La técnica de componentes principales (CP) tiene como objetivo determinar un conjunto pequeño de componentes que expliquen mejor la variación total de una muestra, Barber y Copper (2012), es una técnica que permite aproximar la matriz de covarianzas con un matriz de rango más pequeño. Este objetivo difiere con el de la técnica de análisis de factores (AF), que busca

encontrar un conjunto pequeño de factores que expliquen mejor la variación conjunta de la muestra. Por tanto para analizar la concentración de la variación en diferentes períodos, consideramos que es más adecuada la técnica CP que AF. Además, si los autovectores son una medida aceptable de riesgo sistemático, los componente principales tienen la ventaja que son únicos hasta por una transformación de escala y que no son afectados por rotaciones, mientras que los factores son sólo únicos si la rotación es ortonormal. Esto significa, por ejemplo, que los estadísticos t tienen significado sólo para los componentes, pero no para factores. Además los componentes principales son más simples de calcular que los factores, y si los autovectores son una medida aceptable de riesgo, se preferirán en muchas aplicaciones, Trzcinka (1986) y Schukla y Trzcinka (1990). La técnica también se ha empleado para explicar la estructura de tasas de interés (Barber y Copper, 1996; Golub y Tillman, 2000; Novosyolov y Satchkov, 2008; Barber y Copper, 2012), para proponer estrategias de cobertura de bonos (Falkenstein y Hanweck, 1997), para inmunización (Soto, 2004), estimación de valor en riesgo (Abad y Benito, 2007) o pronóstico de tasas de interés (Reisman y Zohar, 2004). Sin embargo, en estos artículos no se emplea analizando la significancia estadística de los componentes principales de un mercado accionario.

Para ilustrar la técnica de CP, asumamos que mucha, pero no toda, la variación en los retornos X_t , resumida en la matriz de varianzas covarianzas, puede ser explicada por un solo componente principal, U_1 . Si ε_t considera el término de error,

$$X_t = b_{1t}U_1 + \varepsilon_t.$$

Si m es el número de activos, la variación puede ser explicada como,

$$X_t = b_{1t}U_1 + \dots + b_{mt}U_m + \varepsilon_t,$$

donde b_{kt} es la proyección de X_t en U_k .

Lo que se quiere en la práctica es desarrollar un modelo en el cual el número de componentes es pequeño, $K \ll m$, pero que tiene un poder de explicación alto.

$$X_t = b_{1t}U_1 + \dots + b_{Kt}U_K + \varepsilon_t = \tilde{X}_t + \varepsilon_t,$$

donde \tilde{X}_t es la aproximación (o explicación) del valor de X_t por el modelo de K componentes y ε_t es el término de error. Si se asume que el conjunto de vectores de dirección U_1, \dots, U_m son

ortonormales, en el momento t los componentes, por definición serán la proyección de X_t en los vectores U_1, \dots, U_K :

$$b_{kt} = U_K^T X_t \text{ para } k=1, 2, \dots, K,$$

la suma de los errores al cuadrado al tiempo t es igual a la diferencia entre la suma total de cuadrados y la suma de cuadrados explicada

$$\varepsilon_t^T \varepsilon_t = X_t^T X_t - \tilde{X}_t^T \tilde{X}_t$$

Dado que $X_t^T X_t$ es constante, para minimizar los errores al cuadrado basta maximizar la suma de los cuadrados explicados $\tilde{X}_t^T \tilde{X}_t$ o la varianza explicada, esto es:

$$\frac{1}{T} \sum_{t=1}^T \tilde{X}_t^T \tilde{X}_t = \frac{1}{T} \sum_{t=1}^T (U_1^T X_t^T X_t U_1 + \dots + U_K^T X_t^T X_t U_K) = [U_1 \dots U_K]^T V [U_1 \dots U_K], \quad (3.1)$$

donde V es la matriz de varianzas-covarianzas de la muestra.

Si la matriz V es positiva definida, tendrá m autovalores distintos positivos $\lambda_1, \dots, \lambda_m$, los autovectores $Q = [Q_1, \dots, Q_m]$ serán ortonormales y la matriz de varianzas-covarianzas se podrá factorizar (ver, por ejemplo, Strang 1980) como

$$V = Q \Lambda Q^T, \quad (3.2)$$

donde Λ es una matriz diagonal positiva, $\lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$. Si se sustituye la ecuación (3.2) en la ecuación (3.1), se obtiene:

$$\frac{1}{T} \sum_{t=1}^T \tilde{X}_t^T \tilde{X}_t = \frac{1}{T} \sum_{t=1}^T [U_1 \dots U_K]^T Q \Lambda Q^T [U_1 \dots U_K]. \quad (3.3)$$

Si los autovectores y los elementos de λ se ordenan en orden decreciente de tamaño de los autovalores, en el modelo de un componente ($K=1$) la expresión (3.3) se maximiza con $U_1 = Q_1$ y la suma de la varianza explicada es igual al primer autovalor. Más generalmente, si $K=m$, $U_1 = Q_1, \dots, U_m = Q_m$, y la suma de la varianza explicada es igual a la suma de los primeros autovalores y la suma de las varianzas de los errores es igual a la suma de los restantes autovalores.

Otra forma de ver CP, es como un método que aproxima la matriz de varianzas-covarianzas con una matriz de rango menor. La ecuación (3.2) se puede escribir como la suma de matrices $m \times m$ de rango uno:

$$V = \lambda_1 U_1 U_1^T + \dots + \lambda_m U_m U_m^T,$$

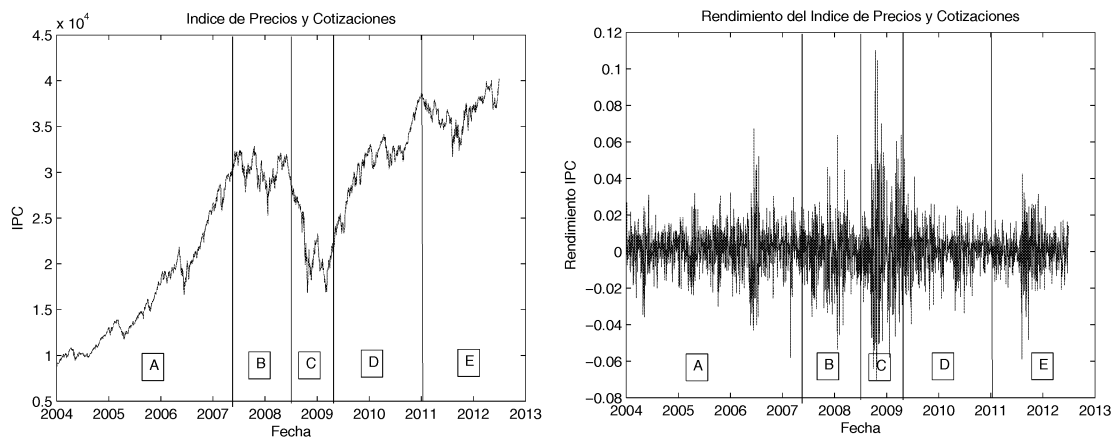
esto es, podemos aproximar V con una matriz de rango $K < m$ igual a la suma de los primeros K términos.

3.3 Análisis empírico

La base de datos es de Econométrica, de la que se estimaron rendimientos diarios de las acciones en base a los precios de cierre ajustados por ajustes corporativos y dividendos, excluyendo días festivos en México de la Bolsa Mexicana de Valores del 2 de enero de 2004 al 29 de junio de 2011, para un total de 2,145 observaciones. Analizando el comportamiento del Índice de Precios y Cotizaciones (IPC), se observaron cinco períodos, un período pre-crisis, de enero de 2004 a mayo de 2007, el período de crisis internacional con incertidumbre en México, de junio de 2007 a mayo de 2008, el periodo de crisis en México de junio de 2008 a febrero de 2009, el período de estabilización, de marzo de 2009 a diciembre de 2010 y el período de la crisis europea, de enero de 2011 a junio de 2012, en base al comportamiento del IPC, ver la Figura 3.1.

En el período de pre-crisis 2004-07, el retorno diario del IPC fue de 0.154 % con desviación estándar de 1.165 %, en el período de crisis internacional con incertidumbre en México 2007-08, el retorno se reduce a 0.017 % y la desviación estándar 1.544 %. También la distribución es menos leptocúrtica, pasa de 6.024 a 4.553 y el sesgo se reduce de 0.114 a -0.021. En el período de la crisis mexicana 2008-09, el retorno se vuelve negativo, -0.278 %, la volatilidad aumenta a 2.524 % y la distribución de los retornos se sesga a la derecha, el sesgo es igual a 0.802. En el período de estabilización 2009-10, el retorno diario es de 0.174 %, la desviación estándar es de 1.292 %, el sesgo es a la derecha, 0.334, y la curtosis de 6.355 es similar a la del período 2004-07. En el período de la crisis europea 2011-12, la media del rendimiento se reduce a 0.016 %, la desviación estándar a 1.107 % y el sesgo se vuelve negativo, -0.412, ver Figura 3.1.

Se seleccionaron las series accionarias que tuvieron al menos 30 rendimientos diferentes a cero en cada sub-periodo y que correspondieran a diferentes emisoras, para un total de 61 series. En el caso de que una emisora tuviera varias series, se eligió la de mayor capitalización al final del período de análisis. Para la muestra general y las muestras en cada sub-período, se estimaron los componentes principales y los auto- valores. Se analiza el porcentaje que explican de la varianza y, estadísticamente, su persistencia a lo largo del período de estudio. Aunque el excluir emisoras que tuvieron poca bursatilidad o no cotizaron durante todo el período de estudio resulta en un sesgo por sobrevivencia, diversos análisis efectuados muestran que los resultados de este estudio se mantendrían aún con este sesgo.



	2004-12	A:2004-07	B:2007-08	C:2008-09	D:2009-10	E:2011-12
Observaciones	2145	865	251	189	463	377
Media	0.080%	0.154%	0.017%	-0.278%	0.174%	0.016%
Desv. Est.	1.405%	1.165%	1.544%	2.524%	1.292%	1.107%
Sesgo	0.227	0.114	-0.021	0.802	0.334	-0.412
Kurtosis	9.009	6.024	4.553	6.555	6.355	6.514
Jarque-Bera	3,245.10	331.44	25.24	119.80	225.77	204.60

Figura 3.1: Retorno del IPC en cada sub-periodo

Se seleccionaron las series accionarias que tuvieron al menos 30 rendimientos diferentes a cero en cada sub-periodo y que correspondieran a diferentes emisoras, para un total de 61 series. En el caso de que una emisora tuviera varias series, se eligió la de mayor capitalización al final del período de análisis. Para la muestra general y las muestras en cada sub-período, se estimaron los componentes principales y los auto- valores. Se analiza el porcentaje que explican de la varianza y, estadísticamente, su persistencia a lo largo del período de estudio. Aunque el excluir emisoras que tuvieron poca bursatilidad o no cotizaron durante todo el período de estudio resulta en un sesgo por sobrevivencia, diversos análisis efectuados muestran que los resultados de este estudio se mantendrían aún con este sesgo.

3.3.1. Prueba de igualdad de autovalores pequeños

Una prueba de aleatoriedad es que los autovalores más pequeños sean estadísticamente semejantes, lo que se requiere para que sean ruido blanco, esto es se puedan considerar errores idénticos e independientemente distribuidos. Para lo cual se efectúan pruebas estadísticas de que los autovalores más pequeños de los componentes principales fueran iguales entre sí,

$$H_K : \lambda_{K+1} = \dots = \lambda_m,$$

para $K = 0, 1, \dots, m-2$.

En el caso extremo de que la hipótesis H_0 fuera cierta, la distribución de los componentes principales X sería esférica, sin ninguna dirección de los retornos más probable que otra. Si se acepta la hipótesis H_1 de que sólo un auto-valor es diferente, los auto-valores $\{\lambda_2, \dots, \lambda_m\}$ son todos iguales y efectivamente un solo componente explicaría el comportamiento de los retornos, lo que no explica este componente se podría interpretar como ruido. Si H_K se acepta para $K \ll m$ y el valor común $\lambda = \lambda_{K+1} = \dots = \lambda_m$ es pequeño, los retornos dependerían de K componentes y las fluctuaciones en direcciones asociadas con λ serían errores de medición.

Para probar la hipótesis H_K , se establece

$$\tilde{l}_q = \frac{1}{q} \sum_{i=K+1}^m l_i,$$

donde los l_i son autovalores de la matriz muestral de covarianzas V y por conveniencia notacional, $q=m-K$, el número de autovalores hipotéticamente iguales. Ahora, sea

$$L_K = \frac{\prod_{i=K+1}^m l_i}{\tilde{l}_q}.$$

Cuando H_K es verdadera, la distribución de $-n \log L_K$ es asintóticamente χ^2 con $(q+2)(q-1)/2$ grados de libertad cuando $n \rightarrow \infty$. Dado que los datos con que se cuentan son 2,145 en el período 2004 a 2012 y los sub-periodos analizados tienen de 189 hasta 865 datos, empleamos el estadístico más preciso desarrollado por Barlett y Lawley, descrito en Muirhead (1982) y Barber y Copper (2012):

$$P_K = - \left[n - K - \frac{2q^2 + q + 2}{6q} + \sum_{i=1}^K \frac{\tilde{l}_q^2}{(l_i - \tilde{l}_q)^2} \right] \log L_K,$$

que tiene la misma distribución asintótica pero más rápida convergencia cuando $n \rightarrow \infty$

Cuadro 3.1: Prueba de H_2 , los dos auto-valores más pequeños son iguales

	2004-07	2007-08	2008-09	2009-10	2011-12	2004-12
N	865	251	189	463	377	2145
$-n \log L_K$	407.47	187.66	195.59	291.84	233.57	990.82
P_K	1,473	977	696	3,893	3,680	19,327
$\chi^2(2)$	5.99	5.99	5.99	5.99	5.99	5.99

La hipótesis nula H_2 es rechazada en todos los casos.

El cuadro 3.1 muestra los estadísticos $-n \log L_k$ y P_k para los subperíodos y el período general de análisis para el caso de H_2 , $q=2$, los dos autovalores más pequeños son iguales, lo cual implica que se rechaza, al igual que todos los casos en que q es mayor a dos.

Cuadro 3.2: Los diez auto-valores mayores en los sub-períodos y el periodo general de análisis.

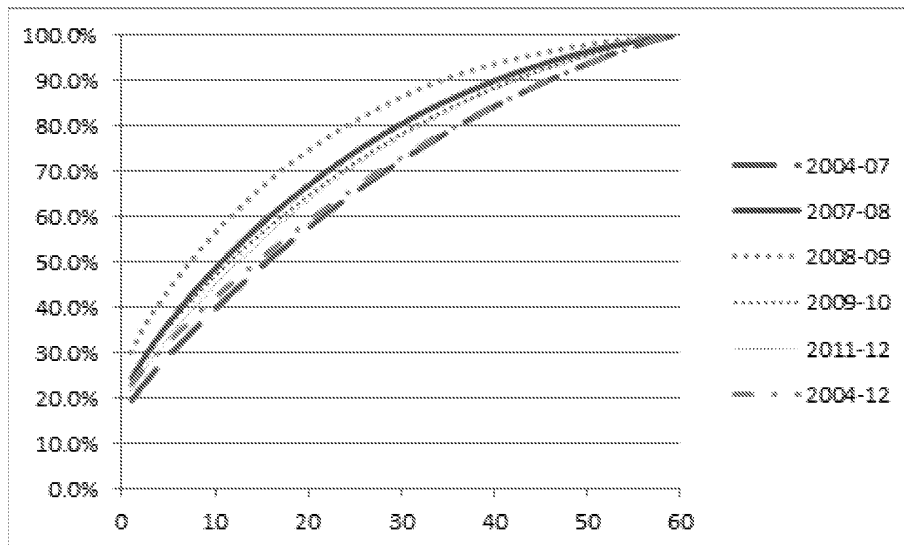
2004-07	2007-08	2008-09	2009-10	2011-12	2004-12
11.5019	14.4559	17.8145	14.4703	13.0507	13.4937
1.593	1.9854	2.7009	2.0782	2.0732	1.6045
1.5214	1.8291	2.303	1.722	1.7287	1.4378
1.4357	1.7349	2.173	1.6767	1.643	1.304
1.3916	1.603	1.7226	1.51	1.5647	1.2265
1.3283	1.578	1.6209	1.4129	1.4795	1.2031
1.2695	1.5545	1.5767	1.3371	1.4304	1.1828
1.2195	1.4631	1.5146	1.3114	1.3477	1.1674
1.1936	1.4224	1.409	1.2672	1.3237	1.1415
1.1665	1.3474	1.3317	1.265	1.2776	1.1192

Del Cuadro 3.2 se puede observar que el primer componente principal tiene un autovalor mucho mayor que los siguientes, que disminuyen gradualmente a partir del segundo autovalor. Si comparamos entre períodos, el autovalor del primer componente principal en el período 2008-09 es mucho más grande, 17.8145, seguidos por los de los períodos 2007-08 y 2009-10, 14.4559 y 14.4703, respectivamente, lo que muestra que la variabilidad de la rentabilidad se explica en mayor proporción por este componente principal en estos períodos. El autovalor del primer componente principal de los períodos 2004-07, 2011-12 y el general, es de magnitud menor, 11.5019, 13.0507 y 13.4937, respectivamente, lo que muestra que la variabilidad de la rentabilidad es tuvo menos concentrada en este factor.

La explicación de la varianza por los cinco y diez mayores componentes principales también está más concentrada durante el período 2008-09 (43.8 % y 56.0 %, respectivamente)

que el resto de los sub-periodos. Esta concentración se observa que aumentó respecto al sub-período anterior de análisis en el período 2007-08 (35.4 % y 47.5 %), en 2009-10 disminuyó ligeramente respecto al sub-período anterior de análisis, (35.2 % y 46.0 %), pero aún para 2011-12 (32.9 % y 44.1 %), no se había regresado a la dispersión del período pre-crisis, ver Figura 3.2.

Figura 3.2: Porcentaje de la varianza explicada por los n mayores autovalores



Período	2004-07	2007-08	2008-09	2009-10	2011-12	2004-12
5 comp.	28.6 %	35.4 %	43.8 %	35.2 %	32.9 %	31.3 %
10 comp.	38.7 %	47.5 %	56.0 %	46.0 %	44.1 %	40.8 %

3.3.2. Prueba estadística de igualdad de componentes principales

Un proceso estacionario presenta componentes principales persistentes en el tiempo. En particular, en este estudio se analiza estadísticamente si el primer componente principal en los sub-períodos y en el período de estudio es el mismo. Esto es, se prueba la hipótesis nula H de que el primer componente principal es igual a un cierto autovector de tamaño unitario,

$$H : U_1 = U_1^o \text{ y } U_1^{oT} U_1^o = 1.$$

T. W. Anderson, citado en Muirhead (1982, sección 9.6) y Barber y Copper (2012), muestra que H es verdadera si los siguientes estadísticos

$$W_1 = nl_1 U_1^{oT} V^{-1} U_1^o - n \text{ y } W_2 = \frac{n}{l_1} U_1^{oT} V^{-1} U_1^o - n,$$

se suman formando un estadístico único $W = W_1 + W_2$ que tiene una distribución limitada χ_{m-1}^2 donde V es una matriz de covarianzas $m \times m$ de una muestra de rendimientos, l_1 es el auto-valor del primer componente principal de V . La hipótesis H dado un tamaño asintótico α se rechaza si $W > F^{-1}(\alpha, m-1)$ donde $F^{-1}(\alpha, m-1)$ es el valor de la cota de la cola superior de una distribución χ_{m-1}^2 con probabilidad α .

Cuadro 3.3: Prueba de igualdad del primer componente principal entre sub-periodos

Período base \ sub-período		2004-07	2007-08	2008-09	2009-10	2011-12
2004-2011	W_1	624.02	227.61	334.77	321.89	509.96
	W_2	-53.91	-11.00	-5.09	-14.78	-28.21
	W	570.12	216.61	329.68	307.11	481.74
2004-2007	W_1		569.30	1,539.70	1,500.00	1,265.20
	W_2		-29.51	-24.50	-66.81	-65.36
	W		539.80	1,515.20	1,430.00	1,199.90

El valor de la cota superior de χ_{m-1}^2 , $F^{-1}(0.05, 60) = 79.08$, por lo que se rechazan todas las hipótesis nulas de igualdad en el primer componente principal.

Se probaron dos hipótesis: la primera — el componente principal en el período es estadísticamente igual al del período 2004-2012, y la segunda, el componente principal en el período es estadísticamente igual al del período 2004-2007. Como se observa en el cuadro 3.3, en ambos casos las hipótesis se rechazan para todos los sub-periodos. Por tanto, podemos rechazar la hipótesis de que hay estacionalidad a lo largo del período de estudio de los retornos de los activos, ya que al menos el primer componente principal no es estadísticamente igual al del período 2004-12 o al del período 2004-07 en cada uno de los otros sub-periodos. Nótese también que en la prueba de igualdad con el componente principal del sub-período 2004-07, el estadígrafo

W es mayor en el subperiodo 2008-09, $W = 1,515.20$ que en los otros sub-períodos, lo que muestra un cambio mayor en las tendencias.

Para probar que la hipótesis $H_0 : \rho = \rho_0$ consideramos el estadígrafo (Ver De Fusco et ál. (2004, capítulo 8):

$$t = \frac{(r - r_0)\sqrt{n-2}}{\sqrt{1-r^2}}$$

La hipótesis H_0 : la correlación es cero entre los primeros componentes principales de cada sub-periodo se rechaza estadísticamente para todos los sub-periodos. La hipótesis H_0 : la correlación es uno entre el primer componente principal de dos sub-períodos no se puede rechazar estadísticamente para el período 2004-2012 y los sub-periodos 2007-08, 2008-09 y 2009-10 y para la correlación entre los sub-períodos 2008-09 y 2009-10, ver el cuadro 3.4.

Cuadro 3.4: Correlación entre los primeros componentes principales de cada subperiodo.

	2004-12		2004-07		2007-08		2008-09		2009-10		2011-12	
2004-12	1		0.847	*	0.892		0.931		0.916		0.8	*
2004-07	0.847	*	1		0.578	*	0.652	*	0.695	*	0.725	*
2007-08	0.892		0.578	*	1		0.663	*	0.797	*	0.825	*
2008-09	0.931		0.652	*	0.663	*	1		0.88		0.736	*
2009-10	0.916		0.695	*	0.797	*	0.88		1		0.799	*
2011-12	0.8	*	0.725	*	0.825	*	0.736	*	0.799	*	1	

La Hipótesis H_0 : La correlación es 0, se rechaza en todos los casos al 95 %. *La hipótesis H_0 : la correlación es 1, se rechaza al 95 %.

3.4 Conclusiones

En este estudio se analiza la significancia y persistencia de los componentes principales de la matriz de varianzas-covarianzas durante el período enero de 2004 a junio de 2012.

Se observa que el peso de los componentes principales que mayormente explican la varianza fue mayor en el sub-período de junio de 2008 a febrero de 2009, cuando las

consecuencias de la crisis internacional sobre México se agravaron, situación que aún para el sub-periodo de enero de 2010 a junio de 2012, aún no regresa a lo observado en el período de pre-crisis, de enero de 2004 a mayo de 2007.

En todos los períodos analizados, el de la muestra general y en los sub-períodos que se dividen, no se puede mostrar estadísticamente que haya autovectores que sean ruido blanco. La persistencia de los componentes principales a lo largo del período de estudio se rechaza estadísticamente, al rechazarse que el primer componente principal en las sub-períodos sea igual al del periodo total o al del primer sub-periodos, con lo que se rechaza la hipótesis de estacionariedad en los rendimientos de la Bolsa Mexicana de Valores.

Sin embargo, pruebas de correlación, asumiendo normalidad, rechazan que el principal componente principal en los sub-períodos tenga correlación de cero con los demás. Aunque sólo no se puede rechazar estadísticamente que la correlación sea igual a uno entre el período general y los sub-períodos de junio de 2007 a mayo de 2008, de junio de 2008 a febrero de 2009 y de marzo de 2009 a diciembre de 2010 y entre los sub-períodos de junio de 2008 a febrero de 2009 y de marzo a 2009 a diciembre de 2010; lo que muestra una cierta continuidad de tendencias en el período de la crisis crediticia y posteriormente.

3.5 Bibliografía

- Abad P. y S. Benito (2007). Value at risk in fixed income portfolios: a comparison between empirical models of term structure, *Journal of American Business Review*, Cambridge, 7(2) 342
- Barber, J. R. y M. L. Copper (1996). Immunization using principal component analysis, *Journal of Portfolio Management*, 23(1), 99-105.
- Barber, J. R. y M. L. Copper (2012). Principal component analysis of yield curve movements, *Journal of Economics and Finance*, 36, 750-765.

Benmelech E. y J. Dlugosz (2009). The credit rating crises, NBER Working Paper No. 15045, Junio.

BIS, Bank of International Settlements (2005). The role of ratings in structured finance: issues and implications, Committee on the Global Financial System (CGFS) Publications, 23, January, disponible en <http://www.bis.org/publ/cgfs23.htm>.

Cacho, J. (2008). Semana de pasión, viernes de dolores, El Confidencial, 12 de octubre de 2008, http://www.elconfidencial.com/cache/2008/10/12/conlupa_46_semana_pasion_viernes_dolores.html

De Fusco, R., D. W. McLeavey, Dennis, J. E. Pinto y D. E. Runkle (2004). Quantitative Methods for Investment Analysis, 2 edition, CFA Institute.

Demyanyk, Y. y O. V. Hemert (2007). Understanding the Subprime Mortgage Crisis, Federal Reserve Bank of Saint Louis, Working Paper 2007-05, http://www.stlouisfed.org/banking/pdf/SPA/SPA_2007_05.pdf, 1-32.

Falkenstein, E. y J. J. Hanweck (1997). Minimizing basis risk from non-parallel shifts in the yield curve part ii: principal components, Journal of Fixed Income, 7(1), 85.

Golub, B. W. y L. M. Tillman (2000). Risk management: approaches for fixed income markets, Wiley, New York.

Gorton, G. B. (2008). The subprime panic, NBER Working Paper No 14398, 2008.

Laffont, B. (2008). Desplome bursátil mundial en medio del pánico a la recesión económica, 25 de octubre de 2008, El País, <http://elpais.com/diario/2008/10/25/economia>

Martínez, R. (2009) Aumenta desempleo en México durante enero de 2009: Inegi, en IBT Times, 25 de febrero de 2009, consultado el 25 de julio de 2009.

Muirhead, R. J. (1982). Aspects of multivariate statistical theory, Wiley, New York.

Novosyolov, A. y D. Satchkov (2008). Global term structure modelling using principal component analysis. Journal of Asset Management, 9(1), 49.

- Perez, C. (2008). Estados Unidos acaba con 80 años de banca de inversión forzado por la crisis, El País, 23 de septiembre, http://elpais.com/diario/2008/09/23/economia/1222120802_850215.html
- Reisman, H. y G. Zohar (2004). Short term predictability of the term structure. *Journal of Fixed Income*, 14(3) 7
- Schukla, R., C. Trzcinka (1990). Sequential tests of the arbitrage pricing theory: a comparison of principal components and maximum likelihood factors, *Journal of Finance*, 1541-1564.
- Soto, G. M. (2004). Duration models and IRR management: a question of dimensions?, *Journal of Banking and Finance*, 28(5), 1089.
- Strang, G. (1980) *Linear algebra and its applications*, Academic Press, New York.
- Trzcinka, C. (1986). On the number of factors in the arbitrage pricing model, *Journal of Finance*, 41, 347-368.
- Violi, R. (2004). Credit ratings transition in structure finance, CGFS Working Group on Ratings in Structure Finance Paper, BIS.
- Violi, R. (2010). Credit ratings in structured finance: some lessons from the current credit crisis and beyond, *Journal of Structured Finance*, Fall, 65-87.

Capítulo 4

Modelos de calificación crediticia: Técnicas estadísticas tradicionales y de aprendizaje maquinal

Adán Díaz Hernández

Facultad de Economía y Negocios de la Universidad Anáhuac México-Norte

John Goddard Close

Departamento de Ingeniería Eléctrica de la Universidad Autónoma Metropolitana, U-Iztapalapa

Resumen.

En este trabajo se presenta una propuesta de modelo de calificación crediticia mediante el uso de técnicas de reconocimiento de patrones y modelos estadísticos tradicionales. La idea principal de este proyecto de investigación consiste en realizar una comparación del desempeño de los modelos estadísticos utilizados tradicionalmente en la industria financiera, con metodologías basadas en clasificadores desarrollados en el área de aprendizaje maquinal y reconocimiento de patrones. Esto se realiza bajo ejercicios de experimentación con distintos conjuntos de datos, tanto públicos como privados, en el contexto nacional como internacional.

Uno de los objetivos del trabajo se enfoca en realizar un análisis para el caso mexicano con información del mercado local que establezca las bases para el uso de alternativas cuyo desempeño sea competitivo, o incluso superior, respecto a los enfoques tradicionales empleados para discriminar el comportamiento y calidad crediticia de los clientes a quienes las instituciones otorgan crédito. Este problema se aborda desde una perspectiva estática (clasificación de buenos y malos clientes en un punto en el tiempo) y presenta las bases para futuros análisis dinámicos (estudio del comportamiento del riesgo asociado a incumplimiento de los clientes).

4.1 Introducción

El problema de calificación crediticia (CS, por sus siglas en inglés: *credit-scoring*) representa una de las primeras aplicaciones de la minería de datos, inclusive antes de que el término *data mining* apareciera. Las instituciones de crédito tradicionalmente emplean metodologías estadísticas para discriminar a los clientes con buen perfil de riesgo de aquellos que potencialmente pueden presentar un mal perfil. Sin embargo, en los últimos años, diversas alternativas se han desarrollado en el área de aprendizaje maquina (*machine learning*) para abordar este problema. El aprendizaje maquina es una rama de la inteligencia artificial en la cual se diseñan y desarrollan algoritmos que permiten generar conclusiones sobre patrones a partir de datos empíricos. Los mecanismos de aprendizaje utilizan ejemplos (datos) para capturar características de interés de la distribución de probabilidad subyacente. Una de las líneas de investigación en aprendizaje maquina se centra en el estudio del aprendizaje automatizado para reconocer patrones complejos y realizar toma de decisiones sobre la base de cierto conjunto de datos. A esta área del aprendizaje maquina se le denomina reconocimiento de patrones.

La capacidad de las instituciones financieras para pronosticar o anticipar la calidad de los portafolios de créditos que mantienen resulta un tema crucial para la continuidad de su negocio y más aún, para garantizar la estabilidad del sistema financiero y el crecimiento económico en su conjunto. Precisamente, la reciente crisis financiera global tuvo sus orígenes en los severos problemas financieros que enfrentó el sector hipotecario estadounidense, producto, entre otros factores, de una mala gestión del riesgo de crédito. Las deficiencias en los sistemas evaluación y calificación crediticia, tanto para contrapartes públicas como los acreditados en los distintos segmentos hipotecarios, condujeron no solamente a malas estimaciones de los riesgos financieros asociados sino a valuaciones erróneas de distintos activos y productos financieros derivados en el mercado¹. El uso de adecuados sistemas de calificación crediticia es entonces un tema central para instituciones y reguladores a nivel local e internacional.

Típicamente, las aproximaciones empleadas en la industria financiera para la construcción de modelos de CS descansan fuertemente en metodologías estadísticas tradicionales como el análisis discriminante, la regresión logística, árboles de decisión, y, en algunos casos, la aplicación de redes neuronales artificiales. Al menos en el contexto del sector financiero local, el uso de técnicas alternativas como las desarrolladas en el campo de acción del reconocimiento de patrones es escaso o casi nulo. Los esfuerzos de este proyecto se enfocan en explorar la aplicabilidad de tales metodologías sobre datos reales del mercado mexicano, evaluando las bondades y deficiencias encontradas, con la finalidad de sentar bases útiles para futuros desarrollos en la materia. Para fines comparativos con otros trabajos previos, en esta

¹ Algunos de los productos financieros que sufrieron impactos importantes por la mala valoración del riesgo de crédito fueron los respaldados por activos riesgosos (*asset-back securities*) y los derivados de crédito (*credit derivatives*). Las pérdidas que enfrentaron los activos respaldados por hipotecas residenciales llevaron a una crisis económica severa. Como referencia, el lector puede consultar el capítulo “Securitization and the Credit Crisis of 2007” en Hull (2012).

investigación se utilizan conjuntos de datos públicos, sobre los cuales se reportan los resultados obtenidos mediante el uso de los distintos clasificadores seleccionados.

En concreto, son cinco los conjuntos de datos que se emplean para los análisis empíricos. Tres de ellos (Alemán, Australiano y PAKDD) se encuentran disponibles de forma pública mientras que los dos restantes (Cerveza y Autos) son de tipo privado, cuyo origen no se proporciona por razones de confidencialidad. Los conjuntos Alemán y Australiano se encuentran disponibles en el repositorio de UCI (<http://archive.ics.uci.edu/ml/>) y han sido utilizados ampliamente en la literatura como *benchmark* para comparar distintas metodologías de clasificación y modelos de CS. Por su parte, el conjunto PAKDD se puede obtener mediante solicitud en la página de la 13th Pacific-Asia Knowledge Discovery and Data Mining conference (<http://sede.neurotech.com.br:443/PAKDD2009/>). El uso de información pública, la cual también ha sido empleada en otros estudios, obedece a la necesidad de contar con un punto de comparación para los resultados obtenidos en este proyecto.

El conjunto Cerveza contiene información de clientes a quienes una empresa mexicana vendió cerveza (durante 2009 y 2010). Para una proporción de estos clientes, la venta del producto se realizó mediante el otorgamiento de líneas de crédito y se monitoreó el comportamiento de perfil de riesgo. Finalmente, el conjunto Autos contiene información socio-demográfica y de seguimiento de los patrones de pago de una cartera de préstamos para la compra de automóviles, otorgados entre 2006 y 2009 en el mercado local. Con este último conjunto, no sólo es posible analizar el perfil de riesgo de forma estática con los modelos de CS, sino que se sientan las bases para realizar un estudio dinámico en el tiempo como futura línea de investigación.

Como parte de uno de los objetivos planteados, se busca comparar la capacidad predictiva de los distintos clasificadores calibrados sobre los conjuntos de datos anteriormente descritos. Los modelos implementados buscan analizar el perfil de riesgo, tanto de contrapartes que solicitan crédito, como de aquellos clientes previamente aceptados. Las características específicas de cada conjunto de información crediticia disponible permiten orientar el análisis bajo un contexto estático, al tiempo de plantear alternativas para implementar en el futuro estudios del comportamiento dinámico del perfil de riesgo de los clientes a lo largo del periodo de observación.

En este documento se analizan diversas técnicas de clasificación para la construcción de modelos de CS. Entre las técnicas estadísticas empleadas, se encuentran el análisis discriminante, regresión lineal, árboles de clasificación y vecinos más cercanos. Asimismo, se estudian diversos métodos del área de aprendizaje maquina para su aplicación en el contexto de los sistemas de calificación del crédito, entre las que destacan las redes neuronales, máquinas de soporte vectorial, clasificadores bayesianos (*naive bayes*, y redes bayesianas), así como modelos híbridos (*bagging* y *boosting*) y máquinas de Boltzmann restringidas. Hasta donde se tiene conocimiento,

esta última técnica no ha sido utilizada anteriormente en el contexto de CS, por lo que se en esta investigación se sentaría un precedente al respecto.

El documento está integrado por cinco secciones. La primera corresponde a la introducción. En la segunda se presentan distintas técnicas que han sido aplicadas para resolver problemas de calificación crediticia, las cuales provienen tanto del área de la estadística como de la inteligencia artificial y el reconocimiento de patrones. Este apartado provee al lector de una base sobre los fundamentos teóricos y características generales de distintos clasificadores, tomando en cuenta sus bondades y deficiencias como herramientas para la construcción de sistemas de calificación crediticia. En la sección tres se realiza una revisión de las distintas metodologías y métricas más utilizadas, tanto en la industria como en la academia, para evaluar el desempeño de los modelos de clasificación. La discusión se centra principalmente en los enfoques empleados en el área de CS. La sección cuatro incluye un análisis descriptivo de las características de los conjuntos de datos utilizados, así como los resultados de la estimación de cada uno de los modelos calibrados y la comparación de su desempeño en cada caso. Al final del documento se incluye un apéndice que contiene información descriptiva adicional de los conjuntos de datos disponibles.

4.2 Técnicas estadísticas y de aprendizaje maquinal en la literatura

En este capítulo se realiza una revisión sobre los fundamentos de algunos de los principales modelos estadísticos y de reconocimiento de patrones que se han utilizado en la literatura e industria para el problema de clasificación en el contexto del *credit-scoring* (CS).

Para ello, se considera un conjunto $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ de N observaciones (ejemplos), cuyos n atributos están contenidos en la componente n -dimensional $\mathbf{x}_i \in X$ y las etiquetas de clases binarias $y_i \in C \equiv \{0,1\}$.

El proceso de entrenamiento de un clasificador $f: X \rightarrow C$ consiste en configurar los parámetros que lo caracterizan, mediante alguna técnica de estimación o aprendizaje sobre la base de cierto conjunto de entrenamiento D . El resultado del proceso de estimación generalmente es un mapeo² \hat{f} que busca aproximar la regla de correspondencia $\mathbf{x}_i \mapsto y_i, i = 1, \dots, N$.

² La regla de correspondencia definida por f contempla el uso de atributos tanto numéricos como no numéricos contenidos en el conjunto X .

4.2.1 Regresión logística, lineal y análisis cuadrático discriminante

Supóngase que se tienen etiquetas de clases binarias y_i . La estrategia de la regresión logística para clasificación (LOG) consiste en estimar la probabilidad $P(y = 1|x)$ mediante la función sigmoidea

$$P(y = 1|x) = \frac{1}{1 + \exp(-(w_0 + \mathbf{w}^T \mathbf{x}))} \quad (4.1)$$

donde $\mathbf{x} \in \mathbb{R}^n$ es un vector de entrada n -dimensional, \mathbf{w} es el vector de parámetros y el escalar w_0 es el intercepto. Los parámetros w_0 y \mathbf{w} son típicamente estimados utilizando el procedimiento de máxima verosimilitud.

Por su parte, el análisis discriminante asigna una observación \mathbf{x} a la clase $y_i \in \{0,1\}$, teniendo la mayor probabilidad posterior $p(y|x)$. Por el teorema de Bayes se puede realizar el cómputo de la probabilidad posterior como

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})} \quad (4.2)$$

Si se supone que las distribuciones condicionales $p(\mathbf{x}|y)$ son Gaussianas multivariadas se puede verificar que

$$p(\mathbf{x}|y = 1) = \frac{1}{(2\pi)^{n/2} |\Sigma_1|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \right\} \quad (4.3)$$

donde $\boldsymbol{\mu}_1$ es el vector de medias de la clase 1 y Σ_1 denota su matriz de varianza-covarianza. Entonces bajo el criterio de clasificación discriminante, se asigna la clase 1 si

$$\begin{aligned} & (\mathbf{x} - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) - (\mathbf{x} - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) < \\ & 2 \log(P(y = 1)) - 2 \log(P(y = 0)) + \log|\Sigma_0| - \log|\Sigma_1|, \end{aligned} \quad (4.4)$$

y la clase 0 en otro caso. La presencia de términos $\mathbf{x}^T \Sigma_1^{-1} \mathbf{x}$ y $\mathbf{x}^T \Sigma_0^{-1} \mathbf{x}$ indican que el límite de decisión es cuadrático en \mathbf{x} y por lo tanto, esta técnica de clasificación es denominada *análisis discriminante cuadrático (CDA)*. En el caso en que $\Sigma_0 = \Sigma_1 = \Sigma$, los términos cuadráticos $\mathbf{x}^T \Sigma_1^{-1} \mathbf{x}$ y $-\mathbf{x}^T \Sigma_1^{-1} \mathbf{x}$ se cancelan y la regla de clasificación se vuelve lineal en \mathbf{x} . A la técnica de clasificación obtenida comúnmente se le refiere como *análisis discriminante lineal (LDA)*

El primer trabajo publicado sobre la aplicación del análisis discriminante lineal (LDA) para construir un sistema de scoring se remite a Durand (1941) quien mostró que esta metodología podría producir buenas predicciones sobre el repago de los créditos. Posteriormente, otras aplicaciones relacionadas fueron las de Myers y Forgy (1963) quienes compararon los resultados del análisis discriminante y la regresión lineal, Lane (1972), Apilado et al (1974), Taffler y Abassi (1984), y Moses y Liao (1987). La regresión lineal también ha sido utilizada en

los modelos de calificación crediticia. Orgler (1970) utilizó regresión lineal en un modelo de créditos comerciales y Orgler (1971) la utilizó también para construir un scorecard que evaluara exclusivamente el comportamiento del nivel crediticio de los clientes actuales. Otros estudios sobre el uso de modelos de regresión son Fitzpatrick (1976), Lucas (1992), y Henley (1995). Existen otros trabajos relacionados con el estudio del comportamiento de la calidad crediticia de las empresas en función de ciertas razones financieras, como por ejemplo, Altman (1968) que construye la denominada función Z de Altman mediante el uso de LDA para pronosticar con tres años de anticipación la probabilidad de incumplimiento de las empresas. Por otro lado, Kumar y Bhattacharya (2006) comparan el desempeño entre el análisis discriminante lineal y las redes neuronales.

Barniv y McDonald (1999) reportaron 178 artículos en revistas de contabilidad y finanzas que entre 1989 y 1996 utilizaron el modelo logístico. Wiginton (1980) fue uno de los primeros en publicar resultados sobre el uso de la regresión logística en los modelos de credit scoring. A la fecha, esta técnica estadística ha sido ampliamente utilizada en las metodologías de calificación crediticia. Algunos de los trabajos en la materia son: Steenackers y Goovaerts (1989) que propone un modelo de credit scoring para préstamos personales, Platt y Platt (1990) aplicó el análisis logístico para predecir incumplimientos con resultados interesantes en términos del desempeño de la clasificación, Laitinen (1999) pronosticó la insolvencia de 3200 empresas finlandesas utilizando 15 variables tomadas de un conjunto de 35 mediante procedimientos de selección automática bajo modelos de regresión lineales y logísticos, Moody's (2001) aplicó la regresión logística para pronosticar el incumplimiento de 4655 empresas públicas europeas de 26 países.

Algunos estudios han encontrado un desempeño inferior en problemas de clasificación, y específicamente en sus resultados de predicción, comparado con otras técnicas. Estos hallazgos han sido corroborados, por ejemplo, por Caiazza (2004) quien comparó LDA, regresión logística, árboles de clasificación, redes neuronales y algoritmos de lógica difusa, también por Borra y Caiazza (2002) donde se compararon los modelos logísticos y árboles de clasificación utilizando modelos aditivos generalizados (bagging y boosting), y por Galindo y Tamayo (2000) quienes los compararon con árboles de clasificación, redes neuronales y k-vecinos más cercanos para datos de créditos hipotecarios. Una posible explicación de estos pobres resultados es proporcionada por Alfo et al (2005) quienes argumentan que cuando el riesgo de incumplimiento es extremadamente alto o bajo, el modelo logístico presenta fallas y como alternativa proponen extender el modelo bajo estructuras aleatorias que incorporen desviaciones respecto del supuesto de residuales i.i.d. mediante su denominado modelo logístico de efectos aleatorios. Tang et al (2005) compara los modelos logístico y de lógica difusa.

Los modelos paramétricos LDA y regresión logística han sido también aplicados en estudios realizados por bancos centrales en Austria, Francia, Alemania, Italia, Reino Unido (ver Ooghe et al, 1999).

Entre algunos de los primeros trabajos que utilizaron otros tipos de regresión no lineal se encuentra por ejemplo, Grablowsky y Talley (1981) compararon el análisis discriminante y el análisis probit utilizando datos de una cadena departamental de EUA.

4.2.2 Máquinas de soporte vectorial (SVM)

De acuerdo con la formulación original de Vapnik (1998), si se emplean etiquetas de clases binarias $y_i \in \{-1, +1\}$, el clasificador SVM, debe satisfacer las siguientes condiciones:

$$\begin{cases} \mathbf{w}^T \varphi(\mathbf{x}_i) + b \geq +1 & \text{si } y_i = +1 \\ \mathbf{w}^T \varphi(\mathbf{x}_i) + b \leq -1 & \text{si } y_i = -1 \end{cases} \quad (4.5)$$

la cual es equivalente a:

$$y_i[\mathbf{w}^T \varphi(\mathbf{x}_i) + b] \geq 1, \quad i = 1, \dots, N \quad (4.6)$$

La función no lineal $\varphi(\cdot)$ mapea el espacio de atributos en un espacio de dimensión mayor (posiblemente infinita). En este espacio de dimensión superior, las desigualdades presentadas previamente, básicamente construyen un hiper-plano $\mathbf{w}^T \varphi(\mathbf{x}_i) + b = 0$ que discrimina entre las dos clases. En el espacio primal de los pesos, el clasificador adquiere la siguiente forma:

$$y(x) = \text{sign}[\mathbf{w}^T \varphi(\mathbf{x}_i) + b], \quad (4.7)$$

Esencialmente, los clasificadores SVM utilizan modelos lineales para establecer regiones que separen clases de manera no lineal. La estrategia que utilizan consiste en transformar el espacio original de los atributos de entrada mediante el mapeo no lineal φ , de manera que en el nuevo espacio las clases queden separadas por un hiperplano. En consecuencia, las regiones de separación en el espacio original son no lineales.

En la Figura 4.1 se presenta un ejemplo de un conjunto linealmente separable en el contexto de credit scoring respecto de dos atributos dados.

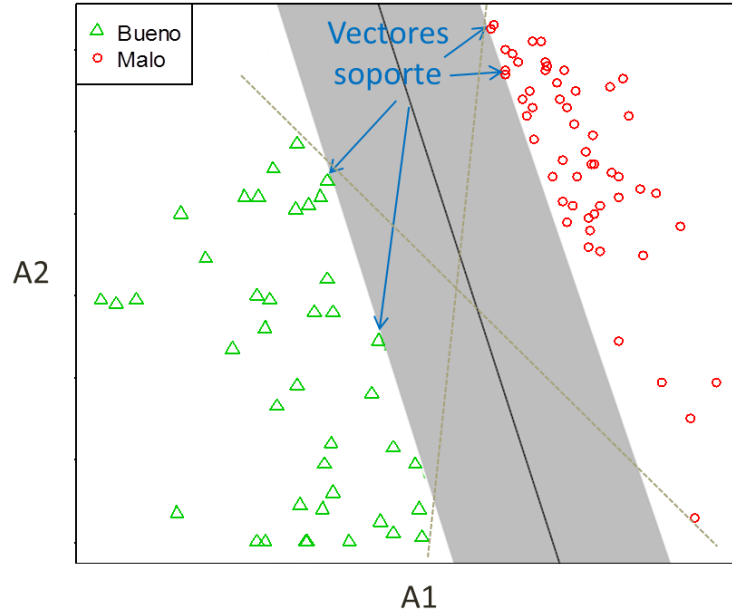


Figura 4.1. Conjunto linealmente separable bajo SVM (dos atributos A1 y A2; dos clases)

No obstante la existencia de una regla de separación dada por la ecuación (4.7), la evaluación no se realiza de esta manera. En cambio, define el problema de optimización convexa.

$$\min_{\mathbf{w}, b, \xi} J(\mathbf{w}, b, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \quad (4.8)$$

sujeto a

$$\begin{cases} y_i [\mathbf{w}^T \varphi(\mathbf{x}_i) + b] \geq 1, & i = 1, \dots, N \\ \xi_i \geq 0 & i = 1, \dots, N \end{cases} \quad (4.9)$$

Las variables de holgura ξ_i se utilizan para permitir malas clasificaciones en el conjunto de desigualdades (por ejemplo, debido a distribuciones traslapadas). La primera parte de la función objetivo busca maximizar el margen entre ambas clases en el nuevo espacio, mientras que la segunda parte minimiza el error incurrido por malas calificaciones. La constante $C > 0$ se considera como un parámetro de refinamiento en el algoritmo. Hay que notar que esta formulación esta cercanamente relacionada con el problema a resolver bajo el método PL. Entre las diferencias más importantes se encuentra la introducción de un término de regularización $\frac{1}{2} \mathbf{w}^T \mathbf{w}$ grande en la función objetivo, asimismo, considera un margen para separar las clases y permite cotas de decisión no lineales mediante el mapeo $\varphi(\cdot)$.

El lagrangiano del problema de optimización (10) y (11) con restricciones está dada por

$$\mathcal{L}(\mathbf{w}, b, \xi, \alpha, \nu) = J(\mathbf{w}, b, \xi) - \sum_{i=1}^N \alpha_i \{y_i [\mathbf{w}^T \varphi(\mathbf{x}_i) + b] - 1 + \xi_i\} - \sum_{i=1}^N \nu_i \xi_i \quad (4.10)$$

La solución al problema de optimización presentado arriba, está dado por el punto silla del lagrangiano, el cual se obtiene al minimizar (4.10) con respecto de \mathbf{w}, b, ξ y maximizándolo con respecto a $\boldsymbol{\alpha}$ y \mathbf{v} . Para resolver el problema

$$\max_{\boldsymbol{\alpha}, \mathbf{v}} \min_{\mathbf{w}, b, \xi} \mathcal{L}(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \mathbf{v}) \quad (4.11)$$

se obtiene:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 & \rightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i y_i \boldsymbol{\varphi}(x_i) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 & \rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \\ \frac{\partial \mathcal{L}}{\partial \xi_i} = 0 & \rightarrow 0 \leq \alpha_i \leq C, i = 1, \dots, N \end{cases} \quad (4.12)$$

Al sustituir la primera expresión en $y(x) = \text{sign}[\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b]$, el clasificador resultante quedaría dado por

$$y(x) = \text{sign}[\sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b] \quad (4.13)$$

donde $K(\mathbf{x}_i, \mathbf{x}) = \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x})$ constituye un núcleo definido positivo que satisface el teorema de Mercer. Los multiplicadores de Lagrange α_i son entonces determinados mediante el siguiente problema de optimización (problema dual):

$$\max_{\alpha_i} -\frac{1}{2} \sum_{i,j=1}^N y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \alpha_i \alpha_j + \sum_{i=1}^N \alpha_i \quad (4.14)$$

sujeto a

$$\begin{cases} \sum_{i=1}^N \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, \quad i = 1, \dots, N \end{cases} \quad (4.15)$$

La construcción del clasificador SVM ahora se reduce a resolver un problema de programación cuadrática convexa en α_i . Como primer aspecto, conviene notar que no será necesario calcular \mathbf{w} o $\boldsymbol{\varphi}(x_i)$ para determinar la superficie de decisión, por ello, no será necesaria la construcción de un mapeo no lineal $\boldsymbol{\varphi}(\mathbf{x})$. Como alternativa, se emplea la función núcleo K . Entre las formas funcionales más comunes para K se tienen la función de base radial (RBF³) $K(\mathbf{x}_i, \mathbf{x}) = \exp\{-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma^2\}$ y el núcleo lineal $K(\mathbf{x}_i, \mathbf{x}) = \mathbf{x}_i^T \mathbf{x}_j$. Además de los clasificadores SVM

³ El operador $\|\cdot\|$ representa la norma euclidiana, la cual se define como $\|\mathbf{x}\| = \sqrt{x_1^2 + \dots + x_n^2}$ para todo vector $\mathbf{x} \in \mathbb{R}^n$.

estándar, en la literatura se han propuesto los clasificadores LS-SVM, los cuales son una versión modificada de los SVMs sugeridos por Suykens y Vandewalle (1999) y Suykens et al. (2002). LS-SVM utiliza una función de costos de mínimos cuadrados y reemplazan las restricciones de desigualdad con restricciones de igualdad para resolver el problema

$$\min_{\mathbf{w}, b, \xi} \mathcal{J}(\mathbf{w}, b, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2 \quad (4.16)$$

sujeito a las restricciones de igualdad:

$$y_i[\mathbf{w}^T \varphi(\mathbf{x}_i) + b] = 1 - e_i, \quad i = 1, \dots, N \quad (4.17)$$

Siguiendo la misma estrategia de solución que con el clasificador SVM estándar, se puede verificar que el clasificador LS-SVM se puede obtener como la solución al sistema lineal de ecuaciones:

$$\left(\begin{array}{c|c} 0 & \mathbf{y}^T \\ \hline - & - \\ \mathbf{y} & \Omega + \gamma^{-1} \mathbf{I} \end{array} \right) \begin{pmatrix} b \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix} \quad (4.18)$$

donde $\mathbf{y} = [y_1; \dots; y_N]$, $\mathbf{1} = [1; \dots; 1]$, $\mathbf{e} = [e_1; \dots; e_N]$, $\boldsymbol{\alpha} = [\alpha_1; \dots; \alpha_N]$. Entonces el teorema de Mercer se puede aplicar directamente a la matriz Ω , definida por $\Omega_{ij} = y_i y_j \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$.

Distintas especificaciones de SVM han sido utilizadas para problemas de clasificación y regresión debido a su buen desempeño en diversas aplicaciones. Recientemente, varios estudios han utilizado SVM en credit scoring con resultados alentadores. El problema de la predicción del incumplimiento de empresas también ha sido abordado aplicando redes neuronales (NN) y SVM en Fan y Palaniswami (2000), Atiya (2001), Shin et al. (2005) y Min et al. (2005). Li et al (2006) desarrolló un modelo de credit scoring para aplicaciones de créditos al consumo identificando mejoras en la generalización respecto a modelos de NN. Para solucionar el problema sobre la interpretación de los resultados de clasificación de SVM, Martens (2007) introdujo métodos de extracción de reglas para SVMs que pueden generar modelos de credit scoring con reglas interpretables y con poca pérdida en la exactitud.

En un estudio comparativo de 17 clasificadores aplicados en riesgo credit scoring, Baesens et al (2003b) analizaron 8 conjuntos de datos de créditos, entre los que se encuentran algunos portafolios de instituciones financieras del Reino Unido, Bélgica, Holanda y Luxemburgo. Los métodos utilizados fueron: regresión lineal y cuadrática, regresión logística, programación lineal, cuatro variantes de SVM, cuatro clasificadores basados en árboles de decisión, dos variantes de vecinos más cercanos, NN, naive Bayes, y árboles bayesianos

aumentados bajo naive Bayes. Los resultados indicaron que los métodos con mejor desempeño fueron SVM con base radial y la versión por mínimos cuadrados, así como también los clasificadores basados en NN.

4.2.3 Redes neuronales

Las redes neuronales (NN) son representaciones matemáticas inspiradas en el funcionamiento del cerebro humano. Muchos tipos de redes neuronales han sido sugeridas en la literatura. Entre los ejemplos de NN más populares se encuentra el denominado perceptrón multicapa (MLP).

Un MLP está compuesto típicamente por una capa de entrada, una o varias capas ocultas y una capa de salida, cada una consistiendo de cierto número de neuronas. Cada neurona procesa sus entradas y genera un valor de salida que es transmitido a las neuronas en la capa subsecuente.

En la siguiente figura se muestra la configuración típica en el contexto del credit scoring, para un MLP con una capa oculta y una capa de salida formada por una sola neurona. En la capa de entrada se incorporan n atributos (como los que usualmente son requeridos en las solicitudes de créditos personales).

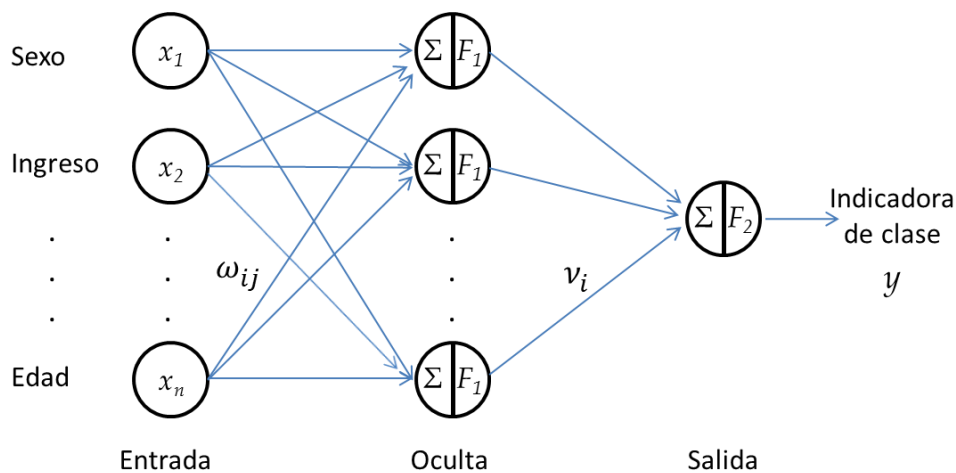


Figura 4.2. Perceptrón multicapa con una capa oculta

La salida h_i para cada neurona oculta i se obtiene de ponderar cada una de sus entradas (atributos) más un término de tendencia o sesgo $b_i^{(1)}$ como se muestra a continuación:

$$h_i = F_1\left(b_i^{(1)} + \sum_{j=1}^n \omega_{ij} x_j\right) \quad (4.19)$$

donde ω_{ij} denota el conector de entrada de pesos j hacia la unidad oculta i . Por su parte, como resultado de la capa de salida se construye la indicadora de clase y de la siguiente manera:

$$y = F_2(b^{(2)} + \sum_{i=1}^{n_h} v_i h_i) \quad (4.20)$$

donde n_h es el número de neuronas ocultas, v_i representa el peso conector de la neurona oculta i hacia la neurona de salida y $b^{(2)}$ el término de sesgo correspondiente. Las entradas de sesgo juegan un rol similar al término de intercepto en un modelo de regresión lineal clásico.

Las funciones F_1 y F_2 que aparecen en las ecuaciones (4.19) y (4.20) se denominan funciones de transición o transferencia, las cuales permiten modelar relaciones no lineales entre los datos. Entre las funciones de transferencia más comúnmente utilizadas se encuentran: la logística o sigmoidea $F(x) = \frac{1}{1+e^{-x}}$, la tangente hiperbólica $F(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ y la función lineal $F(x) = x$.

En el caso de un problema de clasificación binaria, es conveniente utilizar la función de transferencia logística en la capa de salida (F_2), ya que esta salida se limita a un valor entre el rango $[0,1]$ ⁴. Esto permite que la salida y de la MLP sea interpretada por medio de probabilidad condicional. Finalmente, una función umbral es típicamente aplicada para mapear la red de salida y a una etiqueta de clasificación (por ejemplo, bueno/malo, en el contexto del credit scoring).

Nótese que múltiples capas ocultas pueden ser utilizadas pero diversos resultados teóricos han demostrado que las NN con tan una sola capa oculta son capaces de aproximar de manera universal cualquier función continua para cualquier grado de precisión deseado en un intervalo compacto.

Los pesos ω_{ij} y v_i , así como los coeficientes de sesgo $b_i^{(1)}$ y $b^{(2)}$, son parámetros cruciales de la red y necesitan ser estimados durante un proceso de aprendizaje. Al respecto muchos algoritmos han sido sugeridos en la literatura para ello, entre los que destaca el algoritmo de retro-propagación. Como cualquier otro esquema de clasificación, los MLP entrenados con retro-propagación pueden sufrir de un excesivo sobre-ajuste en el conjunto de entrenamiento, específicamente si la red es mucho grande que la estructura necesaria para describir el problema en estudio. Entre las estrategias sugeridas para aminorar este efecto se encuentra: evaluar la tasa de error de clasificación para cierto subconjunto de datos y detener anticipadamente (*early stopping*) el proceso de estimación cuando ésta empiece a disminuir a partir de cierto umbral; el uso de un factor de decaimiento en el que la función de error se penaliza por la suma de cuadrados de los pesos estimados en la red, con la intención de limitar la influencia de conexiones irrelevantes entre neuronas. El autor puede consultar Witten et al. (2011) para profundizar con mayor detalle en las estrategias de calibración de los parámetros de NNs.

⁴ Básicamente, cualquier función de distribución puede ser utilizada como función de transferencia para la capa de salida.

Una desventaja importante del MLP es que contiene unidades internas que esencialmente son poco transparentes. No obstante que existen diversas técnicas para extraer reglas de decisión a partir de redes neuronales entrenadas, no es claro si éstas ofrecen alguna ventaja con respecto a clasificadores estándar que inducen conjuntos de reglas directamente de los datos y cuyo entrenamiento es más rápido.

De acuerdo con Vellido et al (1999) más del 75% de las aplicaciones de NN en el contexto de credit scoring están relacionadas con el uso de perceptrones multicapa con propagación hacia adelante (MLP) y de retropropagación (BP).

Jensen (1992) aplica MLP para pronosticar el comportamiento de pago de aplicaciones de tarjetas de crédito y lo comparó contra un modelo de credit scoring comercial; West (2000) comparó el desempeño de cinco técnicas de NN (MLP, mezcla de expertos, función radial, vector de cuantización y resonancia adaptativa difusa) en los datos de créditos Alemanes y Australianos; por su parte, West et al (2005) incrementa la capacidad del modelo anterior mediante las estrategias de validación cruzada, bagging y boosting.

No obstante las altas tasas de clasificación observadas en las técnicas de NN, se les critica por la ausencia de variables explicativas sobre la clasificación de los créditos y su transparencia explicativa (Baesens et al, 2003a). Algunos de los métodos propuestos en la literatura para solucionar este problema son la combinación con otros métodos (hibridación), extracción de reglas de las redes entrenadas y métodos de agrupamiento. Sobre esta última estrategia, Huysmans et al (2006) explotó la capacidad de visualización y análisis exploratorio de los mapas auto organizados (SOM) en el credit scoring y determinó un mejor desempeño cuando se utilizaba para elevar el poder predictivo de cada neurona en un sistema de NNs. En lo relativo a la extracción de reglas, Baesens et al (2003a) aplicó tres metodologías para la extracción de reglas que simulen el proceso de decisión de NNs entrenadas (NeuroRule, Trepan y Nefclass) para clasificar clientes buenos y malos de tres distintos conjuntos de créditos, realizando una comparación contra el algoritmo C4.5 y regresión logística. Asimismo, Arns et al (2006) utilizaron la técnica de NeuroRule para extraer reglas de NNs multicapa entrenadas para modelar el perfil de riesgo de portafolio de créditos a pequeñas y medianas empresas otorgados por un banco brasileño.

4.2.4 Clasificadores Bayesianos y redes bayesianas

Un clasificador simple que en la práctica suele desempeñarse de manera sorprendentemente aceptable en distintos contextos es el clasificador *naive Bayes*. Este clasificador básicamente aprende las probabilidades condicionales de las clases $p(x_i|y)$ de cada entrada x_i , $i = 1, \dots, n$ dada una etiqueta de clase y . Cada ejemplo o instancia se clasifica utilizando la regla de Bayes para calcular la probabilidad posterior de cada clase y dado el vector \mathbf{x} de los valores observados de sus atributos:

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})}$$

Estos clasificadores trabajan bajo el supuesto de que los atributos son condicionalmente independientes dada la etiqueta de clase, por lo tanto,

$$p(\mathbf{x}|y) = \prod_{i=1}^n p(x_i|y) \quad (4.21)$$

La ecuación (4.21) constituye una simplificación en la estructura de dependencia de los atributos, la cual facilita enormemente los cálculos. Las probabilidades $p(x_i|y)$ son estimadas usando los conteos de frecuencias para los atributos discretos, y un método bajo la distribución normal o una densidad (kernel) de suavizamiento para los atributos continuos. Los clasificadores obtenidos bajo el supuesto de la ecuación (4.21) se denominan clasificadores de Bayes simples (*naive Bayes*).

Friedman et al. (1997) introdujeron clasificadores de Bayes bajo un contexto de árbol aumentado (TAN) como una extensión de los clasificadores de Bayes simples. Los TANs dispensan el supuesto de independencia al permitir dependencias con estructura arbórea entre los atributos. Una dependencia desde x_i hasta x_j implica que el impacto de x_i sobre la variable de clase también depende del valor de x_j . Estos clasificadores pertenecen a un grupo más grande denominado redes bayesianas.

Las redes bayesianas (BN) pertenecen a la familia de modelos gráficos probabilísticos, los cuales son utilizados para representar relaciones entre variables bajo contextos específicos. Específicamente, las BN corresponden a una estructura conocida como gráfica acíclica dirigida (DAG), cuya estructura es definida por dos conjuntos: el conjunto de nodos (vértices) y el conjunto de arcos dirigidos. Los nodos representan variables aleatorias (v.a.) las cuales son representadas por círculos y los arcos dirigidos que representan la dependencia directa entre las v.a. son dibujados como flechas entre los nodos. Cuando el arco es dibujado del nodo X_i (padre) al nodo X_j (hijo), la estructura indica que la v.a. X_j depende de X_i .

La estructura acíclica de la gráfica garantiza que ningún nodo puede ser su propio antecesor o su propio sucesor, de manera que la probabilidad condicional de todas las variables puede escribirse de forma compacta. No obstante que las flechas que gráficamente unen nodos representan una conexión causal entre las variables, el proceso de razonamiento de las BN puede operar propagando información en cualquier dirección. La independencia condicional de cada variable con todos aquellos nodos que no sean sus descendientes, dado el estado de sus padres, reduce significativamente el número de parámetros necesarios para caracterizar la distribución conjunta de las variables, al tiempo de proveer una manera eficiente de calcular probabilidades posteriores.

Adicional a la estructura DAG, para caracterizar completamente a una BN, se requiere especificar los parámetros del modelo de manera tal que se cumple la propiedad de Markov y en

consecuencia la probabilidad condicional en cada nodo depende solamente de sus padres. Para el caso de v.a. discretas, dicha probabilidad condicional se representa mediante una tabla que enlista, para cada combinación de los valores de sus padres, la probabilidad (local) de que un nodo hijo tome sus valores posibles.

De acuerdo con Friedman et al (1997), una red Bayesiana B es una gráfica acíclica que representa probabilidades condicionales sobre un conjunto de v.a. $U = \{Z_1, Z_2, \dots, Z_m\}$. La red queda definida por el par $B = \langle G, \Theta \rangle$, donde G es la DAG cuyos nodos son los elementos en U y la estructura de sus arcos representan la dependencias directas entre las v.a. Z_i . Por su parte, Θ denota al conjunto de parámetros de la red formado por $\theta_{z_i|\pi_i} = P_B(z_i|\pi_i)$ para cada realización z_i de Z_i condicionada sobre π_i , el conjunto de padres de Z_i en G . Si Z_i no tiene padres, se dice que su distribución de probabilidades locales es incondicional, y en cualquier otro caso, se denomina condicional. Entonces

$$P_B(U) = \prod_{i=1}^m \theta_{z_i|\pi_i} \quad (4.22)$$

La ecuación (4.22) permite calcular la distribución conjunta de U de manera más compacta que si la comparamos con la regla de la cadena inducida por la definición de probabilidad condicional

$$P_B(U) = \prod_{i=1}^m P_B(Z_i|Z_{i+1}, \dots, Z_m), \quad (4.23)$$

donde el último término del producto correspondiente al índice $i = m$ es $P_B(Z_m)$.

La factorización (4.22) ofrece ventajas para la inferencia, aprendizaje (estimación de parámetros) y también desde el punto de vista computacional. De esta forma, cualquier probabilidad conjunta o condicional de interés sobre cualesquier subconjunto de v.a. no observables se puede calcular sumando/integrando a (4.22) sobre todos los valores posibles (observados) del complemento del subconjunto de interés.

Siguiendo la notación introducida al inicio del capítulo, se dispone de un conjunto de datos $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$, donde cada y_i y \mathbf{x}_i representan realizaciones de la v.a. de clase Y y los vectores aleatorios $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$, respectivamente. Para efectos prácticos, se supone que todas las X_j son v.a. discretas con soporte finito. El proceso de aprendizaje consiste en encontrar una red apropiada B dado el conjunto D .

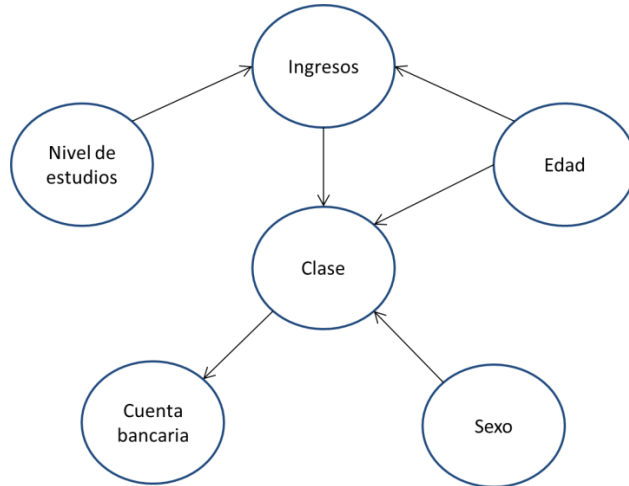


Figura 4.2. Ejemplo de una red bayesiana en el contexto de CS.

En la Figura 4.2 se muestra un ejemplo de BN bajo el contexto de CS. El gráfico indica que, dado el nivel de ingresos del cliente, el perfil de riesgo (variable Clase) y el nivel de estudios son condicionalmente independientes. Similarmente, cuando el perfil de riesgo es conocido, el atributo cuenta bancaria es condicionalmente independiente del resto del resto de sus antecesores: nivel de estudios, ingresos, edad y sexo. Sin embargo, el riesgo de un cliente depende de manera causal directa de su nivel de ingresos, edad y sexo. Por su parte, los atributos nivel de estudios, edad y sexo son marginalmente independientes. Sin embargo, cuando los ingresos son conocidos, los dos primeros resultan ser condicionalmente dependientes, al tiempo que cuando la clase es dada, los dos últimos son también condicionalmente dependientes.

Si consideramos los mismos atributos y variable de clase de la Figura 4.2, el clasificador *naive Bayes* para el problema de clasificación correspondiente, tendría la estructura presentada en la Figura 4.3. Como se puede observar, los arcos y nodos del gráfico satisfacen la ecuación (4.22).

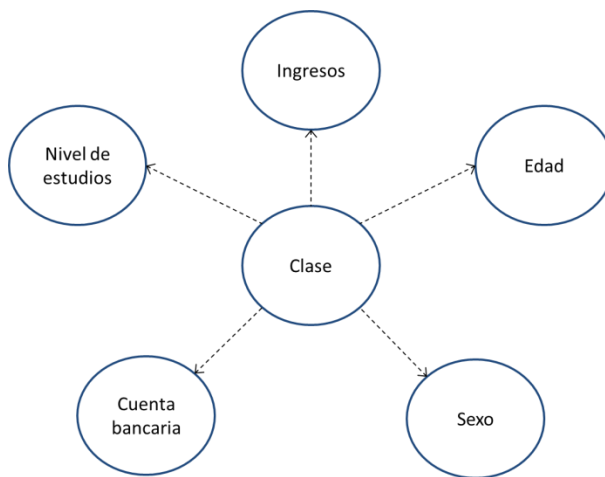


Figura 4.3. Ejemplo de la estructura de un clasificador *naive Bayes*.

Para utilizar una BN como clasificador, simplemente es necesario calcular $\operatorname{argmax}_y P_B(y|\mathbf{x})$ empleando (4.22), la distribución P_B inducida por la red.

Dado que

$$P_B(Y|\mathbf{X}) = P_B(U)/P_B(\mathbf{X}) \propto P_B(U), \quad (4.24)$$

donde $U = \{Y, \mathbf{X}\}$, no es necesario complicar el proceso de inferencia que resulte de calcular $P_B(\mathbf{X})$.

La naturaleza dual de las BN divide a su proceso de aprendizaje en dos etapas. Primero se estima una estructura de red G , y enseguida se estiman las tablas de probabilidad, es decir, los parámetros Θ .

Algunas aproximaciones para realizar el aprendizaje de la estructura son:

- 1) Métricas de puntaje local. Se emplea alguna métrica de calidad $Q(B|D)$ la cual se desea maximizar. Entre las métricas utilizadas se encuentran aquellas construidas con aproximaciones bayesianas, verosimilitud, criterio de información de Akaike (AIC), longitud de descripción mínima (MDL), entropía y otros criterios de información. La propiedad de descomposición en la suma de puntajes individuales para cada nodo que presentan dichas métricas, permite el uso de métodos de búsqueda local.
- 2) Pruebas de independencia local. Este enfoque supone que existe una estructura de red que representa de manera exacta las independencias distribucionales de las v.a. subyacentes a los datos. Si se identifica independencia condicional entre dos nodos, esto indica que no es necesario conectarlos mediante algún arco. Luego de identificar las posiciones de los arcos, la dirección de los mismos se establece de manera que las independencias identificadas.
- 3) Métricas de puntaje global. El desempeño de la BN sobre cierto conjunto de datos se mide estimando alguna métrica de utilidad esperada. Por ejemplo, se puede estimar la tasa de error de la red, o alguna otra medida de precisión, mediante validación cruzada⁵. A diferencia de las métricas de puntaje local, en este caso es difícil realizar una descomposición en puntajes individuales para cada nodo.

Entre los algoritmos de búsqueda comúnmente utilizados para implementar los enfoques mencionados arriba, se encuentran *Hill climbing*, recocido simulado (*Simulated Annealing*), búsqueda tabú, algoritmos genéticos, simulación MCMC (*Markov Chain Monte Carlo*), TAN (*Tree Augmented Naive Bayes*) entre otros.

⁵ En el capítulo 4 se describen distintas alternativas para medir el desempeño de los modelos y métricas posibles, las cuales pueden ser empleadas también en el contexto de BN.

Una vez que se ha determinado una estructura para la BN, la cual por cierto, también pudiera suponerse preestablecida, es necesario llevar a cabo la estimación de las tablas de probabilidad condicional para cada nodo.

Un algoritmo simple para el aprendizaje de las BN es el denominado K2, el cual inicia con un orden preestablecido de los atributos. Se procesa cada nodo considerado y gradualmente considera la incorporación de arcos conectando nodos previamente procesados con el actual. En cada iteración se incorpora el arco que maximice el puntaje de la red.

En el caso en que la estructura de la BN tiene como único padre del resto de los nodos a la clase, el clasificador resultante es el naive Bayes, pues en este caso, todos los atributos son (condicionalmente) independientes dada la clase. Una extensión al clasificador simple de bayesiano propuesta por Friedman et al (1997) es el denominado Tree Augmented Naive Bayes (TAN), los cuales relajan el supuesto de independencia condicional permitiendo la existencia de arcos entre los atributos. En las redes TAN, la variable de clase no tiene padres y cada atributo tiene como padre a la clase y a lo más alguno de los otros atributos. En consecuencia, los atributos forman una estructura de árbol.

El proceso de aprendizaje de las BN requiere un alto costo de conteo. Para cada estructura de red considerada en el algoritmo de búsqueda, los datos deben ser leídos en cada iteración para la construcción de las tablas de probabilidad condicional. Como alternativa, se ha propuesto almacenarlas en una estructura de datos en vez de eliminarlas en cada paso. Una estructura conocida es el árbol de todas dimensiones (AD tree).

El texto seminal de Pearl (1988) estableció a las redes bayesianas como un campo de estudio al consolidar sus propiedades. Este tipo de redes representan un conjunto de clasificadores que modelan las relaciones estructurales de las variables. Davis et al (1992) investigó mecanismos bajos sistemas expertos bayesianos podrían ayudar al problema de clasificación de tarjetas de crédito de Bank of Scotland y comparó la aproximación con ANNs. Baesens et al (2002) evalúa la aplicabilidad de varios tipos de BNs en credit scoring, entre ellos, el clasificador Naive Bayes (NB), árboles aumentados bajo NB, y una BN irrestricta con entrenamiento Monte Carlo via Cadenas de Markov (MCMC). En un ejercicio empírico sobre conjunto de datos alemanes de UCI, este último clasificador tuvo el mejor desempeño comparado con el resto y el algoritmo C4.5. Por su parte, Li y Guo (2006) determinan capacidades similares entre tres modelos de clasificación basados en NB y cinco tipo de ANNs, sobre dos conjuntos de datos de créditos reales. Más recientemente, Antonakis y Sfakianakis (2009) comparan modelos de credit scoring construidos sobre la NBs contra LDA, regresiones logísticas, vecinos más cercanos, árboles de clasificación y ANNs. Los autores utilizan dos conjuntos de datos, uno con créditos de un banco griego y el segundo formado por los créditos australianos de UCI, encontrando un poder predictivo inferior del clasificador NB respecto a los otros métodos.

4.2.5 Árboles de decisión

Muchos algoritmos con árboles de decisión han sido sugeridos previamente en la literatura. Uno de los más populares es el algoritmo C4.5, el cual induce árboles de decisión basados en los conceptos de entropía y ganancia de información. Si se denota por p_c a la proporción de ejemplos de clase $c \in \{0,1\}$ en la muestra $S \subset D$. La entropía de S se define como

$$\text{Entropía}(S) = -p_1 \log_2(p_1) - p_0 \log_2(p_0)$$

donde $p_0 = 1 - p_1$. El valor máximo que la entropía de una muestra puede tomar es 1 (cuando $p_1 = p_0 = 0.5$) en tanto que su valor mínimo lo alcanza en 0 (cuando $p_1 = 0$ o $p_0 = 0$).

Se define la ganancia información de S con respecto al atributo x_i como la reducción en entropía por filtrar (dividir) los ejemplos sobre todos los valores de x_i

$$\text{Ganancia}(S, x_i) = \text{Entropía}(S) - \sum_{v \in \text{valores}(x_i)} \frac{|S_v|}{|S|} \text{Entropía}(S_v) \quad (4.25)$$

donde S_v representa una submuestra de S en la cual el atributo x_i tiene un valor específico v . Cuando el criterio (27) es empleado para decidir hasta qué nodo dividir, el algoritmo favorece la división sobre atributos con muchos valores distintos. Por lo tanto, cuando un conjunto de datos contiene un atributo con un valor distinto para cada ejemplo (como por ejemplo, el caso de un campo identificador), el criterio de ganancia de información lo seleccionará como el mejor criterio de división. Para solucionar esto, el algoritmo C4.5 aplica una normalización y emplea como métrica la razón de ganancia definida como

$$\text{RazonGanancia}(S, x_i) = \frac{\text{Ganancia}(S, x_i)}{\text{Div}(S, x_i)}, \quad (4.26)$$

con

$$\text{Div}(S, x_i) = - \sum_{k \in \text{valores}(x_i)} \frac{|S_k|}{|S|} \log_2 \frac{|S_k|}{|S|}, \quad (4.27)$$

donde S_k representa una submuestra de S donde el atributo x_i tiene un valor específico k . La cantidad (4.27) representa la entropía de S con respecto de los valores de x_i . El uso del criterio (4.26) indica que C4.5 favorece divisiones sobre atributos que producen la mayor razón de ganancia bajo la restricción adicional de que la ganancia de información debe ser al menos tan grande como la ganancia promedio sobre todas las divisiones consideradas. El árbol entonces es construido por medio de partición recursiva. Esta estrategia de crecimiento en forma de árbol resulta entonces en un árbol complejo con muchos nodos intermedios que sobre ajustan a los datos. El algoritmo C4.5 plantea remediar esta situación por medio de un procedimiento de “poda”, el cual se ejecuta retrospectivamente una vez que todo el árbol ha terminado de crecer. Un árbol de C4.5 “no podado” se puede traducir fácilmente en un conjunto de reglas mediante la construcción de reglas individuales para cada camino que vaya desde la raíz del árbol “no

podado” hasta un nodo final (“hoja”). Estas reglas pueden ser posteriormente “podadas” al remover las condiciones basadas en un procedimiento similar al del árbol.

Una de los principales atractivos de los árboles de clasificación es su traducción directa en reglas de decisión que resultan fácilmente entendibles para usuarios que pudieran no estar muy familiarizados con el tema de técnicas de clasificación. Sin embargo, una de sus desventajas más importantes es su inestabilidad, ya que pequeños cambios en los datos pueden llevar a desviaciones importantes en las estimaciones, al tiempo que pueden sobre ajustar los datos si no se utiliza validación cruzada y poda.

Los árboles de clasificación se han desarrollado en varias disciplinas, entre las que destacan las ciencias de la vida, la estadística y la inteligencia artificial. La idea del uso de árboles de decisión para resolver problemas de clasificación general fue desarrollada de forma independiente por Breiman y Friedman en 1973. La selección de los atributos para la subdivisión en cada nodo interno es extremadamente importante en el proceso de construcción del árbol de decisión y determina en gran medida la estructura final del mismo. Una de las referencias más importantes en estadística es Breiman et al (1984) quienes utilizaron como criterio de selección el índice de Gini e introdujeron el término el concepto de árbol de regresión y clasificación (CART). El primer algoritmo para generar un árbol de decisión (ID3) es introducido por Quinlan (1979), posteriormente el C4 de Quinlan (1986) y el C4.5 en Quinlan (1993). Quinlan utilizó el concepto ganancia de información tomando como base una medida de entropía para desarrollar estos algoritmos. Autores como Kass (1980) y Biggs y Ville (1991) utilizaron una prueba ji-cuadrada para realizar particiones multinivel sobre datos categóricos con su metodología CHAID.

El uso de los árboles de clasificación en credit scoring tuvo sus inicios con Makowski (1985) y Coffman (1986) y su uso en la industria continua dada su capacidad de identificar patrones de comportamiento entre los atributos y categorías de clientes ya sea extremadamente riesgosos o con perfil de riesgo muy bajo. Anderson (2007) destaca su transparencia y facilidad de implementación para árboles simples pero critica su poca efectividad en capacidad predictiva. Sin embargo, destaca su uso para la exploración rápida de los datos, ya sea para identificar variables predictivas clave o bien para funcionar como benchmark para otras metodologías.

Algunas referencias más recientes en credit scoring han utilizado como base de sus metodologías diversos árboles de clasificación, por ejemplo, Mues et al (2004) utilizan diagramas de decisión (una generalización de los árboles de decisión) y explotan su uso como herramienta de descripción visual compacta de los datos, Lee et al (2006) identificó mejores resultados con CART y regresión multivariada adaptativa por splines (MARS) que alternativas como LDA, regresión logística, ANNs y SVMs para datos de tarjeta de crédito., Zhao (2007) aplica técnicas de programación genética multi-objetivo para desarrollar árboles de decisión óptimos en el sentido de Pareto, Bastos (2008) implementa árboles de decisión bajo boosting y muestra para datos de tarjeta de crédito que la metodología ofrece resultados con desempeño similar a

preceptrón multicapa y SVM. Por su parte, Li et al (2010) aplica CART para analizar el comportamiento de fallas de negocio en compañías públicas listada en bolsa de China.

4.2.6 Clasificadores de k vecinos más cercanos

El método de k -vecinos más cercanos (KNN) fue aplicado por primera vez en el área de CS por Chatterjee y Barcun (1970) y posteriormente por Henley y Hand (1996). La estrategia de este clasificador consiste en elegir una métrica sobre el espacio del conjunto de datos al que se desea aplicar para medir qué tan distante se encuentra un ejemplo de otro. Con base en un conjunto de entrenamiento dado, un nuevo ejemplo es clasificado con buen o mal perfil de riesgo dependiendo de las proporciones de buenos y malos que mantengan los k -vecinos más cercanos al nuevo ejemplo. En consecuencia, esta técnica de clasificación depende básicamente de tres parámetros: la métrica, el número de instancias k que constituyen el conjunto de vecinos más cercanos, y la proporción de buenos que éstos debieran mantener para que el nuevo ejemplo sea clasificado como bueno.

La selección de la métrica es crucial. Fukunaga y Flick (1984) introdujeron una métrica general de la forma

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j) \mathbf{A}(\mathbf{x}_i) \left((\mathbf{x}_i - \mathbf{x}_j)^T \right)^{\frac{1}{2}} \quad (4.28)$$

donde $\mathbf{A}(\mathbf{x})$ es una matriz de $p \times p$ simétrica definida positiva. $\mathbf{A}(\mathbf{x})$ se llama métrica local si depende de \mathbf{x} , y se denomina métrica global si es independiente de \mathbf{x} . Como la métrica local toma características específicas del conjunto de entrenamiento que pueden no ser apropiadas en general, muchos autores utilizan métricas globales. Henley y Hand (1996) utilizaron mezclas de distancias euclidianas para separar clientes buenos de malos. Específicamente, estos autores sugieren una métrica de la forma

$$d(\mathbf{x}_i, \mathbf{x}_j) = \left((\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{I} + D \mathbf{w} \cdot \mathbf{w}^T) (\mathbf{x}_i - \mathbf{x}_j) \right)^{\frac{1}{2}} \quad (4.29)$$

donde \mathbf{I} es la matriz identidad. La elección de los parámetros k y D se realiza mediante validación cruzada. En la literatura se han propuesto también medidas avanzadas de distancia para mejorar el desempeño de estos clasificadores.

El clasificador de vecinos más cercanos, aunque no es tan ampliamente utilizado en el área de CS como la regresión logística y el discriminante lineal, representa una alternativa útil. Probablemente una de las desventajas del método, desde el punto de vista práctico, es que esta técnica no permite construir un score para las características particulares de cada cliente.

Destacan varias aplicaciones posteriores de método de KNN, entre las que enlistamos, a Henley y Hand (1996) quienes realizaron una investigación detallada de métodos de vecinos más cercanos para datos de una compañía de ventas por correo al tiempo que examinaron la selección de la métrica para definir la propiedad de cercanía y la selección del número de vecinos más cercanos a considerar; Paredes y Vidal (2000) evalúan la mejora en el desempeño del clasificador de vecinos más cercanos con la inclusión de una medida de disimilitud ponderada en varios conjuntos de datos, entre ellos dos correspondientes al otorgamiento de créditos; Islam et al (2007) realiza una comparación contra el clasificador Naive-Bayes para datos de aprobaciones de tarjetas de crédito; Marinakis et al (2008) aplican algoritmos metaheurísticos para clasificar 1411 empresas de un portafolio de créditos de un banco en Grecia mediante vecinos más cercanos y realizan una comparación del método contra SVM, CART, entre otros.

A pesar de ser un método ampliamente utilizado en múltiples aplicaciones relativas a otras áreas, en el caso del credit scoring no ha sido adoptado tan ampliamente dada la percepción en la demanda computacional que requiere (Hand, 1997), aspecto que cada vez resulta un problema menor ante los avances en el cómputo actual.

4.2.7 Máquinas de Boltzmann restringidas

Una máquina de Boltzmann restringida (RBM) es una red neuronal con dos capas, una formada por unidades visibles y otra con unidades ocultas. En esta investigación, tanto las unidades ocultas como las visibles son consideradas como variables aleatorias binarias.

Una RBM es un tipo particular de las denominadas máquinas de Boltzmann (BM). Las BM fueron introducidas por Hinton et al. (1984) para referirse a un método no determinista diseñado para resolver un problema de visualización de imágenes. De acuerdo con los autores, una BM es una red de unidades binarias estocásticas acopladas simétricamente. Contiene un conjunto de unidades visibles $\mathbf{v} \in \{0,1\}^D$, y un conjunto de unidades ocultas $\mathbf{h} \in \{0,1\}^D$. La energía del estado $\{\mathbf{v}, \mathbf{h}\}$ se define como

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\frac{1}{2}\mathbf{v}^t \mathbf{L} \mathbf{v} - \frac{1}{2}\mathbf{h}^t \mathbf{J} \mathbf{h} - \mathbf{v}^t \mathbf{W} \mathbf{h} \quad (4.30)$$

donde $\theta = \{\mathbf{W}, \mathbf{L}, \mathbf{J}\}$ son los parámetros del modelo, los cuales representan las interacciones simétricas entre unidades visibles a ocultas, visibles a visibles y ocultas a ocultas, respectivamente. Los elementos de la diagonal de las matrices \mathbf{L} y \mathbf{J} son cero. El modelo asigna al vector visible \mathbf{v} la probabilidad dada por

$$p(\mathbf{v}; \theta) = \frac{1}{Z(\theta)} \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h}; \theta)} \quad (4.31)$$

donde el término normalizador $Z(\theta) = \sum_v \sum_h e^{-E(v,h;\theta)}$ se denomina función de partición. Las probabilidades condicionales para las unidades ocultas y visibles están dadas, respectivamente, por

$$p(h_j = 1 | \mathbf{v}, h_{-j}) = \sigma(\sum_i W_{ij} v_i + \sum_{k \neq j} J_{jk} h_k) \quad (4.32)$$

$$p(v_i = 1 | \mathbf{h}, v_{-i}) = \sigma(\sum_j W_{ij} h_j + \sum_{k \neq i} L_{jk} v_k) \quad (4.33)$$

donde $\sigma(x) = (1 + e^{-x})^{-1}$ es la función logística. El algoritmo original para BM (Hinton y Sejnowski, 1983) suponía la inicialización aleatoria de cadenas de Markov para aproximar sus distribuciones de equilibrio y estimar una serie de operadores de esperanza matemática resultado de aplicar gradiente descendente a la log-verosimilitud asociada al modelo de BM. El algoritmo desarrollado por estos autores resultó muy lento para fines prácticos, por lo que ha sido necesario hacer simplificaciones como las planteadas en RBM.

A diferencia de BM, en RBM las conexiones solamente se permiten entre unidades de distintas capas pero no entre unidades de la misma capa. En la Figura 4 se muestran las arquitecturas típicas de BM y RBM.

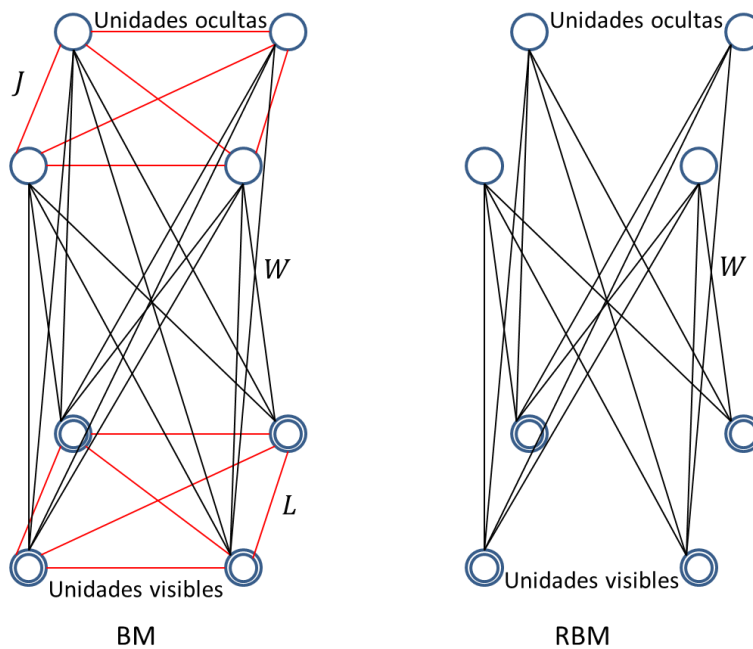


Figura 4.4. Arquitectura de BM (izquierda) y RBM (derecha)

La razón principal que motiva el uso de la arquitectura RBM es que se han desarrollado algoritmos de entrenamiento eficientes, como por ejemplo, el llamado algoritmo de divergencia contrastiva (CDA) propuesto por Hinton (2002).

A la especificación de RBM se le asocia una función de energía definida por

$$E(v, h) = -a'v - b'h - h'Wv, \quad (4.34)$$

donde W es una matriz de pesos simétrica que conecta las unidades visibles (v) y ocultas (h), mientras que a y b son los vectores de sesgo para las capas visible y oculta, respectivamente.

Se puede verificar que toda RBM tiene la propiedad de que las variables visibles son condicionalmente independientes de las variables ocultas, y viceversa, específicamente

$$p(h_j = 1 | v) = \sigma(\sum_i w_{ij}v_i + a_j) \quad (4.35)$$

$$p(v_i = 1 | h) = \sigma(\sum_j w_{ij}h_j + b_i) \quad (4.36)$$

Los parámetros W, a, b se pueden estimar aplicando el algoritmo CDA (ver Hinton, 2002).

Para efectos de utilizar RBM como un clasificador se emplean algún clasificador simple que entrene sobre la capa oculta. Por ejemplo, la regresión logística ha sido una alternativa utilizada muy comúnmente.

La información y búsquedas realizadas en la literatura nos indican que no existe antecedente sobre el uso de RBM para crear clasificadores en el área de CS.

4.2.8 Bagging y Boosting

En el área de aprendizaje maquina existen métodos de clasificación que combinan distintos clasificadores, de tal forma que el desempeño o capacidad predictiva del conjunto completo (*ensemble*) es superior a la que posee cada uno de los algoritmos de clasificación individualmente. En vez de elegir al clasificador con el mejor desempeño sobre conjunto de entrenamiento, este enfoque sugiere utilizar los esquemas de clasificación disponibles, y posiblemente con desempeño individual bajo, para construir clasificadores híbridos. Una de las desventajas de esta aproximación es la interpretación de sus resultados, sin embargo, existen alternativas que permiten obtener descripciones estructuradas sobre los mecanismos del aprendizaje de estos métodos.

Entre los enfoques más prominentes para construir clasificadores combinando clasificadores se encuentran: *bagging*, *boosting* y *stacking*. La mayoría de las veces, estos esquemas han probado mejorar la capacidad predictiva con respecto a los clasificadores individuales. Los dos primeros enfoques utilizan un esquema de ponderación (votación) para construir las predicciones de clase. En el enfoque de *bagging*, los modelos reciben el mismo peso, mientras que en *boosting* los modelos con mejor desempeño son recompensados con ponderaciones mayores que el resto.

En clasificación se consideran dos fuentes de error: 1) el sesgo (o tasa de error) que mide qué tan bien el clasificador ajusta al problema concreto y sobre el cual considera un margen de

error pues en la práctica es desconocido y solamente se puede aproximar; y 2) la varianza, resultado del hecho de que el conjunto de entrenamiento es finito y en consecuencia puede no ser representativo de toda la población de ejemplos. La suma de estas dos fuentes se denomina error total esperado del clasificador. Los métodos de *ensemble*, buscan que con la inclusión de más clasificadores, se tenga una reducción en el componente de varianza.

Bagging busca neutralizar la inestabilidad de los métodos de aprendizaje alterando el conjunto de entrenamiento mediante la eliminación de algunas instancias y la réplica de otras, en vez de seleccionar una nueva muestra independiente en cada iteración. Las instancias se seleccionan aleatoriamente con reemplazo a partir del conjunto original para crear un nuevo conjunto del mismo tamaño, lo cual replica algunos ejemplos y elimina otros. El término bagging, que se utiliza para referirse a un proceso de *bootstrap* agregado, aplica algún esquema de aprendizaje específico sobre cada uno de los conjuntos artificialmente generados y los clasificadores generados a partir de éstos votan sobre la predicción de la clase. No obstante que los conjuntos construidos artificialmente no son independientes entre sí, generalmente mejora el desempeño en comparación con el esquema de aprendizaje individual y no presenta deterioros sustanciales. De hecho, *bagging* ayuda sustancialmente en el caso en que los esquemas de aprendizaje base sean inestables ante pequeños cambios en los datos de entrada. Por ejemplo, utilizar *bagging* sobre árboles de decisión sin proceso de poda ayuda a obtener mecanismos con mayor generalización debido a la inestabilidad de tales árboles. Una mejora a *bagging* se logra cuando los clasificadores utilizados generan probabilidades de clase en sus salidas, con lo que el esquema combinado también generaría probabilidades ponderadas en su salida. *Bagging* también se ha utilizado bajo esquemas sensibles a costos.

Por su parte, *boosting* combina modelos múltiples capaces de complementarse entre sí, es decir, se explotan cada modelo en su parte del dominio donde mejor desempeño tiene respecto del resto. Al igual que *bagging*, este enfoque utiliza votación (para clasificación) o promedio (para predicciones numéricas) para combinar las salidas de los modelos individuales. Combina modelos del mismo tipo, como los árboles de decisión, por ejemplo. A diferencia de *bagging* donde los modelos individuales se construyen de forma separada, el esquema de *boosting* es iterativo, de manera que el desempeño de los modelos en cada iteración es influenciado por el desempeño de los modelos construidos previamente. Asimismo, pondera a cada modelo por su nivel de confianza en vez de utilizar pesos uniformes. Existen diversas variantes del enfoque *boosting*. Una de las más conocidas y específicamente diseñadas para clasificación es la denominada *AdaBoost.M1*. Este enfoque utiliza un peso específico para cada instancia, con lo que el error total de la clasificación se ve modificado como la suma de los pesos de todas las instancias mal clasificadas dividida por el peso de todas las instancias. El uso de pesos específicos sobre las instancias permite que el clasificador se concentre en ciertos subconjuntos de instancias. Inicialmente, el peso asignado a cada instancia es el mismo, sin embargo, conforme se realizan más iteraciones, las ponderaciones cambian para dar más peso a las instancias mal clasificadas, con lo que el algoritmo dedica más esfuerzo en clasificar las instancias más “difíciles”. Se

monitorean los efectos causados sobre las instancias bien clasificadas en pasos anteriores y sobre las instancias cuya clasificación se ha podido corregir. El algoritmo busca un equilibrio en las ponderaciones hasta alcanzar mejoras sustanciales mediante ciertos esquemas de actualización de los pesos (ver Witten et al, 2011).

El enfoque de *stacking* no es tan utilizado como los dos anteriores debido a la dificultad que generalmente involucra su análisis teórico y no tiene una estrategia estandarizada para su implementación. Este método permite combinar clasificadores de distinto tipo, por ejemplo, un árbol de decisión con un clasificador bayesiano *Naive*. De igual forma que en *bagging*, este mecanismo utiliza votación.

Existe una basta literatura sobre los métodos de ensemble y su uso en en contexto de CS ha ido en incremento en la literatura reciente. El término *bagging* fue acuñado por Breiman (1996), mientras que el enfoque *stacking* fue originado por Wolpert (1992).

4.3 Medición del desempeño

En esta sección se revisan algunas de las distintas métricas que se han empleado en la literatura, tanto en el área de minería de datos como de estadística y particularmente en los problemas de clasificación y reconocimiento de patrones. Específicamente, se analizan diversas alternativas que tanto en la industria como en la academia se han utilizado para comparar el desempeño de los modelos de calificación crediticia (CS).

Para evaluar la capacidad predictiva de un modelo que ha sido ajustado utilizando datos históricos, es necesario medir el desempeño que tendrá en el futuro sobre datos que no han sido utilizados en el proceso de estimación. Asimismo, cuando se cuenta con varios modelos que buscan resolver algún problema de pronóstico común, resulta necesario contar con algún criterio de decisión para seleccionar al mejor modelo de acuerdo a poder predictivo.

Usualmente, en el proceso de estimación y evaluación de los modelos de clasificación intervienen tres conjuntos importantes: entrenamiento, validación y prueba, los cuales pueden formar una partición del conjunto de datos⁶. El primero de ellos se utiliza para para calibrar el modelo, en tanto que el segundo sirve para probar diferentes configuraciones de los parámetros o selección de variables. Por su parte, el conjunto de prueba está formado por instancias (ejemplos u observaciones) no evaluadas durante el entrenamiento, el cual se utiliza para evaluar el modelo ajustado.

⁶ El conjunto resultante de la unión de estos tres conjuntos puede coincidir con el conjunto de datos original, o bien, formar un subconjunto propio el cual puede ser obtenido mediante alguna técnica de muestreo algún otro criterio de segmentación.

Como métrica natural del modelo ajustado, generalmente se calcula la tasa de error. La cual se puede entender como la probabilidad de que la clase actual y la predicción de clase obtenida bajo el modelo sean distintas. En la práctica, ésta se estima como el número de casos en que la clase fue mal pronosticada dividido sobre el número total de ejemplos o instancias.

Cuando se busca medir la calidad predictiva o de clasificación del modelo se recomienda hacerlo sobre el conjunto de prueba, ya que, por construcción, la tasa de error obtenida sobre el conjunto de entrenamiento tiende a ser menor que la tasa de error de un conjunto de prueba. Esto se debe principalmente a que el modelo ajustado es construido mediante algoritmos que minimizan el error de clasificación sobre los conjuntos de entrenamiento y/o validación. En consecuencia, la tasa de error no constituye una base de comparación adecuada. Asimismo, dos algoritmos no se pueden comparar con base al comportamiento de los errores obtenidos durante el entrenamiento, ya que es muy probable que el modelo con mayor número de parámetros o aquél con mayor complejidad, presente el mejor desempeño⁷. Estas observaciones nos indican que el proceso de validación del desempeño de un modelo debe realizarse con un conjunto distinto al utilizado en el entrenamiento.

Cuando las etapas de entrenamiento y prueba son realizadas en un solo conjunto, respectivamente, se corre el riesgo de obtener resultados que puedan estar afectados por ruido, observaciones atípicas (outliers) y otros factores aleatorios de las muestras. Para ello, es deseable repetir el experimento un cierto número de veces y promediar los resultados obtenidos. En el caso en que el costo de repetir los procesos de entrenamiento y prueba es demasiado alto, el ejercicio se puede realizar una sola ocasión.

La evaluación del algoritmo de aprendizaje se puede enfocar en el estudio del comportamiento de los errores de validación del modelo (sobre diferentes conjuntos de prueba), cuya distribución se puede comparar con la de los errores de validación de algún modelo alternativo. Sin embargo, de acuerdo con Turney (2000), existen otros criterios que también se deben considerar para efectos de comparación entre modelos:

- Las implicaciones de generalizar la medida de error con funciones de pérdida⁸
- Tiempo de los procesos de entrenamiento y prueba, respecto de la complejidad del espacio paramétrico⁹
- Nivel de interpretación humana de las reglas de decisión inducidas por el modelo

⁷ Esto último sucede debido a que las construcciones de modelos más complejos generalmente se formulan sobre la base de modelos simples, ya sea heredando algunas propiedades de estos últimos, mezclando sus características o aumentando el número de parámetros para obtener generalizaciones de los mismos.

⁸ Una función de pérdida es una función que mapea al error del modelo en un número real que intuitivamente representa el “costo” asociado. Algunas funciones comunes son el valor absoluto y la función cuadrática.

⁹ Se entiende por espacio paramétrico al subconjunto que contiene a todos los valores permisibles (posibles) que pueden tomar los parámetros del modelo.

- Fácil implementación

Finalmente, un aspecto que conviene tener presente, es que las conclusiones obtenidas del algoritmo de aprendizaje son condicionales al conjunto de datos. Como se ha comentado, el conjunto de entrenamiento se utiliza para estimar los parámetros bajo cierto algoritmo de aprendizaje, en tanto que el conjunto de validación generalmente se emplea para calibrar algunos parámetros específicos del algoritmo o de su estructura misma. Con el conjunto de prueba se determina el error del modelo que fue calibrado (entrenado) con la información contenida en los conjuntos de entrenamiento y validación.

4.3.1 Validación cruzada

La validación cruzada es un método ampliamente utilizado en el cual se busca hacer el mayor aprovechamiento del conjunto de datos X mediante el uso de subconjuntos traslapados que conducen a errores dependientes. Se busca generar K parejas de conjuntos de validación/entrenamiento $\{T_i, V_i\}_{i=1}^K$ sobre el mismo conjunto X , de manera que el número de parejas distintas sea tan grande como sea posible, al tiempo que las estimaciones del error sean robustas y los traslapes entre conjuntos diferentes sean lo más pequeño posible. Por otro lado, se debe buscar que la proporción de ejemplos de cada clase permanezca invariante con respecto a la composición del conjunto de datos original (estratificación).

Más específicamente, la validación cruzada consiste en subdividir al conjunto X en K partes de igual tamaño, denotadas por $\{X_i\}_{i=1}^K$. Para construir el conjunto de validación se toma cualquiera de ellas y con el resto de $K - 1$ partes se forma el conjunto de entrenamiento. Este proceso se repite para obtener K parejas de la forma $\{T_i, V_i\}$, donde $V_i = X_i$ y $T_i = \bigcup_{j \neq i} X_j$ para $i = 1, 2, \dots, K$.

La validación cruzada presenta dos problemas principales como consecuencia de su esquema de construcción. El primero tiene que ver con el alto grado de traslape entre conjuntos, ya que, por ejemplo, cualesquiera dos conjuntos de entrenamiento comparten $K - 2$ particiones. Esto implica efectos de dependencia serial entre los errores obtenidos por cada pareja. El segundo problema tiene que ver con el hecho de que para mantener un tamaño grande del conjunto de prueba, el tamaño del conjunto de validación debe ser pequeño.

A medida que el tamaño de K aumenta, el tamaño del conjunto de entrenamiento crece y entonces se obtienen estimadores del error más robustos, sin embargo, el conjunto de validación reduce cada vez más su tamaño. Adicionalmente, el proceso de entrenamiento incurre en un costo que está en relación directa con el valor de K . A medida que N crece, el valor de K puede ser menor; mientras que si N es pequeño, los niveles que debe tomar K deben ser los suficientemente grandes para permitir contar con suficientes conjuntos de entrenamiento.

De acuerdo con Witten et al (2011), se han realizado diversos estudios intensivos sobre numerosos conjuntos de datos, con distintas técnicas de aprendizaje, los cuales han mostrado que alrededor del nivel $K = 10$ se obtienen los mejores estimadores del error. Sin embargo, existe muy poca evidencia teórica que sustente este hecho. No obstante que estos argumentos no son concluyentes y el debate en las áreas de minería de datos y aprendizaje maquinal sobre cuál es el mejor esquema para la evaluación sigue vigente, la validación cruzada con diez particiones se ha posicionado como el método estándar en términos prácticos. Los autores también refieren que los ejercicios empíricos han mostrado que el uso de estratificación mejora los resultados ligeramente, debido a la reducción de varianza del error que se obtiene (aunque no se elimina completamente). Adicionalmente, una estrategia para mejorar la estimación del error consiste en repetir 10 veces el proceso de validación cruzada con diez particiones y promediar los resultados. Esto evidentemente crea una demanda de recursos computacionales importante. Adicionalmente, elecciones de otros valores de K entre 5 y 20 también pueden dar tan buenos resultados como el caso $K = 10$.

4.3.2 Métricas para el desempeño del clasificador

Métricas basadas en la matriz de coincidencias

En este apartado se abordan distintas métricas que sirven para comparar el desempeño de varios clasificadores. En el caso de los problemas de clasificación, la información básica sobre el desempeño del modelo ajustado se puede resumir en la denominada matriz de confusión. Para la situación de un problema de dos clases, los componentes de dicha matriz se pueden describir de la manera siguiente:

		Clase verdadera	
		Positivo	Negativo
Clase pronosticada	Positivo	Cuenta de positivos verdaderos (TP)	Cuenta de positivos falsos (FP)
	Negativo	Cuenta de negativos falsos (FN)	Cuenta de negativos verdaderos (TN)

Figura 4.5. Matriz de coincidencias

De acuerdo con la Figura 4.5, las entradas sobre la diagonal representan el número de aciertos que tiene el modelo cuando se clasifica un cierto conjunto de ejemplos u observaciones de cierta población de los cuales se conoce su verdadera clase. El objetivo de la matriz de confusión

consiste en contabilizar cuatro cantidades importantes: el número de ejemplos correctamente clasificados como clase “positivo” (TP) y los correctamente clasificados en la clase negativo (TN), así como el número de ejemplos de la clase “positivo” que fueron clasificados erróneamente como “negativo” (FN) y los casos clasificados como “negativo” pero que pertenecen a la clase “positivo” (FP). La generalización de esta matriz para el caso de problemas de más de dos clases también es útil para contabilizar el total de aciertos y errores cometidos por el modelo en el pronóstico de la clase a la que pertenece cada ejemplo probado.

Para los problemas de clasificación donde el número de clases es $d > 2$ se pueden diseñar matrices de confusión similares a la Figura 1. En este caso, la matriz de confusión se construye como una matriz de dimensión $d \times d$ cuyas entradas (i, j) contiene el número de instancias que pertenecen a la clase C_i pero han sido asignadas a la clase C_j . En el caso ideal en que no se tiene error de clasificación, todas las entradas fuera de la diagonal debieran ser igual a cero. Alternativamente, se pueden definir d problemas de dos clases, cada uno separando una clase específica de las $d - 1$ clases restantes. En este trabajo de investigación se estudia el problema de credit scoring, en el cual el número de clases es dos bajo el análisis tradicional de buenos y malos clientes. Sin embargo, en el caso en que se requiera definir una escala interna que contemple el estudio de las distintas categorías de clientes, el uso de matrices de confusión de múltiples clases.

Con la información contenida en la matriz de confusión es común construir diversas métricas de desempeño, entre las que destacan:

$$\text{Tasa de correctos (éxito)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.38)$$

$$\text{Tasa de error} = 1 - \text{Tasa de éxito} \quad (4.39)$$

$$\text{Tasa de positivos verdaderos} = \frac{TP}{TP+FN} = \text{Recall} = \text{Sensibilidad} \quad (4.40)$$

$$\text{Tasa de positivos falsos} = \frac{FP}{FP+TN} = \text{Fallout} \quad (4.41)$$

$$\text{Precisión} = \frac{TP}{TP+FP} \quad (4.42)$$

$$\text{Especificidad} = \frac{TN}{TN+FP} = \text{Tasa de negativos verdaderos} \quad (4.43)$$

$$\text{Medida F} = \frac{2}{\frac{1}{\text{Precisión}} + \frac{1}{\text{Sensibilidad}}} \quad (4.44)$$

$$\text{Media armónica} = \frac{2 \cdot \text{Sensibilidad} \cdot \text{Especificidad}}{\text{Sensibilidad} + \text{Especificidad}} \quad (4.45)$$

En las expresiones anteriores se han incluido los nombres alternativos que en distintos contextos se han utilizado para referirse a la misma métrica. En la práctica y la literatura se pueden encontrar métricas adicionales, o bien, variantes de las mismas. Para expresar las medidas anteriores como tasas en %, es necesario multiplicar cada cantidad por 100.

La tasa de éxito, también conocida como porcentaje de clasificación correcta (PCC) de las observaciones, mide la proporción de los casos correctamente clasificados en una muestra de datos. En un cierto número de casos, el PCC puede que no sea el mejor criterio de desempeño. Tácitamente asume que los costos de mala clasificación son equivalentes para predicciones positivamente falsas y negativamente falsas. Este supuesto es problemático, ya que para la mayoría de los problemas en la vida real, un tipo de error de clasificación puede resultar más caro que otro. Otro supuesto implícito del uso del PCC como una forma de criterio de evaluación es que la distribución de clases (priors de la clase) entre los ejemplares, se presume constante en un horizonte de tiempo y es relativamente balanceado. Además, usar el PCC por sí solo, a veces demuestra ser inadecuado toda vez que las distribuciones de clases y los costos de mal calificar clientes son raramente uniformes. De cualquier forma, tomar en cuenta las distribuciones de clases y los costos de mal calificar clientes parece ser un poco duro, ya que en la práctica, éstos raramente pueden ser especificados con precisión y en general se encuentran sujetos a cambios.

La sensibilidad mide la proporción de ejemplos positivos que se predijo iban a ser positivos (es decir, $TP/(TP + FN)$), mientras que la especificidad mide la proporción de ejemplares negativos que se predijo iban a ser negativos (osea, $TN/(FP+TN)$). Note que sensibilidad, especificidad y el PCC varían juntos conforme el umbral en una salida continua de un clasificador varía entre sus extremos.

4.3.3 Receiver operating characteristics curve (ROC)

En el contexto del problema de dos clases, consideremos un sistema que devuelve la probabilidad de que una instancia pertenezca a la clase positiva, la cual denotamos por $P(C_1|x)$. Entonces la probabilidad asociada a la clase negativa está dada por $P(C_2|x) = 1 - P(C_1|x)$. Adicionalmente, supongamos además que $P(C_1|x) > \theta$. Entonces, para θ cercano a 1, resulta poco probable que dada una instancia arbitraria, ésta pertenezca a la clase positiva; es decir, se no habría positivos falsos sino unos pocos positivos verdaderos. A medida que θ disminuye su valor acercándose a cero, el número de positivos verdaderos aumenta pero se incurre en el riesgo de introducir positivos falsos en el análisis de clasificación.

Para distintos valores de θ se pueden obtener las parejas ordenadas de positivos verdaderos y positivos falsos. Al conjunto de puntos que se obtiene la llamada curva o gráfica ROC (receiver operating characteristics) como se muestra en la figura 5.6 ROC es una ilustración gráfica bidimensional de la sensibilidad en el eje Y versus (1-especificidad) en el eje X para varios

valores de la clasificación del umbral. Esto básicamente ilustra el comportamiento de un clasificador sin tener en cuenta la distribución de las clases o el costo de mal calificar clientes, de modo que desacopla efectivamente el desempeño de la clasificación de estos factores.

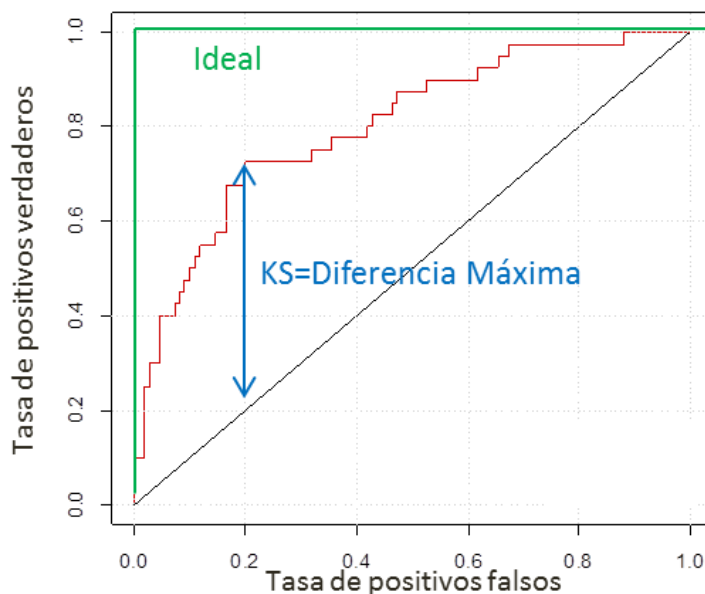


Figura 4.6. Curva ROC

En el caso ideal de un clasificador con cero tasa de error, se tendría una tasa de positivos verdaderos igual a 1, mientras que la tasa de positivos falsos sería 0. Es decir, la forma de ROC sería la representada por la línea verde de la Figura 4.6. La situación ilustrada por la recta identidad (pendiente 1) corresponde a un clasificador que arrojar el mismo número de decisiones verdaderas que falsas, lo cual representaría el peor escenario de desempeño de cualquier clasificador (en el caso de estar por debajo de esta recta bastaría con invertir el orden de las clases para regresar a un comportamiento como el de la curva en rojo). Dados dos clasificadores, se puede decir que uno de ellos es mejor que otro si la curva ROC del primero se encuentra por encima de la del segundo. En el caso en que ambas curvas se intersecten, el criterio de decisión común suele ser el área debajo de ROC (AUC, por sus siglas en inglés). El clasificador ideal tendría AUC igual a 1 y distintos clasificadores se comparan de acuerdo a su AUC comparando así su desempeño promedio sobre distintas condiciones de valores de funciones de pérdida¹⁰. Alternativamente, la máxima distancia entre la curva ROC y la recta identidad define a la estadística KS (Komogorov-Smirnov) asociada a este problema. El valor de KS también ha sido utilizado como criterio de comparación entre modelos. En el contexto de credit scoring, al área AUC también se le denomina coeficiente de Gini. Uno aspecto importante en este contexto tiene

¹⁰ Se puede demostrar que distintos valores de θ corresponden a diferentes matrices de pérdida para ambos tipos de error. En consecuencia, la curva ROC se puede considerar como el comportamiento del clasificador bajo distintas matrices de pérdida (Alpaydin, 2010)

que ver con la utilidad que tiene ROC para determinar puntos de corte para discriminar a los buenos de los malos clientes.

Para efectos de este trabajo, se utilizarán como criterios de desempeño de los modelos calibrados a la tasa de éxito y AUC.

Para el caso de múltiples clases, Hand y Till (2001) propusieron la generalización de la métrica AUC. Por su parte, como una métrica alternativa Hand (2009) propuso una medida coherente alternativa a AUC, la cual denomina como medida H. Existen métricas alternativas como las llamadas gráficas de costos propuestas por Drummond, C.y R. Holte (2006).

4.4. Resultados

En esta sección se realiza un análisis descriptivo de los conjuntos de datos los cuales se utilizarán posteriormente para calibrar modelos de scoring sobre la base de ciertos métodos estadísticos tradicionales y de aprendizaje maquina.

Se cuenta con acceso a tres bases de datos públicas y dos de carácter privado, cuyo origen no se proporciona por razones de confidencialidad. La información pública contiene dos conjuntos de datos disponibles en el repositorio de UCI (<http://archive.ics.uci.edu/ml/>). Uno de ellos, denominado “German credit data”, corresponde a información sobre distintos atributos de 1000 personas clasificadas con un buen o mal perfil de riesgo de crédito. El segundo conjunto “Australian credit approval” contiene 690 instancias de aplicaciones de crédito las cuales están clasificadas como aceptadas o rechazadas. Por su parte, el tercer conjunto público corresponde a información proveniente de la operación de tarjetas de crédito de una gran cadena de retail en Brasil durante el periodo 2003-2008. Dicha información fue solicitada en la página de la 13th Pacific-Asia Knowledge Discovery and Data Mining conference (<http://sede.neurotech.com.br:443/PAKDD2009/>), donde se proporcionan dos subconjuntos, uno de entrenamiento formado por 50,000 instancias y otro para prueba constituido por 10,000 ejemplos. A este conjunto se le referirá como “PAKDD2009”.

Con respecto a la información privada, se cuenta con un conjunto denominado “Cerveza” el cual contiene información relativa a 3,749 clientes a quienes una empresa vendió cerveza durante 2009 y 2010. Para 2,593 de estos clientes, la venta del producto se realizó mediante el otorgamiento de líneas de crédito las cuales fueron monitoreadas para incluir luego indicadores del comportamiento de su perfil de riesgo de los clientes a quienes se otorgó crédito. El segundo conjunto disponible, corresponde a información socio-demográfica y de seguimiento de patrones de pago de una cartera de 48,550 préstamos de auto otorgados entre 2006 y 2009. Este conjunto denominado “Autos”, nos permitirá realizar un análisis dinámico del perfil de riesgo de los clientes.

En la siguiente lista se anota entre paréntesis el nombre que se utilizará en el resto del documento para hacer referencia a cada conjunto:

- “German credit data” (Alemán)
- “Australian credit approval” (Australiano)
- “PAKDD2009” (PAKDD)
- “Cerveza” (Cerveza)
- “Autos” (Autos)

Los conjuntos Alemán y Australiano han sido ampliamente utilizados en la literatura para comparar distintas técnicas de clasificación, tanto en el área del credit-scoring como en contextos más generales. Por su parte, a raíz de la competencia PAKDD 2009, el conjunto PAKDD ha recibido atención algunos artículos recientes. En cambio, la información de los conjuntos Cerveza y Autos representan información proveniente del mercado mexicano. El uso de conjuntos reales y que representen características de la economía local, permitirá evaluar los alcances de técnicas alternativas provenientes del aprendizaje maquina para el caso mexicano.

A manera de resumen sobre las características de los distintos conjuntos se incluye la Tabla 4.1 (ver Apéndice). En el siguiente apartado se proporciona una descripción de las características principales de cada conjunto de datos, así como el tipo de problema específico que cada uno permitirá abordar en el contexto de credit scoring y en línea con los objetivos planteados en este proyecto de investigación.

a) Conjunto Alemán

Este conjunto de datos contiene información relativa a 1,000 solicitantes de crédito, descritos por un conjunto de atributos. El archivo original proporcionado por el Dr. Hans Hofmann se encuentra compuesto por atributos cuyos valores están expresados en categorías mediante símbolos específicos de acuerdo con la Tabla 2 (ver Apéndice).

El atributo de clase indica el perfil crediticio de cada solicitud mediante dos clases: bueno y malo. Sin incluir al atributo de clase, se tienen 20 atributos (7 numéricos y 13 nominales). Del total de ejemplos, 700 tienen buen perfil crediticio y 300 están clasificados con mal perfil. La Figura 4.7 del Apéndice muestra el comportamiento de la densidad de los valores que toma cada uno de los atributos.

Para trabajar con una versión del conjunto alemán con atributos numéricos, la universidad de Strathclyde, Glasgow, produjo un segundo archivo editado el cual cuenta con la inclusión de algunas variables indicadoras para su uso en algoritmos que no trabajan con variables categóricas. Algunos de los atributos de tipo categórico fueron codificados como enteros. El conjunto resultante cuenta con un total de 24 atributos (sin incluir la variable de clase) y, como se ha indicado, todos ellos son de tipo numérico.

La información disponible no indica la presencia de datos faltantes y no proporciona detalle sobre el procesamiento de los datos.

El objetivo de los métodos de clasificación a utilizar consiste ajustar algún modelo que permita clasificar las dos clases en función de los atributos disponibles. Este problema es de tipo estático y con él se busca determinar si con el análisis de los atributos se puede determinar el perfil de riesgo de crédito de nuevas solicitudes.

b) Conjunto Australiano

Contiene información sobre otorgamiento de tarjetas de crédito. Los nombres de los atributos y sus valores originales fueron transformados a ciertos símbolos con la finalidad de proteger la confidencialidad de los datos. El archivo original fue proporcionado por quinlan@cs.su.oz.au.

Este conjunto de datos contiene 690 ejemplos, de los cuales 307 pertenecen a la clase con mal perfil crediticio (malo) y 383 tienen un buen perfil de riesgo (bueno). Se tienen 14 atributos (sin incluir el atributo de clase), 6 de los cuales son numéricos y 8 categóricos. La Tabla 4.3 (ver Apéndice) incluye la descripción de cada uno de los atributos.

De acuerdo con la información proporcionada en el sitio de UCI, los datos faltantes se reemplazaron por el promedio observado en cada atributo en el caso de las variables de intervalo (numéricas) y por la moda del atributo en el caso de las variables categóricas.

Ambos conjuntos de datos, el alemán y australiano, serán utilizados para aplicar los modelos estadísticos tradicionales y algunas técnicas de aprendizaje maquinal para discriminar a los clientes de buen perfil (Buenos) de los clientes (Malos) en un enfoque estático.

Es importante mencionar que las dos fuentes de información han sido ampliamente utilizadas en la literatura para comparar distintos modelos de clasificación. Por mencionar algunos estudios recientes se tienen los trabajos realizados por Baesens et al (2003), Hoffmann et al (2007), Huang et al (2006, 2007), Ong et al (2005), West (2000, 2005), Zhou (2008).

c) Conjunto PAKDD

La información disponible corresponde a la utilizada en una competencia sobre minería de datos organizada por NeuroTech Ltd y el Centro de Informática de la Universidad Federal de Pernambuco, Brazil, presentada en 13th Pacific-Asia Knowledge and Data Mining conference (PAKDD 2009).

Para propósitos de la competencia, se proporcionaron tres conjuntos de información: entrenamiento, tabla de posiciones y prueba. Los tres conjuntos contienen información capturada a lo largo de un año completo, cada uno durante periodos no adyacentes. El conjunto de entrenamiento consta de 50,000 ejemplos con los que se debe extraer información que mejor

explique el perfil de riesgo de los clientes (bueno/malo¹¹), el cual incluye 9,874 clientes con mal perfil de riesgo y 40,126 con buen perfil.

Los otros dos conjuntos sirven para evaluar el desempeño de los modelos ajustados y cada uno cuenta con 10,000 ejemplos. Por razones de la dinámica de la competencia PAKDD 2009, para ninguno de estos dos conjuntos se provee la columna de clase. La intención del conjunto de tabla de posiciones es poder realizar un ejercicio previo, mientras que el conjunto de prueba se utilizó para evaluar las distintas metodologías implementadas con la finalidad de elegir a la ganadora.

Los ejemplos de cada conjunto fueron obtenidos mediante muestreo aleatorio de experiencia histórica registrada en portafolios de tarjetas de crédito otorgadas por una de las cadenas departamentales (retail) más grandes de Brasil. En cada caso, la base inicial contiene información sobre 31 variables explicativas, y solamente en el conjunto de entrenamiento, se dispone del atributo de clase (perfil de riesgo).

La compañía ha operado el otorgamiento de las tarjetas de crédito durante más de 8 años aplicando básicamente dos distintos modelos de evaluación del riesgo, de manera que su tasa de aceptación ha variado entre 50% y 75%. Asimismo, no obstante que las condiciones económicas del periodo en cuestión han mostrado cierta estabilidad, imperfecciones como el ruido, datos faltantes, informaciones atípicas (outliers) y otros cambios en el mercado, pueden reducir el desempeño de los modelos estimados.

En la Tabla 4.4 del Apéndice se incluye la descripción de los atributos, en la que no se incluye al identificador (único) del cliente.

Para el conjunto de prueba se detectaron 8 atributos cuyos valores tienen tasa de faltantes del 100% (ya sea porque todos sus valores están vacíos o bien tienen un solo valor constante para todos los ejemplos), los cuales serán eliminados de la muestra para propósitos del análisis. Como resultado se mantienen 22 atributos (10 de tipo numérico y 22 nominales).

Se propone utilizar alguna transformación logarítmica para disminuir el efecto que tienen las observaciones con valores atípicamente grandes (outliers), así como considerar el uso de algún umbral máximo permisible.

d) Conjunto Cerveza

Este conjunto de datos consta de registros de clientes a los que se les vendió cerveza durante 2009 y 2010. La base de datos original consta de 3,749 clientes, de los cuales a 2,593 se les

¹¹ La forma en que opera este producto de crédito es estándar. Los clientes (aplicaciones de crédito aceptadas) utilizan su tarjeta para realizar compras en la cadena departamental, las cuales se facturan a la cuenta del cliente para pagar entre 10 y 40 días después de realizada la compra. La fecha límite de pago se fija en cierto día del mes. Un cliente se etiqueta como malo si, para una ventana de observación de 11 meses después de su primera compra, éste cuenta con algún incumplimiento de pago (morosidad mayor a 60 días). En otro caso el cliente es considerado como buen cliente. En el caso de algún consumo reciente, se incluyen 60 días más de observación como periodo de maduración de la última facturación.

otorgaron líneas de crédito en función de ciertos atributos valorados por la empresa cervecera en cuestión¹², una de las cervecerías más importantes de México. Se disponen 18 atributos (2 numéricos y 16 nominales), de los cuales dos son de tipo clase: uno caracteriza el otorgamiento de línea de crédito (Si/No) y el otro se describe el comportamiento del perfil de riesgo (bueno/malo). La autorización de líneas de crédito responde a una estrategia de la empresa de financiar a sus clientes para que aumente su capacidad de compra de cerveza y con ello lograr un mejor posicionamiento de la marca en el mercado.

En principio, el estudio busca identificar los patrones de aquellos clientes a quienes se otorgó línea de crédito. Asimismo, interesa discriminar de entre estos clientes aquellos que resultaron con buen comportamiento crediticio de los malos pagadores. La caracterización de los rasgos que mejor describen a cada tipo de cliente permitirá a la empresa realizar una mejor selección de clientes a los que es conveniente vender financiar mediante la asignación de líneas de crédito.

La Tabla 4.5 (ver Apéndice) contiene información sobre el tipo de variables contenidas en la base de datos original.

En un análisis exploratorio de los valores faltantes, se detectó que todos los clientes con valores faltantes en los atributos Esquina, Calle, Zona y Local correspondían a los mismos registros (505) y que además todos formaban parte de los faltantes en el atributo Imagen. Luego de eliminar tales registros, el número de casos con información nula en el atributo Clave se redujo de 291 a solamente 22, en tanto que para los atributos Imagen y Control se encontraron 3 y 10 casos, respectivamente. Finalmente, se eliminaron 4 clientes en los que el atributo Resultado reportó ventas totales negativas. Como resultado de omitir los filtros y consideraciones realizadas, el conjunto de datos a analizar contiene 3,200 clientes. Cabe destacar que de estos, solamente a 2,156 les otorgaron crédito.

En la Tabla 4.6 del Apéndice se muestra la distribución porcentual de los atributos categóricos (no numéricos), de acuerdo con los valores asignados a cada categoría. Por su parte, la Tabla 4.7 del Apéndice contiene estadísticas descriptivas para las únicas dos variables numéricas (de intervalo) del conjunto.

Como ya se ha comentado, este conjunto de datos permitirá realizar dos estudios principales. El primero consistirá en desarrollar modelos de clasificación que expliquen el proceso de otorgamiento de la línea de crédito a los 3,200 clientes totales en función de los primeros 14 atributos mostrados en la primera Tabla 4.6. El segundo, consistirá en evaluar el impacto de cada uno de tales atributos sobre el comportamiento crediticio de los 2,156 clientes a quienes se otorgó línea de crédito, es decir, sobre su buen perfil (cumplido) o mal perfil (incumplido).

¹² Por razones de confidencialidad se ha omitido el origen de la fuente de datos

e) Conjunto Autos

Este conjunto contiene información de 48,550 créditos de auto otorgados entre 2006 y 2009 por cierta institución financiera local.

En la Tabla 4.8 (ver Apéndice) se encuentra la descripción de los atributos sociodemográficos que se dispone de cada uno de los créditos otorgados en todo el periodo de la muestra. La base sociodemográfica cuenta con 24 atributos, de los cuales 17 son de tipo numérico y 7 nominales.

Para algunos de los atributos que presentan ejemplos con valores atípicamente grandes (outliers) se contempla el uso de transformaciones logarítmicas y la eliminación de los casos a partir de cierto rango. Entre estos atributos se encuentran: número de dependientes, ingresos brutos, si cuenta con algún crédito hipotecario, ingresos netos, otros ingresos, deudas, precio del automóvil a financiar.

Se cuenta además, con una base que contiene la información del comportamiento de pago observado en los clientes de la base sociodemográfica entre 2008 y 2010 (ver Tabla 9 del Apéndice).

Con la información contenida en la base de comportamiento de pago, será posible realizar distintos análisis del perfil de los clientes (buenos y malos) en función de las características sociodemográficas, la fecha de otorgamiento del crédito y comportamiento de pago de los clientes por meses de saldo vencido. Este tipo de análisis está alineado con el enfoque dinámico planteado en el proyecto de investigación.

4.4.1 Resultados de la experimentación

Se implementó un primer ejercicio en el que se repiten 100 veces el ajuste de los clasificadores bajo validación cruzada (10-fold) utilizando (1,000 corridas). En la siguiente Tabla 4.10 se presenta la configuración de parámetros utilizada en cada algoritmo para cada conjunto de datos. Se utilizó validación cruzada para determinar las configuraciones que presentaran la menor tasa de error en cada caso.

Clasificador	Australiano	Alemán	Cerveza Línea	Cerveza 90+	PAKDD 2009	Autos 90+	Ventajas/Desventajas
Regresión Logística	Ridge (R) = 2.4242	R = 4.7959	R = 2.0408	R = 4.47	R = 2.2414	R = 3.6578	Traducción de un score en probabilidades
Naive Bayes	Discretización supervisada (DS)	DS	DS	DS	DS	DS	Supuesto simple de independencia condicional

K-vecinos más cercanos	#vecinos más cercanos (K) = 19	K = 17	K= 8	K = 13	K = 25	K = 21	Simple/Difícil traducción a score de nuevos clientes
Árbol J48	Factor de confianza (C) = 0.1189	C = .25	C = .4268	C = .01	C = .15	C = .7393	Interpretación y entendimiento de reglas simples/ Sensible a pequeños cambios portafolio
Red perceptrón multicapa	Capas/unidades ocultas (H) = 8; tasa de aprendizaje (L)=.3	H = 6; L = .01	H = 3; L = .12	H = 5; L = .05	H = 6; L = .15	H = 10; L = .25	Posibilidad de traducción en probabilidades score/ “Caja negra”
Máquina soporte vectorial (SVM)	Costo (E) = 32768 ; gamma (G) = .000488	E = 4096 ; G = .00781	E = 2; G = .25	E = 3.5; G = .01	E = 256; G = .125	E = 173; G = .0274	Posibilidad de traducción en probabilidades score; extracción de reglas/ “Caja negra”
Red bayesiana	Discretización (D); Algoritmo de búsqueda local (BL) K2; No base Naive Bayes	D; Algoritmo BL TAN	D; Algoritmo BL TAN; Criterio Entropía	D; Algoritmo BL Tabu	D; Algoritmo BL Tabu	D; Algoritmo BL TAN; Criterio Entropía	Uso de tabla de distribución de probabilidades condicionales
AdaBoost M1	Clasificador base (CB) = Logística	CB = Logística	CB = Logística	CB = Logística	CB = Logística	CB = Logística	“Caja negra”
Bagging	Clasificador base (CB) = Árbol de decisión rápido (REPTree)	CB = REPTree	CB = REPTree	CB = REPTree	CB = REPTree	CB = REPTree	“Caja negra”
RBM + Logística	Unidades ocultas (H) = 180	H = 90	H = 250	H = 170	H = 260	H = 110	Traducción de un score a probabilidades/ “Caja negra”

Tabla 4.10. Configuración de parámetros utilizados en los clasificadores para cada conjunto de datos

Con excepción del clasificador basado en RBM, las distintas técnicas empleadas fueron calibradas con ayuda del software Weka (Pentaho), el cual está disponible en http://www.cs.waikato.ac.nz/~ml/weka/index_downloading.html. Para el caso de RBM, los análisis se implementaron con apoyo de una librería en MATLAB, la cual se puede descargar, también de forma gratuita, en <http://code.google.com/p/matrbm/>. Finalmente, parte del análisis exploratorio y descriptivo de la información se realizó con el paquete estadístico R y la librería de minería de datos Rattle, ambos accesibles desde el sitio gratuito <http://cran.r-project.org/>.

Derivado de los análisis de los resultados, en la Tabla 4.11 se reportan los valores promedio de las medidas las tasas de clasificación correcta de cada modelo (TC) y el área bajo la curva ROC (AUC)¹³.

Clasificador	Australiano		Alemán		Cerveza Línea		Cerveza 90+		PAKDD2009		Autos 90+	
	% TC	AUC	% TC	AUC	% TC	AUC	% TC	AUC	% TC	AUC	% TC	AUC
Red Bayesiana	86.35	0.92	74.56	0.77	82.83	0.87	60.85	0.65	81.14	0.85	72.82	0.76
NaiveBayes	85.28	0.92	74.55	0.77	81.93	0.84	61.21	0.65	80.75	0.85	70.56	0.72
LibSVM	85.54	0.86	75.65	0.77	84.15	0.87	60.81	0.65	81.44	0.73	71.48	0.72
Regresión Logística	86.38	0.93	77.05	0.79	83.82	0.87	61.27	0.66	81.26	0.85	72.14	0.74
Perceptrón multicapa	83.38	0.90	75.47	0.78	83.10	0.87	58.94	0.63	78.97	0.80	70.87	0.69
Vecinos más cercanos	86.17	0.92	72.49	0.75	83.65	0.89	61.87	0.66	81.45	0.82	69.23	0.70
AdaBoostM1	86.23	0.90	76.92	0.72	83.85	0.84	61.29	0.60	81.25	0.82	72.09	0.76
Bagging	86.32	0.94	76.29	0.79	85.93	0.89	60.55	0.65	81.32	0.82	71.23	0.75
Árboles (J48)	85.13	0.88	73.57	0.68	84.46	0.85	61.53	0.63	81.36	0.75	70.04	0.71
RBM+Logística	86.55	0.87	75.07	0.69	85.02	0.88	61.95	0.66	80.23	0.00	72.59	0.75

Tabla 4.11. Tasas clasificación correcta (TC) y área bajo ROC (AUC) de los clasificadores para cada conjunto de datos

Para evaluar si las diferencias obtenidas en el desempeño son significativas se utilizó una prueba T para muestras pareadas con factor de corrección

¹³ La naturaleza de los valores reportados, puede llevar a que las pruebas t sean no transitivas por diferencias significativas entre las varianzas muestrales de los resultados de experimentación en cada modelo y conjunto de datos.

$$t = \frac{\bar{d}}{s_d \sqrt{\left(\frac{1}{n} + \frac{1}{n_1}\right)}} \quad (4.46)$$

El nivel de significancia utilizado es del 5% y se ha señalado en color rojo los casos en que se tiene el peor desempeño. Asimismo, se ha puesto en color azul las situaciones en que los modelos son igualmente competitivos entre sí (no se encontraron diferencias significativas). En negritas se ha resaltado a los modelos que presentaron los niveles más altos en las métricas de desempeño utilizadas. Los casos sin formato se refieren a casos donde por lo menos alguno de los modelos competidores resultó ser superior.

4.5 Conclusiones y líneas futuras de investigación

Como parte de esta investigación, se ha realizado un análisis detallado sobre el desempeño de distintos clasificadores en el contexto del *credit-scoring* (CS). Para ello, se implementó un ejercicio de experimentación sobre distintos conjuntos de datos disponibles aplicando diversas técnicas de clasificación, algunas de las cuales pertenecen al área de estadística y otras han sido desarrolladas en el área del reconocimiento de patrones. Son cinco los conjuntos de información utilizados, tres de los cuales (Alemania, Australiano y PAKDD) se encuentran disponibles de forma pública mientras que los dos restantes (Cerveza y Autos) son de tipo privado. La información pública, representativa de otros mercados en el extranjero, ha sido utilizada en numerosos estudios de clasificación y CS, en tanto que la base privada pertenece al mercado local mexicano.

Por las características propias de cada conjunto de datos, con los distintos modelos implementados se analizó el perfil de riesgo, tanto de contrapartes que solicitan crédito, como de aquellos clientes previamente aceptados. Asimismo, los análisis empíricos realizados fueron orientados de forma intensiva para resolver el problema de CS bajo un contexto de tipo estático.

Entre las técnicas de clasificación empleadas se encuentran: regresión lineal, árboles de decisión, vecinos más cercanos, *naive Bayes*, redes perceptrón multicapa, máquinas de soporte vectorial, *boosting*, *bagging* y un clasificador híbrido formado por máquinas de Boltzmann restringidas (RBM) y regresión logística. Hasta donde se tiene conocimiento, este trabajo sería pionero en el uso de esta última técnica para atacar el problema de clasificación en el área de CS. Los esfuerzos de este proyecto se enfocaron en explorar la aplicabilidad de tales metodologías sobre datos reales del mercado mexicano, evaluando las bondades y desventajas en cada caso. Específicamente, se sientan bases útiles para el uso de técnicas alternativas de clasificación en portafolios de créditos con las características propias del mercado mexicano, contribuyendo para futuros desarrollos en la materia.

Entre las principales conclusiones obtenidas de los modelos calibrados sobre los distintos conjuntos de datos, sostienen que no existe un clasificador cuyo desempeño sea superior sobre el resto de las metodologías implementadas en este documento. Sin embargo, los resultados obtenidos con base en las medidas de desempeño utilizados (TC y AUC) indican que varios de los algoritmos pertenecientes al área de reconocimiento de patrones son tan competitivos como las técnicas estadísticas tradicionales. Más aún, los modelos híbridos *Bagging* y RBM + logística mantuvieron los niveles de desempeño más alto, por lo menos en alguna de las dos métricas con las que se desarrolló el estudio.

Como futuro trabajo de investigación, se plantean dos líneas principales. La primera se refiere al estudio y aplicación de técnicas de clasificación aún no exploradas en el área del CS, como por ejemplo, las denominadas redes de aprendizaje profundo (DBN), las cuales están formadas a su vez por varias capas de RBM, entre otras alternativas relativas al área de *deep learning*. Como segunda línea de investigación, se plantea la posibilidad de explorar la aplicabilidad de algunos de los clasificadores usados en este documento, para realizar un análisis dinámico del comportamiento de los clientes en el tiempo. Sobre este último aspecto, se pueden precisar varios enfoques del estudio. En primer lugar, se puede citar el análisis de la evolución del perfil de riesgo de los clientes¹⁴. Para ello, técnicas como el análisis de supervivencia y modelos ocultos de Markov (HMM) son un par de alternativas a considerar. Como segundo aspecto que resulta de gran importancia para la gestión de los portafolios de crédito, se considera la inclusión de componentes como la rentabilidad para extender la clasificación a distintos perfiles de clientes en del sentido rendimiento-riesgo que representan para las instituciones. Finalmente, un aspecto complementario a los dos puntos ya mencionados, se refiere a la construcción de escenarios del desempeño de los clientes con base en estrategias de mitigación de pérdidas (morosidad, cobranza y recuperación de cartera) que resulten de la detección de patrones y alertas tempranas bajo múltiples técnicas de clasificación.

4.6 Bibliografía

Alpaydin, E., Introduction to Machine Learning, The MIT Press. 2da. Ed., 2010.

Apilado, V.P., D.C. Warner y J.J. Dauten (1974) Evaluative techniques in consumer finance. Journal of Financial Quantitative Analysis, March, 275-283.

Alfo, M., S. Caiazza y G. Trovato (2005). Extending a Logistic Approach to Risk Modeling through Semiparametric Mixing. Journal of Financial Services Research, 28(1), 163-176.

¹⁴ Este aspecto resulta en principio relevante para las instituciones que desean evaluar los niveles de pérdida futuros en el portafolio, producto del deterioro de la calidad crediticia de las contrapartes y los eventos de incumplimiento. Asimismo, es de gran interés para las compañías, cuantificar los impactos que sobre los requerimientos de capital tiene la evolución del riesgo de los clientes.

- Altman, E.I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, 23(4), 589–609.
- Anderson, R. (2007). *The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation*. Oxford University Press.
- Antonakis, A.C. y M.E. Sfakianakis (2009). Assessing naive Bayes as a method for screening credit applicants. *Journal of Applied Statistics*, 36(5), 537-545.
- Arns, M.T., P.J. Steiner, N.Y. Soma, T. Shimizu y Nievola J.C. (2006). Using neural network rule extraction for credit-risk evaluation. *International Journal of Computer Science and Network Security*, 6(5A), 6-17.
- Atiya, A.F. (2001). Bankruptcy prediction for credit risk using neural networks: a survey and new results. *IEEE Transactions on Neural Networks*, 12(4), 929-935.
- Barniv, R., J.B. McDonald (1999). Review of Categorical Models for Classification issues in accounting and finance. *Review of Quantitative Finance and Accounting*, 13, 39-62.
- Bastos, J. A. (2008) Credit scoring with boosted decision trees. City: Munich Personal RePEc Archive, pp. 262-273.
- Baesens, B., M. Egmont-Petersen, R. Castelo y J. Vanthienen (2002). Learning Bayesian network classifiers for credit scoring using Markov Chain Monte Carlo search. In *Proceedings of the 16th International Conference on Pattern*, 3, 49-52.
- Baesens, B., R. Setiono, C. Mues y J. Vanthienen (2003a). Using neural network rule extraction and decision tables for credit risk evaluation. *Computer Journal of Management Science*, 49(3), 312-329.
- Baesens, B. T. Gestel, S. Viane, M. Stepanova, J. Suykens y J. Vanthienen (2003b). Benchmarking state of the art classification algorithms for credit scoring. *Computer Journal of Operational Research Society*, 54(3), 627-635.
- Banasik J., J.N. Crook y L.C. Thomas (1999). Not if but when borrowers default, *Journal of Operational Research Society*, 50, 1185-1190.
- Mun J., (2006) “Real Options Analysis: Tools and Technique for Valuing Strategic Investment and Decisions.” Nueva York, Nueva York, EUA: John Wiley & Sons.
- Trigeorgis L., (1999) “Real Options: A Primer,” en Alleman J. y E. Noam, eds. *The New Investment Theory of Real Options and Its Implications for Telecommunications Economics*. Boston, Mass. EUA.

- Bicer, I., D. Sevis, y T. Bilgic (2010); Bayesian credit scoring model with integration of expert knowledge and customer data; 24th Mini EURO Conference, pp.324-329.
- Borra, S. y S. Caiazza (2002). Comparative performance of credit scoring models using aggregated predictors. *Data Mining III*, WIT Press, 747-756.
- Breiman, L., Friedman, J.H., Olshen, R.A. y C.J. Stone (1984). *Classification and Regression Trees*. Wadsworth, Belmont, CA.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
- Caiazza, S. (2004). The comparative performance of credit scoring models: an empirical approach. In *Monetary Integration, Markets and Regulation, Research in Banking and Finance*, 4, 17-66.
- Chatterjee, S. y S. Barcun (1970). A nonparametric approach to credit screening, *Journal of the American Statistical Association*, 65, 150-154.
- Chen, M.C. y S.H. Huang (2003). Credit Scoring and Rejected Instances Reassigning through Evolutionary Computational Techniques. *Expert Systems with Applications*, 24(4), 433-441.
- Coffman, J.Y. (1986). The proper role of tree analysis in forecasting the risk behavior of borrowers, *MDS Reports, Management Decision Systems*, Atlanta, GA, 3-9.
- Crook, J.N., D.B. Edelman y L.C. Thomas (2007). Recent developments in consumer credit risk assessment; *European Journal of Operational Research*. 183(3), 1447-1465.
- Davis, R.H., D.B. Edelman, A.J. Gammerman (1992). Machine-learning algorithms for credit-card applications, *IMA Journal of Mathematics applied in Business and Industry*, 4, 43-52.
- Durand, D. (1941). *Risk Elements in Consumer Instalment Financing*. National Bureau of Economic Research, New York. Disponible en <http://www.nber.org/books/dura41-1>
- Drummond, C. y R. Holte (2006). Cost Curves: An Improved Method for Visualizing Classifier Performance. *Machine Learning*, 65, 95-130.
- Embrechts, P, A.J. McNeil y R. Frey (2005). *Quantitative Risk Management: concepts, techniques, and tools*. Princeton University Press, USA.
- Fan, A. y M. Palaniswami (2000). A new approach to corporate loan default prediction from financial statements. In *Proc. Computational Finance/Forecasting Financial Markets Conf. CF/FFM-2000*, London (CD), UK.
- Fantazzini, D. y S. Figini (2009). Random Survival Forests Models for SME Credit Risk Measurement. *Methodology and Computing in Applied Probability* 11(1), 29-45.

- Fisher, R.A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7, 179–188.
- Fix, E. y J. Hodges (1952). Discriminatory analysis, nonparametric discrimination, consistency properties. Report 4, Project 21-49-004, School of Aviation Medicine, Randolph Field, TX.
- Friedman, N., D. Geiger y M. Goldszmidt (1997). Bayesian network classifiers. *Machine Learning* 29, 131–163.
- Fukanaga, K. y T.E. Flick (1984). An optimal global nearest neighbour metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1, 25-37.
- Galindo, J. y P. Tamayo (2000). Credit risk assesment using statistical and machine learning: basic methodology and risk modelling applications. *Computational Economics*, 15, 107-143.
- Glenn, J.J. (2001). Classification accuracy in discriminant analysis: a mixed integer programming approach. *Journal of the Operational Research Society*, 52, 328-339.
- Grablowsky, B.J. y W.K. Talley (1981). Probit and discriminant functions for classifying credit applicants: A comparison: *Journal of Economics and Business*, 33, 254-261.
- Ghodselahe, A. (2011). A hybrid support vector machine ensemble model for credit scoring. *International Journal of Computer Applications*, 17(5).
- Hand, D.J. y W.E. Henley (1997a). Statistical Classification Methods in Customer Credit Scoring: A Review. *Journal of the Royal Statistical Society* 160(3): 523–541.
- Hand, D. J., K. McConway y E. Stanghellini (1997b). Graphical models of applicants for credit. *IMA Journal of Mathematics Applied in Business and Industry*, 8(2), pp. 143–155.
- Hand, D.J. and M.G. Kelly (2001). Lookahead scorecards for new fixed term credit products. *Journal of the Operational Research Society*, 52, 989-996.
- Hand, D.J. y R.J. Till (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning*, 45(2),171–186.
- Hand, D.J (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, 77,103-123.
- Henley, W.E. y D.J. Hand (1996). A k-NN classifier for assessing consumer credit risk. *Statistician*, 65, 77-95.
- Hinton, G. y T. Sejnowski (1983). Optimal perceptual inference. In *IEEE conference on Computer Vision and Pattern Recognition*.

- Hinton, G. E. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771-1800, 2002.
- Hoffmann F., B. Baesens, C. Mues, T. Gestel y J. Vanthienen, “Inferring Descriptive and Approximate Fuzzy Rules for Credit Scoring Using Evolutionary Algorithms”, *European Computer Journal of Operational Research*, 177(1), 2007, pp. 540-555.
- Hsieh, N.C. (2005). Hybrid mining approach in the design of credit scoring models. *Expert Systems with Applications*, 28(4):655-665.
- Huang, Z., H. Chen, C.J. Hsu, W.H. Chen y S. Wu (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems* 37(4), pp. 543-558.
- Huang, J., H. Tzeng y S. Ong (2006). Two stage genetic programming for the credit scoring. *Computer Journal of Applied Mathematics and Computation*, 14(3), 1039-1053.
- Huang, C.L., M.C. Chen y C.J. Wang (2007). Credit Scoring with a Data Mining Approach based on Support Vector Machines. *Expert Systems with Applications*, 33(4), pp. 847-856.
- Hull, J.C (2012). *Options, Futures and Other Derivatives*. Pearson Prentice Hall.
- Huysmans, J., B. Baesens, J. Vanthienen y T. Gestel (2006). *Computer Journal of Expert Systems with Applications*, 30(4), 479-487.
- Islam, M.J., Q.M.J. Wu, M. Ahmadi y M. Sid-Ahmed (2007). Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers. *International Conference on Convergence Information Technology*. IEEE Computer Society.
- Jensen, L. (1992). Using Neural Networks for Credit Card Accounts. *Computer Journal of Managerial Finance*, 18(15), 26-29.
- Johnson, R.W. (2004). Legal, Social, and Economic Issues in Implementing Scoring in the United States. In Thomas, Edelman, and Crook (eds) *Readings in Credit Scoring: Recent Developments, Advances, and Aims*, 5-15. Oxford University Press.
- Kass, G.V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29(2), 119–127.
- Kumar, K., y Bhattacharya, S. (2006). Artificial neural network vs linear discriminant analysis in credit ratings forecast: A comparative study of prediction performances. *Review of Accounting and Finance*, 5(3), 216-227.
- Laitinen, E.K. (1999). Predicting a corporate credit analyst’s risk estimate by logistic and linear models. *International Review of Financial Analysis* 8(2), 97-121.

- Lane, S. (1972). Submarginal credit risk classification. *Journal of Financial Quantitative Analysis*, June, 313-328.
- Lee, T.S. y I.F. Chen (2002). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28(4), 743-752.
- Lee, T.S., C.C. Chiu, C.J. Lu y I.F. Chen (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, 23(3), 245-254.
- Lee, T.S., C.C. Chiu, Y.C. Chou y C.J. Lu (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, 50, 1113-1130.
- Lewis, E.M. (1992). *An Introduction to Credit Scoring*, 2da. Ed. Athena Press, CA.
- Li, X.S. y Y.H. Guo (2006). Personal credit scoring models on naive Bayesian classifier. *Computer Engineering and Applications*, 42(1), 197-201.
- Li, F. C. (2009). The Hybrid Credit Scoring Model based on KNN Classifier. *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*. IEEE Computer Society.
- Li, H., J. Sun, y J. Wu (2010). Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. *Expert Systems with Applications*, 37(8), 5895-5904.
- Li, T., W. Shiue y W. Huang (2006). The evaluation of consumer loans using neural networks. *Omega Computer Journal*, 31(2), 83-96.
- Makowski, P. (1985). Credit scoring branches out. *Credit World*, 75, 30-37.
- Malhotra R. y K. Malhotra (2002). Differentiating between good credits and bad credits using neuro-fuzzy systems. *Computer Journal of Operational Research*, 36(2), 190-201.
- Martens, D., B. Baesens, T.V. Gestel y J. Vanthienen (2007). Comprehensive credit scoring models using rule extraction from support vector machines. *European Journal of Operational Research*, 183, 1466-1476.
- Marinakakis, Y., M. Marinaki, M. Doumpos, N. Matsatsinis y C. Zopounidis (2008). Optimization of nearest neighbor classifiers via metaheuristic algorithms for credit risk assessment. *Journal of Global Optimization* 42(2), 279-293.
- Martens, D., T.V. Gestel, D.B. Manu, R. Haesen, J. Vanthienen y B. Baesens (2008) Credit rating prediction using ant colony optimization. *Journal of the Operational Research Society*, 61, 561-573.

- Mileris R. y V. Boguslauskas (2010). Data reduction influence on the accuracy of credit risk estimation models. *Inzinerine Ekonomika-Engineering Economics*, 21(1), 5-11.
- Min, J.H. y Y.C. Lee (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28, 603-614.
- Moody's (2000). RiskCalc™ for private companies: Moody's default model. www.moody.com
- Moses, D. y S.S. Liao (1987). On developing models for failure prediction. *Journal of Commercial Bank Lending*, 69, 27-38.
- Mues, C., B. Baesens, C.M. Files y J. Vanthienen (2004). Decision diagrams in machine learning: an empirical study on real-life credit-risk data. *Expert Systems with Applications* 27(2), 257-264.
- Myers, J.H. y E.W. Forgy (1963). The development of numerical credit evaluation systems. *Journal of American Statistics Association*, 58, 799-806.
- Narain B. (1992). Survival Analysis and the credit granting decision. In *Credit Scoring and Credit Control*, L.C.Thomas, J.N.Crook, D.B.Edelman (eds.), 109-122, Oxford University Press, Oxford.
- Ong, C.S., J.J. Huang y G.H. Tzeng (2005). Building credit scoring models using genetic programming, *Expert Systems with Applications*, 29(2), 41-47.
- Ooghe H., H. Claus, N. Sierens N. y J. Camerlynck (1999). International comparison of failure prediction models from different countries: an empirical analysis. Working Paper, 99/79, Univesity of Ghen.
- Paredes, R., y E. Vidal (2000). A class-dependent weighted dissimilarity measure for nearest neighbor classification problems. *Pattern Recognition Letters* 21(12), 1027-1036.
- Pearl, J. (1988). *Probabilistic reasoning in Intelligent Systems: networks for plausible inference*. Morgan Kaufmann, San Francisco, CA.
- Piramuthu, S. (1999). Financial credit risk evaluation with neural and neuro fuzzy systema. *Computer Journal of Operational Research*, 112(6), 310-321.
- Platt, H.D. y M.D. Platt (1990). Development of a class of stable predictive variables: the case of bankruptcy prediction. *Journal of Business Finance and Accounting*, 17, 31-51.
- Quinlan, J.R. (1979). Discovering rules from large collections of examples: a case study. In D. Michie (Ed.), *Expert Systems in the Micro-electronic Age*, 68-201.
- Quinlan, J.R. (1986) Induction of decision trees. *Machine Learning* 1(1), 81-106.

- Quinlan, J.R. (1993). C4.5 programs for machine learning. Morgan Kaufmann, San Francisco, CA.
- Sarlija, N., M. Bencic y M.Z. Susac (2009). Comparison procedure of predicting the time to default in behavioural scoring. *Expert Systems with Applications*, 36(5), 8778–8788.
- Sexton, R. S., S. McMurtrey y D.J. Cleavenger (2006). Knowledge discovery using a neural network simultaneous optimization algorithm on a real world classification problem. *European Journal of Operational Research*, 168, 1009-1018.
- Shin, K.S., T.S. Lee y H.J. Kim (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1), 127-135.
- Sohn, S.Y. y H.W. Shin (2006). Reject inference in credit operations based on survival analysis. *Expert Systems with Applications*, 31(1), 26–29.
- Steenackers, A. y M.J. Goovaerts (1989). A credit scoring model for personal loans. *Insurance: Mathematics and Economics*, 8(1), 31-34.
- Stepanova, M. y L.C. Thomas (2002). Survival analysis methods for personal loan data. *Operations Research*, 50(2), 277-289.
- Suykens, J.A.K. and J. Vandewalle (1999). Least squares support vector machine classifiers. *Neural Process Letters*, 9, 293-300.
- Suykens, J.A.K., T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle (2002). *Least Squares Support Machines*. World Scientific Publishing Co., Pte, Ltd. Singapore.
- Taffler, R. J. y B. Abassi (1984). A Model for Predicting Debt Servicing Problems in Developing Countries. *Journal of the Royal Statistical Society*, 147(4), 541-568.
- Tapiero C.S. (2010). *Risk Finance and Asset Pricing: Value, Measurements, and Markets*. Wiley.
- Tang, T. C. y L.C. Chi (2005). Predicting multilateral trade credit risks: comparisons of Logit and Fuzzy Logic models using ROC curve analysis. *Expert Systems with Applications*, 28(3), 547-556.
- Thomas L.C., D.B. Edelman y J.N. Crook (2002). *Credit Scoring and its applications*. Society for Industrial and Applied Mathematics, Philadelphia.
- Tsai C.F. y M.L. Chen (2010). Credit rating by hybrid machine learning techniques. *Applied Soft Computing*, 10, pp.374-380.
- Turney, P. (2000). “Types of Cost in Inductive Concept Learning”. Presentado en Workshop on Cost-Sensitive Learning en Seventeenth International Conference on Machine Learning, Stanford University, Stanford, CA.

- Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley: New York.
- Wang, Y., S. Wang, S. y K.K. Lai (2005). A new fuzzy support vector machine to evaluate credit risk. *IEEE Transactions on Fuzzy Systems*, 13(6), 820-831.
- Wang, G., J. Hao, J. Ma y H. Jiang (2011). A comparative assessment of ensemble learning for credit scoring. *Expert systems with applications*, 38, 223-230.
- West, D. (2000). Neural Network Credit Scoring Models. *Computers and Operations Research*, 27(3), 1131-1152.
- West, D., S. Dellana y J. Qian (2005). Neural Network Ensemble Strategies for Financial Decision Applications. *Computer Journal of Operations Research*, 32(2), 2543-2559.
- Wiginton, J.C. (1980). A note on the comparison of logit and discriminant models of consumer credit behavior. *Journal of Quantitative Analysis*, 15, 757-770.
- Witten I.H., E. Frank y A.H. Mark (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, Morgan Kaufmann.
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5, 241–259.
- Zhao, H. (2007). A multi-objective genetic programming approach to developing Pareto optimal decision trees. *Decision Support Systems*, 43(3), 809-826.
- Zhu, H., P.A. Beiling y G.A. Overstreet (2001). A study in the combination of two consumer credit scores. *Journal of the Operational Research Society*, 52(9), 974-980.

Apéndice

Conjunto	Descripción	#Atributos (Num/Categorico)	#Instancias (Malos/Buenos)	Problema
Alemán	Información relativa a solicitantes de crédito proporcionada por Dr. Hans Hofmann	20 (7/13)	1,000 (300/700)	Estático: determinar el perfil de riesgo de nuevas solicitudes de crédito
Australiano	Otorgamiento de tarjetas de crédito proporcionado por quinlan@cs.su.oz.au omitiendo nombres y descripción atributos	14 (6/8)	690 (307/383)	Estático: Clasificación buenos y malos solicitantes

PAKDD 2009	Información utilizada en competencia de minería de datos presentada en PAKDD 2009 correspondiente a clientes a quienes se otorgaron tarjetas de crédito y donde el conjunto de entrenamiento y prueba son no adyacentes en el tiempo	22 (10/12)	50,000 (9874/40126)	Estático: Clasificación buenos y malos clientes
Cerveza	Registros de clientes de venta de cerveza algunos de los cuales fueron acreedores a línea de crédito durante 2009 y 2010	16 (2/14)	3,200 de los cuales a 2,156 se otorgó línea de crédito (949/1207)	Estático: Clasificación de clientes con y sin línea de crédito así como del perfil de riesgo de los acreditados
Autos	Información de créditos de auto otorgados entre 2006 y 2009	24 (17/7)	48,550 distintos créditos ¹⁵	Estático/Dinámico: Clasificación clientes con bueno y mal perfil, así como su análisis por distintas dimensiones: año de otorgamiento del crédito, morosidad en pagos vencidos, ventanas de observación móviles

Tabla 4.1. Resumen de los cinco conjuntos de datos disponibles

Atributo	Descripción	Medición
Cheques	Saldo cuenta actual en marcos alemanes (DM)	Ordinal; 4 valores: A11-Saldo a favor, A12-[0,200), A13-Mayor o igual a 200, A14-Sin cuenta de cheques
Plazo	Duración en meses	Entero
Historial	Historial crediticio	Ordinal; 5 valores: A30-Sin créditos/AI corriente, A31-AI corriente en este banco, A32-AI corriente en créditos actuales, A33-Pagos retrazados en el pasado, A34-Cuenta crítica/otros créditos en otros bancos
Proposito	Propósito del crédito	Nominal; 11 valores: A40-Auto nuevo, A41-Auto usado, A42-Muebles/Equipo, A43-Radio/Televisión, A44-Electrodomésticos, A45-Reparaciones, A46-Educación, A47-Vacaciones, A48-Reentrenamiento, A49-Negocios, A410-Otros
Monto	Monto del crédito	Intervalo
Ahorros	Saldo en cuentas o inversiones, en DM	Ordinal; 5 valores: A61-[0,100), A62-[100,500), A63-[500,1000), A64-Mayor o igual a 1000, A65-Desconocido/Sin cuenta de ahorros
TiempoEmpleo	Antigüedad en el empleo actual en años	Ordinal; 5 valores: A71-Desempleado, A72-(0,1), A73-[1,4), A74-[4,7), A75-Mayor o igual a 7
TasaPago	Tasa de pago como porcentaje del ingreso disponible	Intervalo
Personal	Estatus civil y sexo	Nominal; 5 valores: A91-hombre divorciado/separado, A92-mujer divorciada/separada/casada, A93-hombre soltero, A94-hombre casado/viudo, A95-mujer soltera
Deudores	Otros deudores/garantes	Nominal; 3 valores: A101-Ninguno, A102-Co solicitante, A103-Garante
TiempoResidencia	Tiempo en la residencia actual	Intervalo
Propiedad	Tipo de propiedad	Nominal; 4 valores: A121-Bienes raíces, A122-Ahorro vivienda, A123-Auto u otro no considerado en atributo 'Ahorros', A124-Desconocido/Sin propiedad
Edad	Edad en años	Entero
OtrosBienes	Otros planes de inversión	Nominal; 3 valores: A141-Banco, A142-Tienda, A143-Ninguno
Vivienda	Tipo de vivienda	Nominal; 3 valores: A151-Rentada, A152-Propia, A153-Gratis
NumCreditos	Número de créditos en el banco actual	Entero
Empleo	Tipo de empleo	Ordinal; 4 valores: A171-Desempleado/No residente sin capacitación, A172-Residente no capacitado, A173-Empleado Capacitado, A174-

¹⁵ La distribución Malos/Buenos depende de la definición de mal perfil de riesgo (incumplimiento) y el periodo al que correspondan las observaciones. El análisis dinámico identificará los efectos bajo las distintas dimensiones.

		Administrador/Autoempleado/Altamente calificado/Director
Dependientes	Número de dependientes económicos	Entero
Telefono	Estatus de número telefónico	Binaria: A191-No tiene, A192-Sí tiene y está registrado a su nombre
Extranjero	Trabajador extranjero	Binaria: A201-Sí, A202-No
Clase	Indicadora del perfil crediticio	Binaria: 0-Bueno, 1-Malo

Tabla 4.2. Atributos del conjunto Alemán

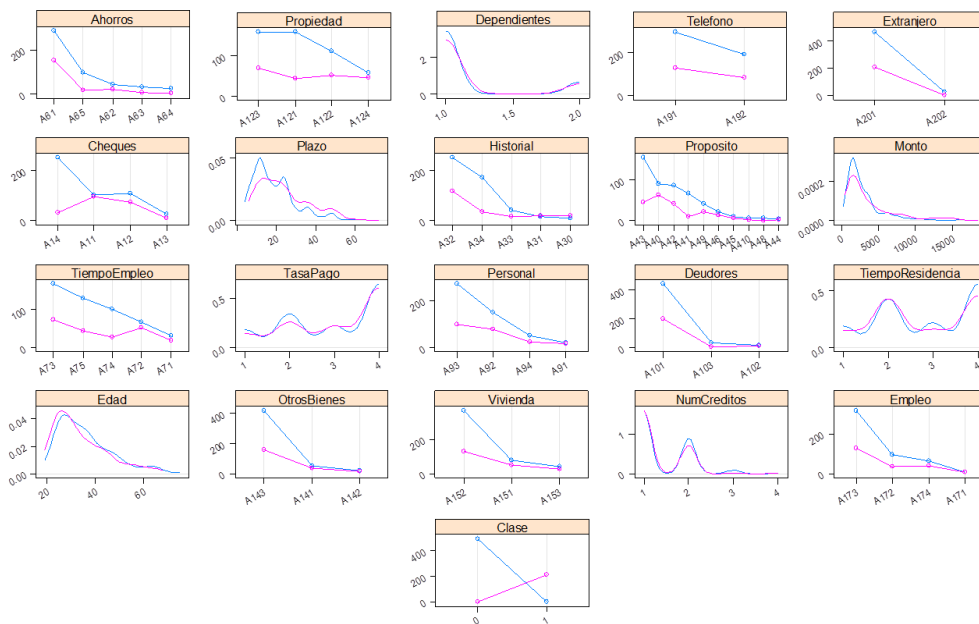


Figura 4.7. Diagramas de densidades (atributos numéricos) y frecuencia (atributos categóricos) del conjunto alemán por clase

Atributo	Medición	Tasa Faltantes
A1	Categórica; 2 valores: 0-a, 1-b	1.7%
A2	Intervalo	1.7%
A3	Intervalo	-
A4	Categórica; 3 valores: 1-p, 2-g, 3-gg	0.9%
A5	Categórica; 14 valores: 1-ff, 2-d, 3-i, 4-k, 5-j, 6-aa, 7-m, 8-c, 9-w, 10-e, 11-q, 12-r, 13-cc, 14-x	0.9%
A6	Categórica; 9 valores: 1-ff, 2-dd, 3-j, 4-bb, 5-v, 6-n, 7-o, 8-h, 9-z	1.3%
A7	Intervalo	1.3%
A8	Categórica; 2 valores: 0-t, 1-f	-

A9	Categoría; 2 valores: 0-t, 1-f	-
A10	Intervalo	-
A11	Categoría; 2 valores: 0-t, 1-f	-
A12	Categoría; 3 valores: 1-s, 2-g, 3-p	-
A13	Intervalo	1.9%
A14	Intervalo	-
A15	Binaria: 0-Bueno(''), 1-Malo('')	-

Tabla 4.3. Atributos conjunto de datos Australiano

Atributo	Descripción	Medición	Tasa Faltantes	
			Entrenamiento	Prueba
ID_SHOP	Código de la tienda donde se realizó la solicitud	Entero	-	-
SEX	Sexo	Binaria: M=Masculino, F=Femenino	0.006%	-
MARITAL_STATUS	Estatus civil	Nominal; 5 valores: S=Soltero, C=Casado, D=Divorciado, V=Viudo, O=Otro	-	-
AGE	Edad del aplicante	Entero	-	-
QUANT_DEPENDANTS	Número de dependientes económicos	Entero	100%	-
EDUCATION	Nivel de educación del solicitante	Nominal; 4 valores: 1, 2, 3, 4 (categorías no especificadas)	100%	10.51%
FLAG_RESIDENCIAL_PHONE	Cuenta con teléfono residencial	Binaria: Y=Si, N=No	-	-
AREA_CODE_RESIDENCIAL_PHONE	Código de área del teléfono residencial	Entero	-	-
PAYMENT_DAY	Día (fijo) de pago en el mes	Entero	-	-
SHOP_RANK	Calificación que la compañía asigna a la tienda	Nominal; 3 valores entrenamiento: 0, 2, 3 y 4 valores prueba: 0, A, B, C, D	-	-
RESIDENCE_TYPE	Tipo de residencia	Nominal; 4 valores: P=Propia, A=Rentada, C=DE los padres, O=Otra	-	0.36%
MONTHS_IN_RESIDENCE	Tiempo en residencia actual en meses	Entero	-	-
FLAG_MOTHERS_NAME	Se proporciona el nombre de la madre del solicitante	Binaria: Y=Si, N=No	-	-
FLAG_FATHERS_NAME	Se proporciona el nombre del padre	Binaria: Y=Si, N=No	-	-
FLAG_RESIDENCE_TOWN=WORKING_TOWN	Trabaja en el mismo pueblo donde vive	Binaria: Y=Si, N=No	-	-
FLAG_RESIDENCE_STATE=WORKING_STATE	Trabaja en el mismo estado donde vive	Binaria: Y=Si, N=No	-	-
MONTHS_IN_THE_JOB	Tiempo en empleo actual en meses	Entero	-	-
PROFESSION_CODE	Código de la profesión del aplicante	Entero	-	-
MATE_INCOME	Ingreso neto mensual de la pareja del aplicante en moneda brasileña (R\$)	Intervalo	-	-
FLAG_RESIDENCIAL_ADDRESS=POSTAL_ADDRESS	Recibe el correo postal en el domicilio donde vive	Binaria: Y=Si, N=No	-	-
FLAG_OTHER_CARD	Si cuenta con otro crédito o tarjeta de marca propia	Binaria: Y=Si, N=No	100%	-
QUANT_BANKING_ACCOUNTS	Número de cuentas bancarias del solicitante	Entero	100%	0.36%
PERSONAL_REFERENCE_#1	Nombre de la referencia personal #1 (en portugués)	Cadena	0.004%	1.45%

PERSONAL_REFERENCE_#2	Nombre de la referencia personal #2 (en portugués)	Cadena	10.40%	1.45%
FLAG_MOBILE_PHONE	Posee teléfono celular	Binaria: Y=Si, N=No	100%	100%
FLAG_CONTACT_PHONE	Posee teléfono de contacto	Binaria: Y=Si, N=No	100%	-
PERSONAL_NET_INCOME	Ingreso neto mensual del solicitante en moneda brasileña (R\$)	Intervalo	-	-
COD_APPLICATION_BOOT	Código de módulo donde se realizó la solicitud	Entero	100%	-
QUANT_ADDITIONAL_CARDS_IN_THE_APPLICATION	Número de tarjetas adicionales solicitadas	Entero	-	100%
FLAG_CARD_INSURANCE_OPTION	Solicita seguro de tarjeta	Nominal igual a N=No	100%	-
TARGET_LABEL_BAD=1	Perfil de riesgo	Binaria: 1-Malo, 0-Bueno	-	-

Tabla 4.4. Atributos conjunto de datos PAKDD 2009

Atributo	Descripción	Medición	Tasa Faltantes
Clave	Cliente clasificado como clave (por alto volumen de ventas)	Binaria: 1-Clave, 0-No Clave	7.8%
Imagen	Cliente clasificado como imagen (por ubicación o tipo de sector que atiende)	Binaria: 1-Cliente Imagen, 0-No Imagen	13.6%
Exclusivo	Estatus de exclusividad de la marca en los productos de cerveza que el cliente ofrece	Binario: 1-Exclusivo, 0-Mixto	-
Esquina	Ubicación del establecimiento	Binaria: 1-Esquina, 0-Otra ubicación	13.5%
Calle	Tipo de calle	Nominal; 3 valores: 1-Principal, 2-Secundaria, 3-Carretera	13.5%
Zona	Tipo de zona	Nominal; 4 valores: 1-Comercial, 2-Habitacional, 3-Industrial, 4-Mixta	13.5%
Trafico	Nivel de tráfico vehicular a que se expone el establecimiento	Ordinal; 3 valores: 1-Alto, 2-Medio, 3-Bajo	-
Dimension	Dimensión de la fachada del establecimiento en metros	Intervalo	-
Local	Tipo de local	Nominal; 3 valores: 1-Empresa, 2-Propio, 3-Rentado	13.5%
Region	Estado de la república donde se ubica el establecimiento	Nominal; 3 valores: 1-Chiapas, 2-Campeche, 3-Tabasco	-
Giro	Giro del establecimiento	Nominal; 8 valores: 1-Abarrotes, 2-Cantina, 3-Cervecería, 4-Depósito, 5-Evento aire libre, 6-Minisuper, 7-Restaurante, 8-Otros	-
Control	Grado de control o influencia en el establecimiento	Ordinal; 5 valores: 1-Blindado, 2-Alto, 3-Mediano, 4-Bajo, 5-Sin control	0.3%
Tamaño	Tamaño del cliente (en función de sus ventas)	Ordinal; 6 valores: 1-Exitoso, 2-Productivo, 3-Promedio, 4-Cobertura, 5-Apoyo, 6-Minimo	12.6%
NivelSocEcon	Nivel socioeconómico de la zona donde se ubica el establecimiento	Ordinal; 3 valores: 1-Nivel AB, 2-Nivel C, 3-Nivel DE	-
Resultado	Ventas totales realizadas al cliente (kilolitros)	Intervalo	-
GEC	Tipo de cliente en función del promedio de ventas	Ordinal; 5 valores: 1-Platino, 2-Oro, 3-Plata, 4-Bronce, 5-Customizado	-
SVencido90	Indicadora de incumplimiento (saldo vencido 90+ días)	Binario: 1-Incumplido, 0-No incumplido	-
LineaCredito	Indicadora de otorgamiento de línea de crédito al cliente	Binario: 1-Con crédito, 0-Sin crédito	-

Tabla 4.5. Atributos conjunto de datos Cerveza

Atributo	Distribución								
	0	1	2	3	4	5	6	7	8
Clave	76%	24%							
Imagen	94%	6%							
Exclusivo	21%	79%							
Esquina	84%	16%							
Calle		23%	33%	44%					
Zona		15%	79%	1%	5%				
Trafico		31%	27%	42%					
Local		2%	61%	37%					
Region		26%	1%	74%					
Giro		17%	4%	14%	25%	4%	9%	12%	15%
Control		16%	7%	5%	21%	50%			
Tamaño		5%	16%	27%	16%	37%			
NivelSocEcon		2%	77%	21%					
GEC		7%	34%	37%	12%	11%			
LineaCredito	33%	67%							
SaldoVencido90 ¹⁶	56%	44%							

Tabla 4.6. Distribución de los atributos categóricos conjunto Cerveza

	Dimensión	Resultado
Min	1.00	0.04
Max	650.00	29,169.91
Promedio	11.09	496.86
Desv Est	20.38	1,142.17

Tabla 4.7. Estadísticas descriptivas de los atributos numéricos conjunto Cerveza

¹⁶ Por definición, las proporciones de clientes con saldo vencido son calculadas sobre el universo de clientes a quienes se les otorgó línea de crédito

Atributo	Descripción	Medición	Tasa Faltantes
educacionnueva	Máximo nivel de estudios	Nominal; 9 valores: 21-Primaria, 22-Secundaria, 23-Preparatoria, 24-Carrera, 25-Universidad, 26-Diplomados, 27-Posgrado, 28-Maestria, 29-Doctorado	
EDOCIVIL	Estado civil	Nominal; 5 valores: 0-Soltero, 1-Divorciado, 2-Viudo, 3-Unión libre, 4-Casado	
SEXO	Género	Nominal; 2 valores: 0-Masculino, 1-Femenino	
EDAD	Años cumplidos al solicitar el crédito	Entero	
DEPENDIENTES	Número de personas que dependen del solicitante	Intervalo	
INGRESOS	Monto de ingresos mensuales	Intervalo	0.7%
HIPOTECA	Cuenta con crédito hipotecario	Binario: 0-No tiene, 1-Si tiene	
INGRESOSNETOS	Monto de ingresos netos mensuales	Intervalo	0.7%
OTROSINGRESOS	Otros ingresos (mensuales)	Intervalo	0.2%
DEUDAS	Monto de obligaciones (deudas) actuales	Intervalo	0.1%
DOMICILIACION	Cuenta con domiciliación de pagos	Binario: 0-No tiene, 1-Si tiene	
POP	Probabilidad de pago al solicitar el crédito	Intervalo	
SCORE	Score calculado al solicitar crédito	Intervalo	
plazo	Plazo del crédito en meses	Intervalo	
montofin	Monto a financiar	Intervalo	
mensualidad	Monto de la mensualidad	Intervalo	
Precio	Precio del vehículo	Intervalo	
Enganchep	Tasa de enganche (%)	Intervalo	
TasaSTD	Tasa teórica a la que le prestaría la institución (%)	Intervalo	
TasaCliente	Tasa a la que la institución prestó al cliente (%)	Intervalo	
Marca	Marca del automóvil a financiar	Nominal; 8 valores: 1-FORD, 2-CHEVRO, 3-LAND R, 4-LAND ROVER, 5-LINCOLN, 6-MAZDA, 7-SEAT, 8-VOLVO	
año	Año de inicio del contrato	Intervalo; en formato aaaa	
mes	Mes de inicio del contrato	Intervalo; en formato mm	
subsidiado	Se otorgó subsidio en el préstamo	Binaria: 0-Ni tiene, 1-Si tiene	

Tabla 4.8. Atributos conjunto de datos Autos (Base socio-demográfica)

Atributo	Descripción
año	Año del reporte
mes	Mes del reporte
dpd	Días de saldo vencido
pagadas	Mensualidades pagadas
remterm	Mensualidades faltantes
saldoinsoluto	Saldo insoluto
perdida	Indicadora de incumplimiento (+90días de mora)

Tabla 4.9. Atributos conjunto datos Autos (Base comportamiento)

Capítulo 5

El riesgo del Sistema Financiero Internacional sobre el Crecimiento

Salvador Rivas-Aceves

Centro de Investigación Avanzada en Finanzas, UAEH

Claudia Estrella Castillo Ramírez

Centro de Investigación Avanzada en Finanzas, UAEH

José Francisco Martínez Sánchez

Centro de Investigación Avanzada en Finanzas, UAEH

Resumen.

Mediante un modelo de crecimiento endógeno estocástico, de economía abierta, con agentes económicos que inicialmente tienen dotaciones iniciales distintas pero preferencias iguales entre sí y con un sistema financiero, que pone a disposición de los agentes económicos diversos instrumentos para la toma de sus decisiones, se caracteriza el equilibrio macroeconómico y se verifica la existencia de una relación entre el sistema financiero y el sector real de la economía. Asimismo, se analiza el impacto que tiene la tasa de interés internacional sobre la tasa de crecimiento económico, la cual está directamente relacionada con los rendimientos medios esperados de los instrumentos financieros aquí analizados.

5.1 Introducción

El vínculo entre el sistema financiero y el sector real de la economía ha sido tema de gran relevancia en los últimos años, sobre todo a raíz de la crisis financiera y económica de tipo mundial que se originó a finales del 2008 en Estados Unidos de Norte América. Debido a dicha crisis, la relación entre el sistema financiero y la economía real se manifestó de una manera muy violenta y con consecuencias funestas. Desde un inicio, con el papel cada vez más importante del sistema financiero, y sobre todo en épocas de crisis financieras o cambiarias, los estudios teóricos pioneros que mostraron relación entre el sector financiero y el sector real fueron desarrollados por Bagehot (1968), Gurley y Shaw (1955) y McKinnon (1973) y muestran que un sistema financiero bien desarrollado puede generar impactos positivos sobre la economía real.

Si el sistema financiero presenta una buena estructura económica, es decir si es eficiente y cumple la función de intermediación financiera, entonces el sector financiero debe promover la acumulación del capital físico y del capital humano bajo la presencia de restricciones al crédito, y así estimular el crecimiento, de acuerdo con Banerjee y Newman (1993), Galor y Zeira (1993), Aghion y Bolton (1997) y Piketty (1997). En contra parte, las restricciones al crédito en realidad pueden impedir el proceso de acumulación de capital y, por lo tanto, tener un impacto negativo sobre el crecimiento económico, véase Benabou (1993, 1996a, 1996b), Durlauf (1996a, 1996b), Fernández y Rogerson (1994, 1996), Kremer y Maskin (1994) y Rivas-Aceves y Martínez (2009, 2011a, 2011b).

La presente investigación muestra que existe un vínculo entre el sistema financiero y el sector real de la economía, dicho vínculo se da a través de los instrumentos que el sistema financiero pone a disposición de los agentes económicos los cuales tienen efectos sobre la tasa de crecimiento de la economía. Se encuentra que los rendimientos medios esperados tienen un impacto positivo sobre la tasa de crecimiento, mientras que las volatilidades asociadas a estos instrumentos, que se interpretan como riesgo, generan efectos negativos sobre el crecimiento. Asimismo, se encuentra que los rendimientos medios esperados en los instrumentos financieros están directamente relacionados con la tasa de interés internacional y que un incremento en la misma afecta directa y positivamente a la tasa de crecimiento de la economía.

Para mostrar los resultados obtenidos, el presente documento se organiza de la siguiente manera. El segundo apartado establece las condiciones básicas de la economía y analiza el comportamiento de los hogares, mientras que el tercer apartado estudia el comportamiento de las

empresas. Por su parte, el cuarto apartado establece las condiciones bajo las cuales el sistema financiero opera y con todo ello se analiza el comportamiento global de la economía en el apartado cinco, para poder establecer el equilibrio macroeconómico mostrado en el apartado seis. La relación existente entre el sector financiero y el sector real de la economía es estudiada con más detalle en el apartado siete, mientras que en el apartado ocho se presentan las conclusiones. Por último, las referencias y un apéndice matemático son incluidos en los apartados nueve y diez.

5.2 Los hogares

Suponga que existe una economía abierta que está habitada por N número de personas, las cuales tienen las mismas preferencias pero dotaciones iniciales distintas y divididas en dos sectores, los hogares y las empresas. Respecto a los hogares, éstos están compuestos por consumidores que buscan maximizar el valor esperado de su utilidad debido al consumo del bien perecedero, el cual está medido por:

$$E \int_0^{\infty} u(c_t) e^{-\rho t} dt, \quad (5.1)$$

en donde la función de utilidad $u(c_t)$ cumple con que $u'(c_t) > 0$ y $u''(c_t) < 0$, en otras palabras, presenta rendimientos marginales decrecientes. Bajo esta definición, el consumo per cápita está representado por c_t y ρ es la tasa subjetiva de descuento que mide las preferencias por el consumo presente del individuo representativo. Suponga una utilidad logarítmica del tipo $u(c_t) = \ln c_t$ y, además, que el consumidor tiene acceso a información para la toma de decisiones, la cual está disponible en cada momento t . En consecuencia, la función de utilidad de tipo von Neumann-Morgenstern separable en $t = 0$ que mide dicho comportamiento es:

$$V_0 = E \left\{ \int_0^{\infty} \ln c_t e^{-\rho t} dt \middle| \mathcal{F}_0 \right\}, \quad (5.2)$$

con \mathcal{F}_0 como el conjunto de información inicial o las condiciones iniciales de las variables económicas, el cual será definido más adelante. El bien de consumo es comercializable tanto en el interior como en el exterior de la economía, gracias a la ausencia de barreras arancelarias. Lo anterior implica que dicho bien puede ser adquirido a un precio determinado por la Paridad del Poder de Compra, es decir:

$$P_t = P_t^* E_t, \quad (5.33)$$

en donde P_t representa el precio nacional, P_t^* es el precio internacional y E_t es el tipo de cambio nominal, con E_0 conocido. Si se supone que el precio internacional del bien es igual a la unidad, entonces los agentes pueden conocer la tasa esperada de inflación determinada por la dinámica del tipo de cambio, la cual es:

$$\frac{dP_t}{P_t} = \frac{dE_t}{E_t} = \varepsilon dt + \lambda dU_t. \quad (5.4)$$

La tasa de devaluación media esperada está dada por ε , mientras que λ mide la dispersión en el tipo de cambio y U_t es una variable aleatoria, que se distribuye de manera normal, con incrementos temporales cuya media es cero y varianza igual al incremento temporal. Es bajo estas condiciones que los consumidores toman decisiones.

5.3 Las empresas

Los productores utilizan todo el capital disponible, k_t , para producir el bien de consumo, y se considera disponible en el sentido de que es aquel que no ha sido consumido, lo que implica que el bien de consumo puede utilizarse como bien de capital. Asimismo, la producción per cápita, que está cuantificada por y_t , se lleva a cabo a través de una tecnología dada para todos los productores, de tal forma que las condiciones de producción son:

$$dy_t = k_t \alpha (dt + \sigma_y dW_t + \nu_y dZ_t), \quad (5.5)$$

en donde α mide la media esperada de la productividad marginal del capital, σ_y la dispersión en la productividad esperada y ν_y un salto en la productividad. Por su parte, dW_t es otro proceso de Wiener definido en un espacio fijo de probabilidad cuya filtración aumentada cumple que $(\Omega, \mathcal{F}, (F_t)_{t \geq 0}, \mathbb{P})$, con incrementos temporales independientes, media cero y varianza igual al incremento temporal, mientras que dZ_t es un proceso de Poisson que caracteriza la dinámica del salto en la productividad con una intensidad θ_y de manera que:

$$\begin{cases} \Pr(\text{un salto durante } dt) = \Pr(dZ_t = 1) = \theta_y dt, & (5.6) \\ \Pr(\text{ningun salto durante } dt) = \Pr(dZ_t = 0) = 1 - \theta_y dt + o(dt), & (5.7) \\ \Pr(\text{más de un salto durante } dt) = \Pr(dZ_t > 1) = o(dt). & (5.8) \end{cases}$$

Se supone adicionalmente que dW_t y dZ_t no están correlacionados, que el número inicial de saltos en la productividad es $Z_0 = 0$, que existe un nivel de capital inicial positivo, k_0 , y además

se considera que $o(dt)$ mide efectos de las variables en cuestión que persistan más allá de $t + dt$, los cuales son insignificantes y con $o(dt)/t \rightarrow 0$ cuando $t \rightarrow 0$.

5.4 El sector financiero

Para poder solventar las decisiones de consumo y de producción de los agentes, existe un sistema financiero que pone a disposición de los agentes económicos un conjunto de instrumentos financieros. En consecuencia, los agentes poseen tres formas para poder financiar su consumo, dos de estos instrumentos generan una vinculación con el sector financiero y los agentes, mientras que el último con el sector real y los agentes. Los dos primeros son activos uno sin riesgo y otro con riesgo, mientras que el último instrumento es un crédito bancario, y todos pueden ser adquiridos de manera nacional o internacional.

5.4.1 Activo sin riesgo

Suponga que un agente económico tiene acceso a una cuenta de inversión en donde mantiene recursos monetarios, los cuales le generan un rendimiento constante μ_S por unidad de tiempo sin riesgo alguno, de esta manera el activo sin riesgo S_t está representado por:

$$dS_t = S_t \mu_S dt. \quad (5.9)$$

Es importante señalar que μ_S representa un rendimiento de tipo determinista y que la condición inicial de este activo es $S_0 > 0$.

5.4.2 Activo con riesgo

Por otro lado, los agentes pueden tener acceso a un activo con riesgo cuya dinámica en su rendimiento está dada por:

$$dR_t = R_t(\mu_R dt + \sigma_R dX_t), \quad (5.10)$$

en donde R_t es el activo con riesgo y tiene una condición inicial $R_0 > 0$, además μ_R es el rendimiento medio esperado, σ_R la volatilidad asociada y X_t nuevamente es otro proceso de Wiener definido en un espacio fijo de probabilidad cuya filtración aumentada cumple que $(\Omega, \mathcal{F}, (F_t)_{t \geq 0}, \mathbb{P})$, con incrementos temporales independientes, media cero y varianza igual al incremento temporal.

5.4.3 Crédito bancario

Suponga que el sector productivo de la economía ahora está compuesto por dos tipos de productores, a saber: un primer productor que posee el capital necesario para realizar sus actividades de producción, más un remanente resultado de sus superávits registrados en periodos previos, y un segundo productor que no posee el capital necesario para producir.

Lo anterior implica que del total de la población de la economía, N , que se supone permanece constante periodo tras periodo, los dueños de los medios de producción son de dos tipos, los prestamistas (l) y los prestatarios (b), es decir, $N = N^l + N^b$. Los agentes prestamistas tienen un stock de capital mayor al de los agentes prestatarios y de manera conjunta componen el stock de capital total de la economía, de tal suerte que $K_t = K_t^l + K_t^b$ con $K_t^l > K_t^b$. Ambos agentes representativos, productor prestamista y prestatario, necesitan un nivel k_t para llevar a cabo el proceso productivo (véase ecuación 5), el primero de ellos posee:

$$k_t^l = k_t + \tilde{k}_t, \quad (5.11)$$

en donde \tilde{k}_t mide el nivel de capital adicional con $\tilde{k}_0 > 0$, mientras que el segundo sólo posee menos de lo necesario para producir. Bajo estas condiciones, el productor prestamista puede decidir otorgar bajo la forma de crédito su capital adicional, dando lugar al sistema financiero. Lo anterior permite que el productor prestatario acuda a solicitar en préstamo el capital que le falta para realizar sus actividades de producción. Se supone que la cantidad de capital que el agente prestamista coloca en el sistema financiero es exactamente igual a la cantidad de capital requerida por el agente prestatario, de tal forma que no existen recursos ociosos en la economía. Esto implica que para producir el productor prestatario necesita:

$$k_t = k_t^b + \tilde{k}_t. \quad (5.12)$$

Dicho crédito tiene una dinámica determinada por su costo, de tal forma que el capital \tilde{k}_t puede ser adquirido pagando un costo $\delta > 0$ de la forma siguiente:

$$d\tilde{k}_t = \tilde{k}_t \delta (dt + \sigma_\delta dV_t + \nu_\delta dQ_t), \quad (5.13)$$

en donde δ es el rendimiento medio esperado del crédito, σ_δ representa la volatilidad en el rendimiento, ν_δ un salto en el rendimiento medio esperado, dV_t es un proceso de Wiener con comportamientos similares a los previamente definidos, mientras que dQ_t es un nuevo proceso de

Poisson que caracteriza la dinámica del salto en el rendimiento del crédito, con una intensidad θ_δ de manera que:

$$\left\{ \begin{array}{l} \Pr(\text{un salto durante } dt) = \Pr(dQ_t = 1) = \theta_\delta dt, \end{array} \right. \quad (5.14)$$

$$\left\{ \begin{array}{l} \Pr(\text{ningun salto durante } dt) = \Pr(dQ_t = 0) = 1 - \theta_\delta dt + o(dt), \end{array} \right. \quad (5.15)$$

$$\left\{ \begin{array}{l} \Pr(\text{más de un salto durante } dt) = \Pr(dQ_t > 1) = o(dt). \end{array} \right. \quad (5.16)$$

Los procesos dV_t y dQ_t no están correlacionados y el número inicial de saltos en el rendimiento del capital es $Q_0 = 0$. En consecuencia, el prestamista obtendrá al final del periodo $\tilde{k}_t + dR_k \tilde{k}_t = \tilde{k}_t(1 + dR_k)$ y evidentemente, el prestatario pagará esa misma cantidad.

5.4.5 Comportamiento global del sistema financiero

Debido a que estos todos los instrumentos financieros pueden ser adquiridos a nivel nacional o internacional, entonces es fácil deducir que el comportamiento de los rendimientos medios esperados de cada instrumento financiero está determinado por las ecuaciones (5.3), (5.4), (5.9), (5.10) y (5.13), de tal forma que:

$$\mu_R = \mu_S = \delta = i^* + \varepsilon, \quad (5.17)$$

con i^* como la tasa de interés nominal internacional determinada de manera exógena.

5.5 La economía en conjunto

Para cerrar el modelo es necesario establecer un par de condiciones más. Primero, que todos los agentes económicos mantienen un acervo nominal de dinero nacional o saldos reales constante (m_t) con el objetivo de realizar sus operaciones de consumo, inversión y producción, cuyo un retorno real resulta de aplicar el lema de Itô a la ecuación (5.4), es decir:

$$\frac{dm_t}{m_t} = \frac{d\left(\frac{M_t}{P_t}\right)}{\frac{M_t}{P_t}} = \frac{d\left(\frac{1}{P_t}\right)}{\frac{1}{P_t}} = P_t d\left(\frac{1}{P_t}\right) = (-\varepsilon + \lambda^2)dt - \lambda dU_t. \quad (5.18)$$

Y segundo, suponer que los agentes económicos toman de manera simultánea decisiones de producción y de consumo, en otras palabras, que los hogares son los dueños de los medios de producción, generando así dos agentes representativos a saber: el consumidor-productor prestamista y el consumidor-productor prestatario. En consecuencia, la acumulación del capital

depende del nivel de riqueza que cada agente posea, el cual está determinado por los instrumentos financieros a su alcance, por su nivel de producción y por su nivel de consumo, por lo que:

$$a_t^l = y_t + S_t + R_t + m_t + \tilde{k}_t - c_t. \quad (5.19)$$

$$a_t^b = y_t + S_t + R_t + m_t - \tilde{k}_t - c_t. \quad (5.20)$$

Se define un nivel de riqueza inicial a_0 de manera exógena. Lo anterior implica que las dinámicas en la acumulación de la riqueza correspondientes son:

$$\frac{da_t^l}{a_t^l} = \omega_{1t} \frac{dk_t}{k_t} + \omega_{2t} \frac{dS_t}{S_t} + \omega_{3t} \frac{dR_t}{R_t} + \omega_{4t} \frac{dm_t}{m_t} + \omega_{5t} \frac{d\tilde{k}_t}{\tilde{k}_t} - \frac{c_t^l}{a_t^l}, \quad (5.21)$$

$$\frac{da_t^b}{a_t^b} = \omega_{1t} \frac{dk_t}{k_t} + \omega_{2t} \frac{dS_t}{S_t} + \omega_{3t} \frac{dR_t}{R_t} + \omega_{4t} \frac{dm_t}{m_t} - \omega_{5t} \frac{d\tilde{k}_t}{\tilde{k}_t} - \frac{c_t^b}{a_t^b}, \quad (5.22)$$

$$\omega_{1t} = \frac{k_t}{a_t}, \quad \omega_{2t} = \frac{S_t}{a_t}, \quad \omega_{3t} = \frac{R_t}{a_t}, \quad \omega_{4t} = \frac{\tilde{k}_t}{a_t}, \quad \omega_{5t} = \frac{m_t}{a_t}, \quad (5.23)$$

$$\omega_{1t} + \omega_{2t} + \omega_{3t} + \omega_{4t} + \omega_{5t} = 1. \quad (5.24)$$

Al sustituir las ecuaciones (5.5), (5.9), (5.10), (5.13) y (5.18) en (5.21) y (5.22) se encuentran las identidades contables que determinan la dinámica de la acumulación de la riqueza en la economía siguiente:

$$\begin{aligned} \frac{da_t^l}{a_t^l} = & \omega_{1t} \alpha (dt + \sigma_y dW_t + \nu_y dZ_t) + \omega_{2t} \mu_S dt + \omega_{3t} (\mu_R dt + \sigma_R dX_t) \\ & + \omega_{4t} [(-\varepsilon + \lambda^2) dt - \lambda dU_t] + \omega_{5t} \delta (dt + \sigma_\delta dV_t + \nu_\delta dQ_t) - \frac{c_t^l}{a_t^l}, \end{aligned} \quad (5.25)$$

$$\begin{aligned} \frac{da_t^b}{a_t^b} = & \omega_{1t} \alpha (dt + \sigma_y dW_t + \nu_y dZ_t) + \omega_{2t} \mu_S dt + \omega_{3t} (\mu_R dt + \sigma_R dX_t) \\ & + \omega_{4t} [(-\varepsilon + \lambda^2) dt - \lambda dU_t] - \omega_{5t} \delta (dt + \sigma_\delta dV_t + \nu_\delta dQ_t) - \frac{c_t^b}{a_t^b}. \end{aligned} \quad (5.26)$$

Equivalentemente,

$$\begin{aligned} \frac{da_t^l}{a_t^l} = & \left[\delta + \omega_{1t} (\alpha - \delta) + \omega_{2t} (\mu_S - \delta) + \omega_{3t} (\mu_R - \delta) + \omega_{4t} (\lambda^2 - \varepsilon - \delta) - \frac{c_t^l}{a_t^l} \right] dt \\ & + [\delta \sigma_\delta + \omega_{1t} (\alpha \sigma_y - \delta \sigma_\delta) - \omega_{2t} \delta \sigma_\delta + \omega_{3t} (\sigma_R - \delta \sigma_\delta) - \omega_{4t} (\lambda + \delta \sigma_\delta)] dW_t \\ & + [\delta \nu_\delta + \omega_{1t} (\alpha \nu_y - \delta \nu_\delta) - \omega_{2t} \delta \nu_\delta - \omega_{3t} \delta \nu_\delta - \omega_{4t} \delta \nu_\delta] dQ_t, \end{aligned} \quad (5.25')$$

$$\begin{aligned} \frac{da_t^b}{a_t^b} = & \left[\omega_{1t}(\alpha + \delta) + \omega_{2t}(\mu_S + \delta) + \omega_{3t}(\mu_R + \delta) + \omega_{4t}(\lambda^2 - \varepsilon + \delta) - \delta - \frac{c_t^b}{a_t^b} \right] dt \\ & + [\omega_{1t}(\alpha\sigma_y + \delta\sigma_\delta) + \omega_{2t}\delta\sigma_\delta + \omega_{3t}(\sigma_R + \delta\sigma_\delta) + \omega_{4t}(\delta\sigma_\delta - \lambda) - \delta\sigma_\delta] dW_t \\ & + [\omega_{1t}(\alpha\nu_y + \delta\nu_\delta) + \omega_{2t}\delta\nu_\delta + \omega_{3t}\delta\nu_\delta + \omega_{4t}\delta\nu_\delta - \delta\nu_\delta] dQ_t. \end{aligned} \quad (5.26')$$

De las ecuaciones anteriores se deduce que la acumulación de la riqueza para cada uno de los agentes económicos depende de la productividad media esperada del capital, de los rendimientos medios esperados de todos los instrumentos financieros, de la volatilidad asociada al instrumento con riesgo y al crédito bancario, y de los saltos en la productividad y en el crédito bancario.

5.5.1 Equilibrio macroeconómico

Bajo estas condiciones, tanto el agente prestamista como el agente prestatario realizan su toma de decisiones basándose en la misma *sigma algebra* o conjunto de información disponible en $t = 0$, el cual es :

$$\mathcal{F}_0 = \{E_0, P_0, k_0, Z_0, S_0, R_0, Q_0\}. \quad (5.27)$$

De esta manera, el equilibrio macroeconómico depende de los equilibrios resultantes de las condiciones establecidas en (5.2) y (5.25') para el agente prestamista, y de (5.2) y (5.26') para el agente prestatario.

5.5.2 Equilibrio del agente prestamista

La condición *Hamilton-Jacobi-Bellman (HJB)* asociada al problema de control óptimo establecido por (5.2) y (5.25'), igualada a cero es la siguiente¹:

$$\begin{aligned} 0 = & \ln c_t^l - \rho V(a_t^l) \\ & + a_t^l V'(a_t^l) \left(\delta + \omega_{1t}(\alpha - \delta) + \omega_{2t}(\mu_S - \delta) + \omega_{3t}(\mu_R - \delta) + \omega_{4t}(\lambda^2 - \varepsilon - \delta) - \frac{c_t^l}{a_t^l} \right) \\ & + \frac{1}{2} V''(a_t^l) a_t^{l2} [\delta\sigma_\delta + \omega_{1t}(\alpha\sigma_y - \delta\sigma_\delta) - \omega_{2t}\delta\sigma_\delta + \omega_{3t}(\sigma_R - \delta\sigma_\delta) - \omega_{4t}(\lambda + \delta\sigma_\delta)]^2 \\ & + [V(a_t^l) [\delta\nu_\delta + \omega_{1t}(\alpha\nu_y - \delta\nu_\delta) - \omega_{2t}\delta\nu_\delta - \omega_{3t}\delta\nu_\delta - \omega_{4t}\delta\nu_\delta] - V(a_t^l)] dQ_t. \end{aligned} \quad (5.28)$$

5.La ecuación anterior tiene las siguientes condiciones de optimalidad:

¹ véase el apéndice matemático.

$$\begin{aligned}\frac{\partial HJB}{\partial c_t^l} &= \frac{1}{c_t^l} - V'(a_t^l) = 0 & V''(a_t^l) &= -\frac{\beta_1^l}{a_t^{l2}} \\ V(a_t^l) &= \beta_0^l + \beta_1^l \ln a_t^l & c_t^l &= \frac{a_t^l}{\beta_1^l} \\ V'(a_t^l) &= \frac{\beta_1^l}{a_t^l}\end{aligned}$$

Al sustituir las condiciones de optimalidad anteriores en la ecuación (5.28) y suponer que $\beta_1^l = 1/\rho$, se obtiene que:

$$\begin{aligned}0 &= \ln \rho - \rho \beta_0^l \\ &+ \frac{1}{\rho} [\delta + \omega_{1t}(\alpha - \delta) + \omega_{2t}(\mu_S - \delta) + \omega_{3t}(\mu_R - \delta) + \omega_{4t}(\lambda^2 - \varepsilon - \delta) - \rho] \\ &- \frac{1}{2\rho} [\delta \sigma_\delta + \omega_{1t}(\alpha \sigma_y - \delta \sigma_\delta) - \omega_{2t} \delta \sigma_\delta + \omega_{3t}(\sigma_R - \delta \sigma_\delta) - \omega_{4t}(\lambda + \delta \sigma_\delta)]^2 \\ &+ \frac{1}{\rho} \{(\alpha \theta_y - \delta \theta_\delta) \omega_{1t} \ln[1 + (\alpha v_y - \delta v_\delta)] \\ &- (\omega_{2t} + \omega_{3t} + \omega_{3t} - 1) \delta \theta_\delta \ln(1 + v_\delta)\} \quad (5.30)\end{aligned}$$

Al derivar la ecuación anterior respecto a las proporciones de la riqueza e igualar a cero los resultados obtenidos, se encuentran los siguientes premios al riesgo que el agente económico prestamista obtiene por destinar una parte de su riqueza al proceso productivo, a la tenencia del activo sin riesgo, a la tenencia al instrumento financiero con riesgo y a la tenencia de saldos reales, respectivamente:

$$\frac{(\alpha - \delta) + (\alpha \theta_y - \delta \theta_\delta) \ln[1 + (\alpha v_y - \delta v_\delta)]}{\alpha \sigma_y - \delta \sigma_\delta}, \quad (5.31)$$

$$\frac{\delta - \mu_S - \delta \theta_\delta \ln(1 + v_\delta)}{\delta \sigma_\delta}, \quad (5.32)$$

$$\frac{\mu_R - \delta + \delta \theta_\delta \ln(1 + v_\delta)}{\sigma_R - \delta \sigma_\delta}, \quad (5.33)$$

$$\frac{\delta + \varepsilon - \lambda^2 + \delta \theta_\delta \ln(1 + v_\delta)}{\lambda - \delta \sigma_\delta}. \quad (5.34)$$

Es importante señalar, como se aprecia en las ecuaciones (5.31)-(5.34), que el premio al riesgo no es otra cosa que los rendimientos medios esperados más los saltos en dichos rendimientos, descontados por las volatilidades asociadas. Evidentemente, entre mayor sea el rendimiento

medio esperado mayor o menor sea la intensidad del salto en el mismo, mayor será el premio al riesgo. Bajo estas condiciones el equilibrio macroeconómico es:

$$a_t^l = a_0^l e^{(\psi^l)t + d\Psi^l}, \quad (5.35)$$

$$c_t^l = \rho a_0^l e^{(\psi^l)t + d\Psi^l}, \quad (5.36)$$

$$\psi^l = \frac{1}{\rho} [\delta + M^l + \Omega^l - \rho] - \frac{1}{2\rho} [\delta\sigma_\delta + \Sigma^l]^2, \quad (5.37)$$

$$d\Psi^l = [\delta\sigma_\delta + \Sigma^l] dW_t, \quad (5.38)$$

$$\omega_1^l = \frac{(\alpha - \delta) + (\alpha\theta_y - \delta\theta_\delta) \ln[1 + (\alpha v_y - \delta v_\delta)]}{\alpha\sigma_y - \delta\sigma_\delta}, \quad (5.39)$$

$$\omega_2^l = \frac{\delta - \mu_S - \delta\theta_\delta \ln(1 + v_\delta)}{\delta\sigma_\delta}, \quad (5.40)$$

$$\omega_3^l = \frac{\mu_R - \delta + \delta\theta_\delta \ln(1 + v_\delta)}{\sigma_R - \delta\sigma_\delta}, \quad (5.41)$$

$$\omega_4^l = \frac{\delta + \varepsilon - \lambda^2 + \delta\theta_\delta \ln(1 + v_\delta)}{\lambda - \delta\sigma_\delta}, \quad (5.42)$$

$$\omega_5^l = 1 - \omega_1^l - \omega_2^l - \omega_3^l - \omega_4^l, \quad (5.43)$$

en donde $M^l = \omega_1^l(\alpha - \delta) + \omega_2^l(\mu_S - \delta) + \omega_3^l(\mu_R - \delta) + \omega_4^l(\lambda^2 - \varepsilon - \delta)$ y representa la parte de la riqueza media esperada del agente prestamista que no está destinada a financiar el crédito bancario del agente prestatario. Asimismo, el riesgo asociado a la riqueza que no se destina para financiar el crédito bancario del agente prestatario está medido a través de $\Sigma^l = \omega_1^l(\alpha\sigma_y - \delta\sigma_\delta) - \omega_2^l\delta\sigma_\delta + \omega_3^l(\sigma_R - \delta\sigma_\delta) - \omega_4^l(\lambda - \delta\sigma_\delta)$. Por último, los saltos que se producen en los rendimientos medios esperados del crédito bancario así como de las condiciones de producción de las empresas correspondientes, al interior de esta economía, son $\Omega^l = (\alpha\theta_y - \delta\theta_\delta)\omega_1^l \ln[1 + (\alpha v_y - \delta v_\delta)] - (\omega_2^l + \omega_3^l + \omega_4^l - 1)\delta\theta_\delta \ln(1 + v_\delta)$.

Por su parte, la tendencia determinista de la tasa de crecimiento económico del sector prestamista está medida por ψ^l , mientras que $d\Psi^l$ mide el comportamiento estocástico de la misma. Respecto a las proporciones de equilibrio, se puede apreciar que éstas son invariantes en el tiempo y que dependen de manera directa del premio al riesgo asociado a su instrumento financiero y a la producción.

5.5.3 Equilibrio del agente prestatario

La condición *Hamilton-Jacobi-Bellman (HJB)* asociada al problema de control óptimo establecido por (5.2) y (5.20'), igualada a cero es la siguiente²:

$$\begin{aligned}
0 = & \ln c_t^b - \rho V(a_t^b) \\
& + a_t^b V'(a_t^b) \left(\omega_{1t}(\alpha + \delta) + \omega_{2t}(\mu_S + \delta) + \omega_{3t}(\mu_R + \delta) + \omega_{4t}(\lambda^2 - \varepsilon + \delta) - \delta - \frac{c_t^b}{a_t^b} \right) \\
& + \frac{1}{2} V''(a_t^b) a_t^{b^2} [\omega_{1t}(\alpha \sigma_y + \delta \sigma_\delta) + \omega_{2t} \delta \sigma_\delta + \omega_{3t}(\sigma_R + \delta \sigma_\delta) + \omega_{4t}(\delta \sigma_\delta - \lambda) - \delta \sigma_\delta]^2 \\
& + [V(a_t^l [\omega_{1t}(\alpha v_y + \delta v_\delta) + \omega_{2t} \delta v_\delta + \omega_{3t} \delta v_\delta + \omega_{4t} \delta v_\delta - \delta v_\delta]) - V(a_t^l)] dQ_t, \quad (5.44)
\end{aligned}$$

con sus condiciones de optimalidad:

$$\left. \begin{aligned}
\frac{\partial HJB}{\partial c_t^b} = \frac{1}{c_t^b} - V'(a_t^b) = 0 & \quad V''(a_t^l) = -\frac{\beta_1^b}{a_t^{b^2}} \\
V(a_t^b) = \beta_0^b + \beta_1^b \ln a_t^b & \quad c_t^b = \frac{a_t^b}{\beta_1^b} \\
V'(a_t^b) = \frac{\beta_1^b}{a_t^b} &
\end{aligned} \right\} \quad (5.45)$$

Se sustituyen las condiciones de optimalidad anteriores y $\beta_1^b = 1/\rho$ en la ecuación (5.44) para encontrar:

$$\begin{aligned}
0 = & \ln \rho - \rho \beta_0^b \\
& + \frac{1}{\rho} [\omega_{1t}(\alpha + \delta) + \omega_{2t}(\mu_S + \delta) + \omega_{3t}(\mu_R + \delta) + \omega_{4t}(\lambda^2 - \varepsilon + \delta) - \delta - \rho] \\
& - \frac{1}{2\rho} [\omega_{1t}(\alpha \sigma_y + \delta \sigma_\delta) + \omega_{2t} \delta \sigma_\delta + \omega_{3t}(\sigma_R + \delta \sigma_\delta) + \omega_{4t}(\delta \sigma_\delta - \lambda) - \delta \sigma_\delta]^2 \\
& + \frac{1}{\rho} \{ (\alpha \theta_y + \delta \theta_\delta) \omega_{1t} \ln[1 + (\alpha v_y + \delta v_\delta)] + (\omega_{2t} + \omega_{3t} + \omega_{4t} - 1) \delta \theta_\delta \ln(1 + v_\delta) \} \quad (5.46)
\end{aligned}$$

Análogamente, se deriva la ecuación (5.46) respecto a las proporciones de la riqueza y se igualan a cero los resultados obtenidos, con ello se obtienen los siguientes premios al riesgo que el agente económico prestatario obtiene por destinar una parte de su riqueza al proceso productivo, a la tenencia del activo sin riesgo, a la tenencia al instrumento financiero con riesgo y a la tenencia de saldos reales, respectivamente:

² Véase el apéndice matemático.

$$\frac{(\alpha + \delta) + (\alpha\theta_y + \delta\theta_\delta) \ln[1 + (\alpha\nu_y + \delta\nu_\delta)]}{\alpha\sigma_y + \delta\sigma_\delta}, \quad (5.47)$$

$$\frac{\delta\theta_\delta \ln(1 + \nu_\delta) - \mu_S - \delta}{\delta\sigma_\delta}, \quad (5.48)$$

$$\frac{\mu_R + \delta + \delta\theta_\delta \ln(1 + \nu_\delta)}{\sigma_R + \delta\sigma_\delta}, \quad (5.49)$$

$$\frac{\varepsilon + \delta\theta_\delta \ln(1 + \nu_\delta) - \delta - \lambda^2}{\delta\sigma_\delta - \lambda}. \quad (5.50)$$

Una vez más el premio al riesgo para cada instrumento financiero y proceso productivo están determinados por los rendimientos medios esperados más los saltos en los mismos descontados por sus volatilidades asociadas, aunque con una dirección contraria a las correspondientes al agente prestamista. Bajo estas condiciones el equilibrio macroeconómico es:

$$a_t^b = a_0^b e^{(\psi^b)t + d\Psi^b}, \quad (5.51)$$

$$c_t^b = \rho a_0^b e^{(\psi^b)t + d\Psi^b}, \quad (5.52)$$

$$\psi^b = \frac{1}{\rho} [M^b + \Omega^b - \delta - \rho] - \frac{1}{2\rho} [\Sigma^b - \delta\sigma_\delta]^2, \quad (5.53)$$

$$d\Psi^b = [\Sigma^b - \delta\sigma_\delta] dW_t, \quad (5.54)$$

$$\omega_1^b = \frac{(\alpha + \delta) + (\alpha\theta_y + \delta\theta_\delta) \ln[1 + (\alpha\nu_y + \delta\nu_\delta)]}{\alpha\sigma_y + \delta\sigma_\delta}, \quad (5.55)$$

$$\omega_2^b = \frac{\delta\theta_\delta \ln(1 + \nu_\delta) - \mu_S - \delta}{\delta\sigma_\delta}, \quad (5.56)$$

$$\omega_3^b = \frac{\mu_R + \delta + \delta\theta_\delta \ln(1 + \nu_\delta)}{\sigma_R + \delta\sigma_\delta}, \quad (5.57)$$

$$\omega_4^b = \frac{\varepsilon + \delta\theta_\delta \ln(1 + \nu_\delta) - \delta - \lambda^2}{\delta\sigma_\delta - \lambda}, \quad (5.58)$$

$$\omega_5^b = 1 - \omega_1^b - \omega_2^b - \omega_3^b - \omega_4^b. \quad (5.59)$$

Ahora en este equilibrio macroeconómico, la parte de la riqueza media esperada del agente prestatario que no está destinada a obtener un crédito bancario está determinada por $M^b = \omega_1^b(\alpha + \delta) + \omega_2^b(\mu_S + \delta) + \omega_3^b(\mu_R + \delta) + \omega_4^b(\lambda^2 - \varepsilon + \delta)$, por su parte, el riesgo asociado a esta proporción de la riqueza es $\Sigma^b = \omega_1^b(\alpha\sigma_y + \delta\sigma_\delta) + \omega_2^b\delta\sigma_\delta + \omega_3^b(\sigma_R + \delta\sigma_\delta) + \omega_4^b(\delta\sigma_\delta - \lambda)$, mientras que los saltos en la riqueza ocasionados por el crédito bancario y el

proceso productivo para el agente prestatario es $\Omega^b = (\alpha\theta_y + \delta\theta_\delta)\omega_1^b \ln[1 + (\alpha\nu_y + \delta\nu_\delta)] + (\omega_2^b + \omega_3^b + \omega_4^b - 1)\delta\theta_\delta \ln(1 + \nu_\delta)$. De igual manera, la tendencia determinista de la tasa de crecimiento económico del sector prestatario está medida por ψ^b , mientras que $d\Psi^b$ mide el comportamiento estocástico de la misma. Respecto a las proporciones de equilibrio, se puede apreciar que éstas son invariantes en el tiempo y que dependen de manera directa del premio al riesgo asociado a su instrumento financiero y a la producción.

5.6 Crecimiento y sistema financiero

La tasa de crecimiento de la economía depende del promedio del crecimiento de los sectores productivos prestamista y prestatario, de tal forma que al considerar las ecuaciones (5.19), (5.20), (5.37) y (5.53) el componente determinista del crecimiento de esta economía es:

$$\psi = \frac{\psi^l + \psi^b}{2}. \quad (5.60)$$

Equivalentemente,

$$\psi = \frac{(\delta + M^l + \Omega^l - \rho) + (M^b + \Omega^b - \delta - \rho) - \frac{1}{2}[(\delta\sigma_\delta + \Sigma^l)^2 + (\Sigma^b - \delta\sigma_\delta)^2]}{2\rho}, \quad (5.60')$$

Debido a que ambos agentes dividen su riqueza en el mismo número de partes, entonces la ecuación anterior se puede simplificar de la siguiente manera:

$$\psi = \alpha + \mu_S + \mu_R + \lambda^2 + \alpha\theta_y\nu_y + \delta\theta_\delta\nu_\delta - \varepsilon - \rho. \quad (5.61)$$

La economía crecerá si $\alpha + \mu_S + \mu_R + \lambda^2 + \alpha\theta_y\nu_y + \delta\theta_\delta\nu_\delta > \varepsilon + \rho$, es decir, si hay un aumento en la productividad media esperada, o en el rendimiento medio esperado del activo sin riesgo, o en el rendimiento medio esperado del activo con riesgo, o en la volatilidad en los precios, o en la intensidad de los saltos asociados al crédito bancario o al proceso productivo, o en todos en conjunto, la tasa de crecimiento incrementará siempre y cuando se mantenga constante el tipo de cambio y las preferencias de los consumidores, en caso contrario la economía decrecerá. Por su parte, la variabilidad en el componente determinista de la tasa de crecimiento es:

$$E \left[\frac{da_t}{a_t} \right]^2 = \alpha^2 (\sigma_y^2 + v_y^2 \theta_y) (\sigma_R^2 + \lambda^2) dt. \quad (5.62)$$

Lo anterior significa que la varianza en la tasa de crecimiento depende de la varianza en la productividad media, de los saltos en la productividad media, de la varianza del rendimiento del activo con riesgo y de la varianza en el tipo de cambio. De acuerdo con las ecuaciones (5.37) y (5.53), dicha variabilidad tiene un efecto negativo sobre el crecimiento económico de cada sector de la economía y, por ende, en la economía en conjunto. En consecuencia el componente estocástico de la economía es:

$$\Psi_W = (\sigma_y + v_y \theta_y) (\sigma_R + \lambda) \xi. \quad (5.63)$$

La ecuación anterior muestra que cada incremento temporal en la economía afectará directamente a la desviación estándar estocástica de la tasa de crecimiento. Finalmente consideremos el escenario internacional del sistema financiero, por lo que se sustituye la ecuación (5.11) en (5.61), de tal forma que ahora:

$$\psi = 2i^* + \alpha + \varepsilon + \lambda^2 + \alpha \theta_y v_y + \delta \theta_\delta v_\delta - \rho. \quad (5.60)$$

La ecuación anterior muestra que la tasa de crecimiento de económico aumentará si se presenta un incremento de la tasa de interés internacional, la cual está directamente relacionada con los rendimientos medios esperados de los instrumentos financieros a nivel internacional, lo que implica que a mayores recursos que el agente obtenga del sistema financiero mayor será el crecimiento. Asimismo, incrementos en la intensidad de los saltos o en el número de saltos genera un mayor ritmo de crecimientos en las actividades económicas. Desafortunadamente, en la actualidad el sistema financiero a nivel global se caracteriza justo por hacer todo lo contrario, es decir, por generar la mayor cantidad de ganancias posibles de operaciones financieras que no tiene que ver directamente con la actividad económica.

5.6 Conclusiones

Al utilizar un modelo de crecimiento endógeno estocástico, de economía abierta, con rendimientos constantes a escala y rendimientos marginales de crecientes se mostró que las volatilidades asociadas a la tenencia de saldos reales, a la tenencia de un activo financiero

riesgoso y a los ingresos obtenidos del proceso productivo con una tecnología estocástica, tienen efectos negativos sobre la tasa de crecimiento. Más todavía, cuando las dispersiones antes mencionadas son menores a los rendimientos promedios, se verifica una caída en la acumulación de la riqueza por parte de los agentes prestamistas y prestatarios. En contra parte, los rendimientos medios esperados de todos los instrumentos financieros así como la productividad media esperada al interior de las actividades económicas, generan un efecto positivo sobre el crecimiento. Lo mismo ocurre cuando se incrementa la intensidad en el salto o el número de saltos en el proceso productivo y en el crédito bancario, por separado o de manera conjunta.

Por su parte, al considerar el entorno del sistema financiero internacional se encontró que un incremento en los rendimientos medios esperados de los activos financieros a nivel nacional, pero también a nivel internacional, incrementa la tasa de crecimiento. En consecuencia, es deseable que un sistema financiero al interior de una economía se comprometa a proporcionar instrumentos financieros que generen efectos positivos sobre la actividad económica y no al revés.

Es importante resaltar que los resultados aquí encontrados dependen fuertemente de los supuestos explicitados en la sección, respecto del comportamiento del sector financiero y del sector real. Asimismo, las principales limitaciones de la investigación radican principalmente en el papel que el sistema financiero desempeña, por lo que modificaciones en su comportamiento pueden ampliar el análisis. Por otro lado, es necesario un análisis en donde se incorpore al gobierno para medir el impacto que éste tenga sobre el crecimiento y posiblemente sobre el sistema financiero.

5.7 Bibliografía

Aghion, P. y P. Bolton, (1997), “A Trickle Down Theory of Growth and Development”, *Review of Economic Studies*, 64, 151-172.

Bagehot, W. (1968), *Lombard Street: el Mercado monetario de Londres*. Fondo de Cultura Económica. México.

- Banerjee, A. y A. Newman, (1993), “Occupational Choice and the Process of Development”, *Journal of Political Economy*, 101, 274-298.
- Benabou, R. (1993), “Workings of a City: Location, Education and Production”, *Quarterly Journal of Economics*, 108, 619-652.
- Benabou, R. (1996a), “Equity and Efficiency in Human Capital Investment: The Local Connection”, *Review of Economic Studies*, 62(2), 237-264.
- Benabou, R. (1996b), “Inequality and Growth”, *NBER Working Paper 5658*, 1-56.
- Durlauf, S. (1996a), “A Theory of Persistent Income Inequality”, *Journal of Economic Growth*, Vol. 1, No. 1, 75-94.
- Durlauf, S. (1996b), “Neighborhood Feedbacks, Endogenous Stratification, and Income Inequality”, forthcoming in “*Proceedings of the Sixth International Symposium on Economic Theory and Econometrics*”, W. Barnett, G. Gandolfo y C. Hillinger, eds., Cambridge University Press.
- Fernández, R. y R. Rogerson, (1994), “Public Education and the Dynamics of Income Distribution: A Quantitative Evaluation of Education Finance Reform”, *NBER Working Paper 4883*.
- Fernández, R. y R. Rogerson, (1996), “Income Distribution, Communities and the Quality of Public Education”, *Quarterly Journal of Economics*, 111, 135-164.
- Galor, O. y J. Zeira, (1993), “Income Distribution and Macroeconomics”, *Review of Economic Studies*, 60, 35-52.
- Gurley, J. G., y E. S. Shaw, (1995), “Financial Aspects of Economic Development”, *American Economic Review*, 45, 515-538.

Kremer M. y E. Maskin (1994), “Segregation of Workers by Skill and the Rise of Inequality”, *Working Paper Massachusetts Institute of Technology*.

Mckinnon, R. I. (1973), *Money and Capital in Economic Development*, Washington, DC: Brookings Institution.

Piketty, T. (1997), “The Dynamics of Wealth Distribution and the Interest Rate with Credit Rationing”, *Review of Economic Studies*, 64, 173-189.

Rivas-Aceves S. y J. F. Martínez Pérez, (2009), “El Impacto de la Regulación Gubernamental en el Mercado de Capitales sobre el Crecimiento”, capítulo en libro *Economía teórica y aplicada, México y el mundo*, Editorial UAM-Eón, México 2009, pp.59-68. ISBN UAM: 978-607-477-176-3, ISMB Eón: 978-607-7519-53-9.

Rivas-Aceves S. y J. F. Martínez Pérez, (2011a), “La Crisis Financiera Global, el mercado de capitales y el Crecimiento Económico: Algunos aspectos para la economía de México”, capítulo en libro *La integración económica ante la primera crisis financiera del siglo XX*, Editorial Eón, México, 2011, pp. 73-98.

Rivas-Aceves S. y J. F. Martínez Pérez, (2011b), “La Crisis Financiera Internacional y el Crecimiento Económico: La Banca de Desarrollo, una Posible Solución”, capítulo en libro *La integración económica: balance de un proceso inconcluso*, Editorial Eón, México, 2011, pp. 199-224.

Apéndice matemático

Derivación de la ecuación (5.22)

El problema planteado por (5.2) y (5.19), arroja que:

$$J(a_t^l, t) = \text{Max}_{c_s|_{[t, t+dt]}} E\{\ln(c_t^l)e^{-\rho t} + o(dt) + J(a_t^l, t) + dJ(a_t^l, t)|\mathcal{F}_t\}. \quad (\text{A.1})$$

Al aplicar el Lema de Íto se tiene:

$$\begin{aligned} dJ(a_t^l, t) = & \left[\frac{\partial J}{\partial t} \right. \\ & + \frac{\partial J}{\partial a_t^l} a_t^l \left(\delta + \omega_{1t}(\alpha - \delta) + \omega_{2t}(\mu_S - \delta) + \omega_{3t}(\mu_R - \delta) + \omega_{4t}(\lambda^2 - \varepsilon - \delta) \right. \\ & \left. \left. - \frac{c_t^l}{a_t^l} \right) \right] dt \\ & + \left[\frac{1}{2} \frac{\partial^2 J}{\partial a_t^l{}^2} a_t^l{}^2 [\delta\sigma_\delta + \omega_{1t}(\alpha\sigma_y - \delta\sigma_\delta) - \omega_{2t}\delta\sigma_\delta + \omega_{3t}(\sigma_R - \delta\sigma_\delta) - \omega_{4t}(\lambda + \delta\sigma_\delta)]^2 \right] dt \\ & + \frac{\partial J}{\partial a_t^l} a_t^l [\delta\sigma_\delta + \omega_{1t}(\alpha\sigma_y - \delta\sigma_\delta) - \omega_{2t}\delta\sigma_\delta + \omega_{3t}(\sigma_R - \delta\sigma_\delta) - \omega_{4t}(\lambda + \delta\sigma_\delta)] dw \\ & + \frac{\partial J}{\partial a_t^l} a_t^l [\delta v_\delta + \omega_{1t}(\alpha v_y - \delta v_\delta) - \omega_{2t}\delta v_\delta - \omega_{3t}\delta v_\delta - \omega_{4t}\delta v_\delta] dQ \end{aligned} \quad (\text{A.2})$$

Si se establece que $J(a_t^l, t) = V(a_t^l)e^{-\rho t}$, entonces:

$$\begin{aligned} 0 = & \ln c_t^l - \rho V(a_t^l) \\ & + a_t^l V'(a_t^l) \left(\delta + \omega_{1t}(\alpha - \delta) + \omega_{2t}(\mu_S - \delta) + \omega_{3t}(\mu_R - \delta) + \omega_{4t}(\lambda^2 - \varepsilon - \delta) - \frac{c_t^l}{a_t^l} \right) \\ & + \frac{1}{2} V''(a_t^l) a_t^l{}^2 [\delta\sigma_\delta + \omega_{1t}(\alpha\sigma_y - \delta\sigma_\delta) - \omega_{2t}\delta\sigma_\delta + \omega_{3t}(\sigma_R - \delta\sigma_\delta) - \omega_{4t}(\lambda + \delta\sigma_\delta)]^2 \\ & + [V(a_t^l [\delta v_\delta + \omega_{1t}(\alpha v_y - \delta v_\delta) - \omega_{2t}\delta v_\delta - \omega_{3t}\delta v_\delta - \omega_{4t}\delta v_\delta]) - V(a_t^l)] dQ_t. \end{aligned} \quad (\text{A.3})$$

Para encontrar (A.3) se necesita sustituir $J(a_t^l, t) = V(a_t^l)e^{-\rho t}$ aplicar (A.2), tomar esperanzas y fijar $E(dw) = 0$, dividir entre dt y tomar el límite cuando $t \rightarrow 0$ a (A.1). El mismo método se aplica para encontrar la ecuación (5.41), sólo que con el problema de optimización dado por (5.2) y (5.20).

Capítulo 6

Determinación de la probabilidad de incumplimiento para empresas que cotizan en la BMV utilizando el modelo de Merton con activos con características de persistencia

Guillermo Sierra Juárez

*Universidad de Guadalajara, CUCEA
Departamento de Métodos Cuantitativos*

Katia García Loredó

CUCEA Posgrado de Economía

Resumen

En este trabajo se analiza la probabilidad de incumplimiento para empresas que cotizan en la BMV con un prueba de retrospectiva para el período 2007-2010 y un pronóstico para el año 2013; la probabilidad de incumplimiento se calcula mediante el modelo de Merton tradicional y con una modificación del modelo considerando características de persistencia mediante brownianos fractales. Concretamente se busca realizar una comparación entre el modelo tradicional y el modelo fraccional. Los resultados indican una subestimación de la probabilidad de incumplimiento por parte del modelo tradicional en todas las empresas evaluadas al obtener una probabilidad de incumplimiento mayor con el modelo fractal.

6.1 Introducción

Algunas de las más importantes crisis bancarias, han tenido su origen en la falta de controles sobre la concentración en una u otra forma del riesgo crediticio y de su incumplimiento. El desarrollo de la supervisión bancaria y de los mercados de transferencia de riesgo crediticio han obligado que la industria financiera desarrollo nuevas herramientas para medir y manejar el riesgo.

Según el Banco de México (Banxico, 2009) en los últimos años se ha observado un aumento importante en el valor en riesgo (VaR¹) de crédito de la banca múltiple. En el 2007 el incremento en el VaR obedeció al aumento en el saldo de la cartera de crédito, el incremento verificado en 2008 ha sido producto de un aumento en los factores de riesgo que conforman al VaR. Los factores que conforman el VaR son las probabilidades de incumplimiento, la correlación entre los incumplimientos de los deudores y la concentración de la cartera.

La crisis financiera del 2008 y la situación de aumento en la probabilidad de incumplimiento antes mencionada, ha generado una serie de cuestionamientos sobre las políticas de administración de riesgos y de gobierno corporativo de algunas empresas a nivel mundial, al igual que de las empresas mexicanas. Una muestra de la relevancia que debe tener una correcta medición del riesgo es el comité de supervisión bancaria de Basilea, formado en 1974 por los gobernadores centrales del G-10, cuyo principal interés es establecer lineamientos para una adecuada medición del riesgo de crédito.

El riesgo de crédito se define por De Lara (2003) como la pérdida potencial que se registra con motivo del incumplimiento de una contraparte en una transacción financiera.

Las metodologías actuales para calcular el riesgo de crédito dependen de sobremanera de métodos numéricos; entre las diversas herramientas y modelos de estimación para medir la probabilidad de incumplimiento de las empresas y para poder medir su riesgo crediticio, existe el modelo de Merton.

Este trabajo desarrolla la aplicación empírica del modelo de Merton utilizando un grupo de empresas que cotizan en la Bolsa Mexicana de Valores (BMV), además de incorporar el movimiento browniano fraccional, que es una generalización al modelo tradicional.

¹ VaR por su expresión en inglés, Value at Risk (Valor en Riesgo), es una medida de riesgo utilizada en el riesgo de mercado de activos financieros. Se define como la máxima pérdida esperada a un nivel de confianza y en un horizonte de tiempo.

El modelo de Merton, desarrollado en 1974, supone que existe información completa, esta definido en un plazo fijo y la probabilidad de incumplimiento solo se presenta cuando los activos están por debajo de la deuda. Es un modelo ya desarrollado que es base para la medición de otros modelos de riesgo de incumplimiento, por lo que se busca proponer una generalización al introducir los brownianos fraccionales, con los cuales se incorpora memoria o persistencia a la serie, con el fin de mejorar las estimaciones.

El análisis empírico desarrollado en este trabajo esta conformado por 10 empresas mexicanas que cotizan en la BMV y emiten bonos corporativos. El análisis se divide en dos períodos: el primero es una prueba de retrospectiva en el cual se abarca del 2007 al 2010, y el segundo es un pronóstico del año 2013. Se utilizan datos corrientes obtenidos de los estados financieros de cada empresa y de ciertas empresas proveedoras de precios. Además se incluye un análisis de tres métodos de volatilidad: la volatilidad histórica, el proceso ARMA y el proceso GARCH. Se espera que al realizar la modificación al movimiento browniano fractal se consiga una mejor aproximación de la probabilidad de incumplimiento de las empresas con el modelo de Merton.

El trabajo se divide en cinco secciones, siendo este la primera. La sección 6.2 presenta el marco teórico que proporciona una visión general del riesgo crediticio, sus modelos de obtención y sus clasificaciones, además se mencionan investigaciones que emplean el modelo de Merton y sus resultados.

En la sección 6.3 se encuentra la metodología del modelo de Merton tradicional, la ecuación de Black-Scholes fractal, el modelo de Merton Fraccional, el proceso de Hurst y la determinación de la muestra.

La sección 6.4 incorpora los resultados de la probabilidad de incumplimiento; por último en la sección 6.5 se presentan las conclusiones.

6.2 Marco teórico

Según Basilea (1988) el riesgo de crédito se define como el riesgo de que la contraparte incumpla en la obligación; también se concibe como un deterioro en la calidad crediticia de la contraparte o en la garantía o colateral pactado originalmente.

La International Actuarial Association (2004) define el riesgo de crédito como el riesgo al que una compañía está expuesta por el incumplimiento o impago de una obligación de crédito, ó por el cambio en la calidad de las garantías del crédito de la contraparte o de los intermediarios que inicialmente habían sido acordadas.

El riesgo de crédito se refiere a la pérdida potencial derivada de que la contraparte no pueda cumplir con sus obligaciones financieras en las condiciones definidas contractualmente, Sánchez (2001). La versión de mercado indica que el riesgo de crédito se define como la pérdida potencial que podría sufrir un tenedor de un portafolio de préstamos, instrumentos financieros o derivados, como consecuencia de que el valor, al final del horizonte de inversión, es inferior al valor actual del portafolio, ambos valuados a precios de mercado. Es decir, hay un riesgo de crédito aunque la contraparte no sufra quebranto alguno. Se han desarrollado diferentes modelos para estimar el riesgo de crédito, algunos orientados a estimar el riesgo en su versión de mercado.

6.2.1 Modelos de Riesgo Crédito

Los modelos de incumplimiento se diferencian de acuerdo con las técnicas de medición del riesgo de crédito, realizando una clasificación entre modelos tradicionales, estructurales y de forma reducida (Jackson, Nickell y Perraudim, 1999; Márquez, 2006) esto con base a los supuestos que cada tipo de modelo tiene.

Los modelos tradicionales de administración de riesgo se basan en conceptos de tipo fundamental, su característica es que utilizan una ponderación de factores para determinar la probabilidad de incumplimiento de las obligaciones. Las técnicas de tipo fundamental parten de variables económicas y financieras, así como del desempeño de la empresa. Estos modelos involucran el criterio subjetivo de cada analista, que por lo general se base en valoraciones de acuerdo a su experiencia. Los modelos tradicionales comprenden: el modelo de las 5 Cs, el modelo de Altman (1968) y sus modificaciones, y el modelo estándar de Basilea (1988, 2001 y 2004). Este tipo de modelos se ha quedado atrás por las condiciones cambiantes del ámbito financiero y se ha sustituido por técnicas probabilísticas y estadísticas más sofisticadas.

Los modelos estructurales son de amplio reconocimiento en los mercados financieros, utilizados por grandes firmas y de amplia divulgación en la literatura financiera, su característica es que suponen que los inversionistas tienen toda la información del mercado y tienen un conocimiento completo del valor de los activos y la deudas de todas las firmas (Jarrow y Protter,

2004). Los modelos estructurales incluyen el modelo de Merton (1974), Geske (1977), el modelo CreditMetrics de JP Morgan (1977) y el modelo Credit Portafolio Manager de KMV Moody's.

Por otro lado, los modelos de forma reducida asumen que los inversionistas no disponen de toda la información sobre el manejo de la firma, y dicha falta de conocimiento no permite predecir la probabilidad de incumplimiento de las firmas. Entre los modelos de forma reducida más importantes se encuentra el modelo Credit Portafolio View de McKinsey, Márquez (2006). En esta categoría también están otros trabajos realizados por distintos autores que se basan en modelos logit, modelos probit, modelos lineales de probabilidad y simulación.

El modelo de Merton (1974) calcula la probabilidad de incumplimiento, la cual se presenta cuando el valor de los activos es inferior al monto total de la deuda financiera de la empresa. El valor de los activos se considera como una opción de compra con precio de ejercicio igual a el monto de la deuda (conocido como call europeo²). Además el modelo considera que el valor de los activos de la firma sigue una opción que demanda información específica sobre la estructura de vencimiento de la deuda.

El modelo de Merton es el modelo base para los cálculos de las probabilidades de incumplimiento en el modelo de KMV Moddy's y del modelo de Geske. Sin embargo, en el modelo de Merton el incumplimiento se presenta cuando el valor de los activos es inferior al monto total de la deuda financiera, mientras que en el modelo de KMV Moddy's el incumplimiento se presenta cuando el valor de los activos está por debajo de un umbral definido entre los pasivos totales y la deuda de corto plazo. Esta diferencia en el establecimiento del umbral afecta directamente la probabilidad de incumplimiento, pues manteniendo los demás factores constantes, mientras más alejado se ubique el umbral del monto total de la deuda menor será la probabilidad de incumplimiento.

En el modelo de Geske se considera que el valor de los activos de la firma es una opción compuesta determinada por los montos y vencimientos de las deudas de corto y largo plazo de la compañía, mientras que en el modelo de Merton el valor de los activos se considera como una opción de compra con precio de ejercicio el monto de la deuda.

La principal crítica que ha recibido el modelo de Merton es que los spread³, que se estiman a través del modelo son muy bajos en comparación con los que se dan en la práctica.

² Un call europeo es una opción de compra que solo se puede ejercer en la fecha de vencimiento. El put europeo es la opción de venta que de igual forma solo se puede ejercer en la fecha de vencimiento.

³ Premio al riesgo por la deuda.

Chen y Panjer (2003) deducen que el spread estimado a través del modelo de Merton es equivalente al spread obtenido en los modelos de forma reducida, por esta razón podría usarse la información de los spread del mercado para llegar a estimaciones de la probabilidad de incumplimiento ajustada a la información del mercado. Los spread de mercado considerados por Chen y Panjer (2003) son dados por las agencias calificadoras como Moody's que se basan en las calificaciones obtenidas de las matrices de transición y estas se establecen con las probabilidades de incumplimiento calculadas a través del modelo de Merton.

El trabajo de Cerezo, Bielsa y Ramón (2011) tiene como objetivo medir el riesgo de crédito de empresas que cotizan en la bolsa de valores a través del modelo estructural de Merton; se analizan empresas que cotizaron sus acciones en el mercado colombiano durante el período 2005-2007. Además, en este trabajo se analiza si hay diferencias en los indicadores de riesgo de crédito cuando se agrupa a las empresas por sectores económicos. Como conclusiones expresan que durante el período de estudio se puede apreciar una relación positiva entre la tasa libre de riesgo y la tasa de crecimiento del PIB, en la medida en que se presentó una mayor capitalización bursátil también se incrementó el endeudamiento de las empresas consideradas en el estudio y se aprecia una correlación negativa entre el valor de los activos de las firmas y su volatilidad.

Para el caso de México, está el trabajo de Moreno (2007), en este se utiliza el modelo basado en el método de valuación de opciones de Black-Scholes-Merton que permite valorar las acciones de una empresa endeudada como opciones de compra en un horizonte de tiempo, así es posible el cálculo de la distancia al incumplimiento. Se aplica al grupo financiero Bancomer para el período de enero de 1998 a junio de 2000 y al grupo financiero BBVA-BANCOMER de julio de 2000 al diciembre de 2003. Utiliza dos modelos, el de valuación de opciones de Black y Scholes (1973) y el de Merton. Como resultados obtiene que el método desarrollado hace evidente que en el período de 1998-1999, la institución transitaba estadísticamente por un período de quiebra técnica (pasivos mayores a activos), y que tal situación hizo necesario aceptar la oferta de Banco Bilbao Vizcaya Argentaria, fusionando las instituciones bancarias. Señala que la adecuada alineación de los métodos descritos, específicamente del basado en la aplicación del modelo Merton para el cálculo de la distancia de incumplimiento, hace posible tomar posturas de precaución o cautela y anticiparse a situaciones de incumplimiento que tienen un gran costo para la sociedad cuando tal pone en riesgo la estabilidad del sistema financiero.

6.3 Metodología

El modelo Black - Scholes valúa opciones de compra y de venta europea sobre acciones que no pagan dividendos. Merton realizó una extensión de este modelo de valuación de opciones hacia la valuación de la capacidad de pago de las empresas. En dicha extensión supone que la deuda corporativa se toma como una opción de compra sobre el capital de la empresa, es decir, el precio de una opción equivale al valor de la empresa; al momento del vencimiento pueden ocurrir dos cosas, si la deuda es menor al valor de la empresa los acreedores no ejercen su opción porque la deuda se puede liquidar, pero si la deuda llega a ser mayor que el valor de la empresa entonces los acreedores pueden ejercer su opción y tendrían derecho a quedarse con la empresa.

6.3.1 Modelo de Merton para el cálculo de probabilidades de incumplimiento

El modelo de Merton se clasifica como estructural, supone que los inversionistas tienen información perfecta del mercado y un conocimiento completo del valor de los activos y de la deuda de todas las firmas. El incumplimiento se presenta cuando el valor de los activos es inferior al monto total de la deuda financiera. El valor de los activos se considera como una opción call europea que tiene como precio de ejercicio el monto de la deuda.

Se utiliza la fórmula de valuación de Black y Scholes para calcular la probabilidad de incumplimiento. El modelo de Merton⁴ supone que una empresa emite bonos cupón cero que vencen en una fecha establecida llamada T . Las variables del modelo son:

t : Fecha de referencia, hoy.

T : Fecha de vencimiento.

S_t : Valor de mercado de las acciones de la empresa, hoy.

S_T : Valor de mercado de las acciones de la empresa en T .

$B_c(t, T)$: Valor de mercado de los bonos (con vencimiento en T) de la empresa, hoy.

V_t : Valor de mercado de los títulos, de capital y deuda, de la empresa, hoy.

V_T : Valor de mercado de los títulos, de capital y deuda, de la empresa en T .

D : Principal e interés que la empresa debe pagar en T .

σ_s : Volatilidad de las acciones de la empresa.

σ_v : Volatilidad del valor de la empresa en el mercado.

⁴ Desarrollo del modelo tomado de Venegas (2008).

V_t esta compuesto por el valor de mercado de las acciones (S_t) y el valor de mercado de los bonos (B_c) de la empresa:

$$V_t = S_t + B_c(t, T) \quad (6.1)$$

Se supone además que

$$dV_t = \mu_V V_t dt + \sigma_V V_t dW_t \quad (6.2)$$

$$dS_t = \mu_S S_t dt + \sigma_S S_t dW_t \quad (6.3)$$

Dado que el proceso $(W_t)_{t \in [0, T]}$ es un movimiento Browniano definido sobre un espacio fijo de probabilidad con su filtración aumentada, existe una correlación perfecta entre las ecuaciones (6.2) y (6.3); estas ecuaciones hablan sobre los cambios en el valor de la empresa y en el valor de mercado de las acciones de la empresa. La limitación de la ecuación (6.2) es que los parámetros $\mu_V V_t$, $\sigma_V V_t$ y V_t no son observables.

Si el valor de la empresa en el tiempo T es menor que la deuda, $V_T < D$, entonces la empresa incumple, por lo menos parcialmente, con el pago de su deuda, y en este caso $S_T = 0$. Si por el contrario $V_T \geq D$, la empresa paga su deuda en T y $S_T = V_T - D$. Esto se resume en:

$$S_T = \max(V_T - D, 0) \quad (6.4)$$

Esto muestra que el valor de mercado de las acciones puede verse como una opción europea de compra sobre el valor de mercado de la empresa con precio igual al pago de su deuda. La formulación de Black-Scholes proporciona el valor inicial de las acciones de la empresa en t , en un mundo neutral al riesgo, el cual está dado por:

$$S_t = V_t \Phi(d_1) - D e^{-r(T-t)} \Phi(d_2) \quad (6.5)$$

Donde

$$d_1 = \frac{\ln\left(\frac{V_t}{D}\right) + \left(r + \frac{1}{2}\sigma_v^2\right)(T-t)}{\sigma_v \sqrt{T-t}} \quad (6.6)$$

$$d_2 = d_1 - \sigma_v \sqrt{T-t} \quad (6.6)$$

El valor de la deuda hoy es $B_c(t, T) = V_t - S_t$.

$$\begin{aligned} S_t &= e^{-r(T-t)} \int_{-\infty}^{\infty} \max(v - D, 0) f_{V_T|V_t}(v|V_t) dv \\ &= e^{-r(T-t)} E[V_T 1_{\{V_T > D\}} | V_t] - D e^{-r(T-t)} \mathbb{P}\{V_T > D | V_t\} \end{aligned}$$

Donde $f_{V_T|V_t}(v|V_t)$ es la función de densidad de V_T , condicional al valor inicial V_t . La probabilidad neutral al riesgo, de que la empresa cumpla con el pago de su deuda es:

$$\mathbb{P}\{V_T > D|V_t\} = \Phi(d_2) \quad (6.7)$$

Siendo Φ la función de distribución normal estándar. Por lo tanto la probabilidad al riesgo, de que la empresa incumpla con el pago de su deuda es

$$\mathbb{P}\{V_T < D|V_t\} = 1 - \Phi(d_2) = \Phi(-d_2) = \int_{-\infty}^{d_2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad (6.8)$$

Para calcular esta probabilidad se debe conocer a σ_V como V_t . Sin embargo, ninguna de las dos es directamente observable. No obstante, si la empresa emite acciones que son colocadas públicamente, entonces S_t puede estimarse como el número de acciones por su precio de mercado. Por el lema de Itô:

$$\mu_S S_t dt + \sigma_S S_t dW_t = \left(\frac{dS_t}{dt} + \mu_V V_t \frac{dS_t}{dV_t} + \frac{1}{2} \sigma_V^2 V_t^2 \frac{d^2 S_t}{dV_t^2} \right) dt + \sigma_V V_t \frac{dS_t}{dV_t} dW_t$$

Dado que los componentes estocásticos deben ser iguales, se sigue que:

$$\sigma_S S_t = \sigma_V V_t \frac{dS_t}{dV_t}$$

$$\sigma_S S_t = \sigma_V V_t \Phi(d_1)$$

Si se cuenta con un estimador de la desviación estándar de los rendimientos, σ_S , entonces la ecuación anterior es otra ecuación que σ_V y V_t deben satisfacer.

$$\mu_S S_t = D e^{-r(T-t)} \left(-r \Phi(d_2) + \frac{\sigma_V \Phi'(d_2)}{2\sqrt{T-t}} \right) + \mu_V \sigma_V \Phi(d_1) + \frac{\sigma_V V_t \Phi'(d_1)}{2\sqrt{T-t}}$$

Ya que σ_V y V_t se han determinado, se sustituyen los valores y se obtiene el valor de μ_V . Se debe realizar un pequeño ajuste en las variables del modelo dado que para el valor de la empresa, V_T , existen diferentes definiciones pero no una fija. Se toman las cuentas financieras y se asigna al valor de la empresa los activos totales. Entonces para la ecuación (6.1) se tiene:

$$A_t = C_t + P_t$$

Donde

A_t : Activo total de la empresa en tiempo t .

C_t : Capital de la empresa en tiempo t .

P_t : Pasivo de la empresa en tiempo t , y por pasivo entenderemos el valor de los bonos en el tiempo t .

$$C_T = \max(A_T - P_T, 0) \quad (6.10)$$

La formulación de Black-Scholes sigue siendo la misma con el ajuste de las variables de la ec. 6.5 a la ec. 6.7:

$$C_t = A_t \Phi(d_1) - P_T e^{-r(T-t)} \Phi(d_2) \quad (6.9)$$

Donde

$$d_1 = \frac{\ln\left(\frac{A_t}{P}\right) + \left(r + \frac{1}{2}\sigma_v^2\right)(T-t)}{\sigma_v \sqrt{T-t}} \quad (6.10)$$

$$d_2 = d_1 - \sigma_v \sqrt{T-t} \quad (6.11)$$

Y para obtener la probabilidad de incumplimiento del pago de la deuda, antes descrita por la ec. 6.9 se tiene :

$$\mathbb{P}\{A_T < P | A_t\} = 1 - \Phi(d_2) = \Phi(-d_2) = \int_{-\infty}^{d_2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad (6.12)$$

La razón principal por la cual se realiza el ajuste de las variables sobre las razones financieras como activo, pasivo y capital, es que esta información esta al alcance público.

6.3.2 Modelo de Merton considerando movimiento browniano fraccional

La motivación de considerar la teoría de fractales y de movimientos brownianos fraccionales es la de mejorar la modelación de ciertos fenómenos financieros, esto porque últimamente se ha encontrado que algunos de los supuestos financieros y/o matemáticos no describen la realidad, por lo que se requiere de teorías más generales que expliquen estas diferencias y que incluyan como casos particulares a las que ya existen.

El movimiento browniano fraccional (MBF) es un proceso más general, que como un caso particular contempla variables con características de independencia y descritas con movimientos brownianos tradicionales.

Para considerar el precio de un call europea con movimientos brownianos fraccionales podemos considerar a Sierra (2007), en este trabajo se desarrolla la fórmula para el precio de una opción call europea en un mercado Browniano fraccional. La ecuación original de Black-Scholes asume que el comportamiento del subyacente asociado al derivado puede modelarse con un movimiento browniano tradicional.

Para poder llegar a la ecuación de Black-Scholes Fraccional, Sierra considera una cuenta bancaria libre de riesgo y una acción en un mundo neutral al riesgo modelado por un movimiento geométrico browniano fraccional.

El precio de la opción call europea dentro de un período de tiempo $0 \leq t \leq T$, con un precio de ejercicio K , una tasa r libre de riesgo, una volatilidad σ^2 , una distribución normal estándar $N(\cdot)$ es, un vencimiento en el tiempo T y un coeficiente de Hurst⁵ (H), esta dado por⁶:

$$C(t, S(t)) = S(t)N(d_1) - Ke^{-r(T-t)}N(d_2) \quad (6.13)$$

Donde:

$$d_1 = \frac{\ln\left(\frac{S(t)}{K}\right) + r(T-t) + (\sigma_v^2/2)\sqrt{(T^{2H}-t^{2H})}}{\sigma_v\sqrt{(T^{2H}-t^{2H})}} \quad (6.14)$$

$$d_2 = \frac{\ln\left(\frac{S(t)}{K}\right) + r(T-t) - (\sigma_v^2/2)\sqrt{(T^{2H}-t^{2H})}}{\sigma_v\sqrt{(T^{2H}-t^{2H})}} \quad (6.15)$$

Para desarrollar el modelo de Merton con la aportación de los Brownianos Fraccionales es necesario usar la metodología del modelo tradicional e introducir la fórmula de Black-Scholes fraccional desarrollada por Sierra (2007).

Básicamente se parte de las ecuaciones originales, con la diferencia de que σ_v ya no será una varianza fija, este supuesto se romperá para poder tener una varianza que pueda cambiar en el tiempo y que pueda ser dependiente de los sucesos anteriores.

Con un proceso de movimiento browniano fraccional, las ecuaciones y soluciones no cambian radicalmente, la diferencia es que dependen del valor del coeficiente Hurst y del tiempo inicial del proceso.

Se tiene de nuevo que :

$$C_T = \max(A_T - P_T, 0)$$

Donde A_T es el activo total de la empresa, C_T es el capital de la empresa y P_T es el valor de los bonos, todos en el tiempo T . Para la formulación de la ecuación del precio de compra de una opción europea, por Black-Sholes, se tiene:

$$C_t = A_t\Phi(d_1) - P_T e^{-r(T-t)}\Phi(d_2)$$

Antes de definir a d_1 y d_2 es necesario señalar que aquí es donde entran el valor del exponente Hurst y por lo tanto el movimiento Browniano fraccional.

⁵ El parámetro Hurst sirve para detectar características de persistencia en la serie.

⁶ Si se quiere conocer más sobre el desarrollo o demostración de la fórmula de Black-Scholes fraccional, revisar con detenimiento a Sierra (2007).

$$d_1 = \frac{\ln\left(\frac{A_t}{P_T}\right) + r(T-t) + (\sigma_v^2/2)\sqrt{(T^{2H}-t^{2H})}}{\sigma_v\sqrt{(T^{2H}-t^{2H})}} \quad (6.16)$$

$$d_2 = \frac{\ln\left(\frac{A_t}{P_T}\right) + r(T-t) - (\sigma_v^2/2)\sqrt{(T^{2H}-t^{2H})}}{\sigma_v\sqrt{(T^{2H}-t^{2H})}} \quad (6.17)$$

Recordando que al asignar el valor de $H=0.5$ se obtiene la ecuación de brownianos tradicionales.

Y por último la fórmula para obtener probabilidad de incumplimiento del pago de la deuda (ec. 6.14) :

$$\mathbb{P}\{A_T < P|A_t\} = 1 - \Phi(d_2) = \Phi(-d_2) = \int_{-\infty}^{d_2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Para obtener el modelo fraccional es necesario antes obtener el exponente de Hurst de cada serie, el método de obtención se explica a continuación.

6.3.3 Fractales y Exponente Hurst (H)

La geometría fractal esta formada por todo ese conjunto de estructuras irregulares que no pueden ser descritas por las matemáticas tradicionales, también es conocida como geometría de la naturaleza y sus principales características son la autosimilaridad⁷ y la dimensión fractal.

El movimiento fractal es utilizado por Sierra (2007) para deducir una forma general de la valuación de derivados, de la ecuación de Black-Scholes, de la ecuación general de bonos y de la estructura de plazos del modelo de tasas de Vasicek. Dicha fórmula de Black-Scholes fractal será utilizada para desarrollar el modelo de Merton fractal.

El coeficiente de Hurst (H) debe su nombre a Harold Edwin Hurst quien fue el primero en estudiar las series fractales. H tiene como objetivo medir el grado de independencia o memoria que presenta la serie en sus observaciones, en otras palabras, que tanta relación existe entre las observaciones a través del tiempo; distingue entre una serie tradicional que es independiente y una serie fractal que presenta características de persistencia o anti persistencia. Peters (1994) explica el coeficiente, el significado y como obtenerlo.

Dependiendo del valor de H la serie indica:

- $H = 0.5$, la serie presenta características tradicionales de independencia.

⁷ Propiedad que posee un objeto en el que todo es aproximadamente o exactamente similar a una parte de sí mismo.

- $0.5 < H \leq 1$, la serie es persistente, caracterizada por efectos de memoria larga. Lo que suceda hoy impactará el futuro. Se presentan comúnmente en la naturaleza, en el mercado de capitales y en la economía.
- $0 \leq H < 0.5$, implica antipersistencia. Algunos teóricos igualan este comportamiento con un proceso de reversión a la media, se vuelve a lo niveles previos medios.

La técnica para encontrar el coeficiente de H se conoce como rango escalado y es explicada en Peters (1994) o Sierra (2007).

Dado que el número de elementos en la muestra puede alterar la significancia del exponente, se debe de plantear una prueba t. Con un número limitado de elementos no se asegura la significancia del coeficiente, se recomienda que sean más de 10. El exponente de Hurst es sensible al tamaño de la muestra y al momento de medición.

Se plantea una prueba de hipótesis tradicional.

H_0 : $H = 0.5$ Comportamiento de caminata aleatorio o browniano tradicional. Características de independencia.

H_a : $H \neq 0.5$ Comportamiento persistente o antipersistente.

Se espera que el coeficiente de Hurst tenga un valor⁸ de:

$$E\left(\frac{R}{S_N}\right) = \left(\frac{n-0.5}{n}\right) \left(n * \pi/2\right)^{-0.50} * \sum_{r=1}^{n-1} \sqrt{\frac{n-r}{r}}$$

$$Var(H)_n = 1/T$$

A partir de estas dos ecuaciones podemos determinar el nivel de significancia. Es importante señalar que el exponente depende del tamaño de la muestra y del momento de medición. Si H va en decremento la serie muestra tendencia a alejarse de las características de persistencia.

6.3.4 Determinación de la Muestra

Para conformar la muestra utilizada se seleccionan entre las empresas mexicanas que cotizaban en la BMV a las que emitan bonos corporativos no perpetuos con vencimiento fijo debido al supuesto que el modelo de Merton plantea, esto debido a la disponibilidad de información⁹.

⁸ Para una explicación más extensa ver Peters (1994).

El análisis empírico está formado por un período de backtesting o prueba de retrospectiva de años pasados y otro período de pronóstico. El período de backtesting comprende del 2007 al 2010, y el período de pronóstico sólo abarca el año 2013; ambos períodos implementa el modelo de Merton en sus dos versiones.

El período de estudio backtesting se centra en el año 2007 con la intención de hacer una simulación acerca de la probabilidad de incumplimiento para el período de la crisis inmobiliaria. Es decir, analizaremos si estas empresas podían incumplir con sus obligaciones contraídas en los años siguientes (2008-2010). Para ello se utilizan datos de los estados financieros anuales en miles de pesos corrientes de cada empresa para el período comprendido. Se cuenta con datos trimestrales desde el 2000 hasta el 2012.

Para realizar el pronóstico se utiliza a las empresas que tengan bonos con vencimiento en el 2013, y se analiza la situación desde el último estado financiero, que es el primer trimestre del 2012.

Para seleccionar las empresas se toman a las que están incluidas en el Índice de Precios y Cotizaciones (IPC), ya que del total de las que cotizan en la BMV, estas son las que presentan mayor bursatilidad. Es importante señalar que las empresas deben de contar con bonos corporativos con vencimiento en 2008, 2009 y 2010, así como 2013. Sólo diez empresas de las 35 incluidas en el IPC cumplen con esta características, pero representan 53.26% del índice.

La muestra final está conformada por diez empresas de diferentes sectores que cumplen con los requisitos del período de evaluación. Las empresas analizadas son: Arca Continental, América Móvil, Grupo Bimbo, Cemex, Fomento Económico Mexicano, Corporación Geo, Grupo Financiero Banorte, Kimberly-Clark de México, Grupo Televisa y Urbi Desarrollos Urbanos.

Para el cálculo de la deuda se emplea un procedimiento más complejo, debido a que la empresa puede tener varios o ningún bono en cada año analizado. La deuda se calcula sumando los bonos que cada empresa con vencimiento en cada uno de los años, cada bono se lleva a un fecha en común, utilizando la fórmula de interés simple, esto con el fin de facilitar el cálculo de la suma.

⁹ Para la obtención de los datos se utilizan diferentes proveedores de precios, tales como VALMER y BLOOMBERG; y de software financieros como Economática e Infosel.

La volatilidad de los activos financieros es la forma tradicional de considerar la varianza de los rendimientos de estos activos¹⁰. Para el cálculo de la volatilidad se plantean tres procedimientos: la volatilidad histórica, el proceso ARMA y el proceso GARCH. Para los dos últimos casos fue necesario realizar un análisis para obtener la volatilidad de cada empresa en cada período. Los criterios de decisión entre los diferentes modelos fueron: coeficientes significativos, varianza mínima, errores ruido blanco y criterio Akaike.

La **volatilidad histórica** es la volatilidad del precio de un activo calculada a partir de datos históricos, es por lo tanto una medición ex post. Esta es una medida estadística del movimiento pasado de los precios. Es un proceso en el cual las ponderaciones que se dan a la información pasada son fijas. Este procedimiento no explota de forma óptima el efecto clustering¹¹.

El proceso autorregresivo y de promedios móviles (**ARMA**) es utilizado en los pronósticos del análisis econométrico. Como su nombre lo indica, el proceso incluye un proceso autorregresivo (AR) y un proceso de medias móviles (MA). El proceso AR es en el cual el valor en el período actual depende de los valores en períodos anteriores. Por su parte el proceso MA es el resultado de una combinación lineal de una constante más un promedio móvil de los términos de error del presente y del pasado.

Este proceso permite estimar la estructura intertemporal de la volatilidad, tiende a un nivel de equilibrio de largo plazo (volatilidad paramétrica), optimiza el efecto de clustering, pero las estimaciones pueden ser estadísticamente inconsistentes.

El modelo ARMA es una generalización del modelo de volatilidad histórica. La diferencia es que el modelo de volatilidad histórica tiene ponderaciones fijas y el ARMA las determina por los datos mediante un mecanismo estadístico, es así como disminuye el problema de sobreestimación del efecto clustering.

El proceso autorregresivo generalizado con heterocedasticidad condicional (**GARCH**), supone que la varianza cambia a través del tiempo¹². Estos procesos se caracterizan por la acumulación de la volatilidad, esto significa que existen lapsos con amplias variaciones durante prolongados períodos seguidos por un intervalo de tranquilidad relativa. Su propósito es estimar la varianza no

¹⁰ Según la teoría de los mercados eficientes, los precios de los activos financieros siguen un comportamiento aleatorio. El rendimiento de los instrumentos financieros muestran una media y desviación estándar estables a través del tiempo.

¹¹ Las volatilidades elevadas persisten por períodos largos antes de disminuir a sus niveles de largo plazo.

¹² Desarrollado por Engle (1982).

condicional de los rendimientos de los activos financieros. El modelo *GARCH* (1,1) es el modelo más común en la estimación de la volatilidad.

El modelo *GARCH* a diferencia de los modelos *ARMA*, donde la varianza se estima con base en los rendimientos observados, utiliza los rendimientos esperados¹³. Otra diferencia es que el modelo *GARCH* no sobreestima el efecto de apalancamiento¹⁴, por lo cual aprovecha mejor el capital.

En los modelos de incumplimiento usualmente se toma la volatilidad del precio de las acciones para los estudios financieros, pero en este trabajo se analizó que la volatilidad del precio de las acciones y la volatilidad de los activos no tiene una relación directa como tal. Esto se puede deber a que la volatilidad del precio de las acciones trae consigo especulación de los mercados y correlación con otros tipos de acciones. Dado que se observa que la volatilidad de los activos es más estable que la de las acciones, se utilizará la volatilidad de los activos para el análisis.

6.4 Resultados

La razón financiera de Activo-Pasivo nos indica la capacidad de sustentabilidad que tiene una empresa; en otras palabras, esta razón muestra cuantas veces la empresa puede pagar sus pasivos con sus activos, a mayor sea el número mayor sustentabilidad, a menor número menor sustentabilidad. La gráfica 6.1 muestra la tendencia que ha tenido esta razón para las empresas analizadas con la finalidad de evaluar dichas empresas.

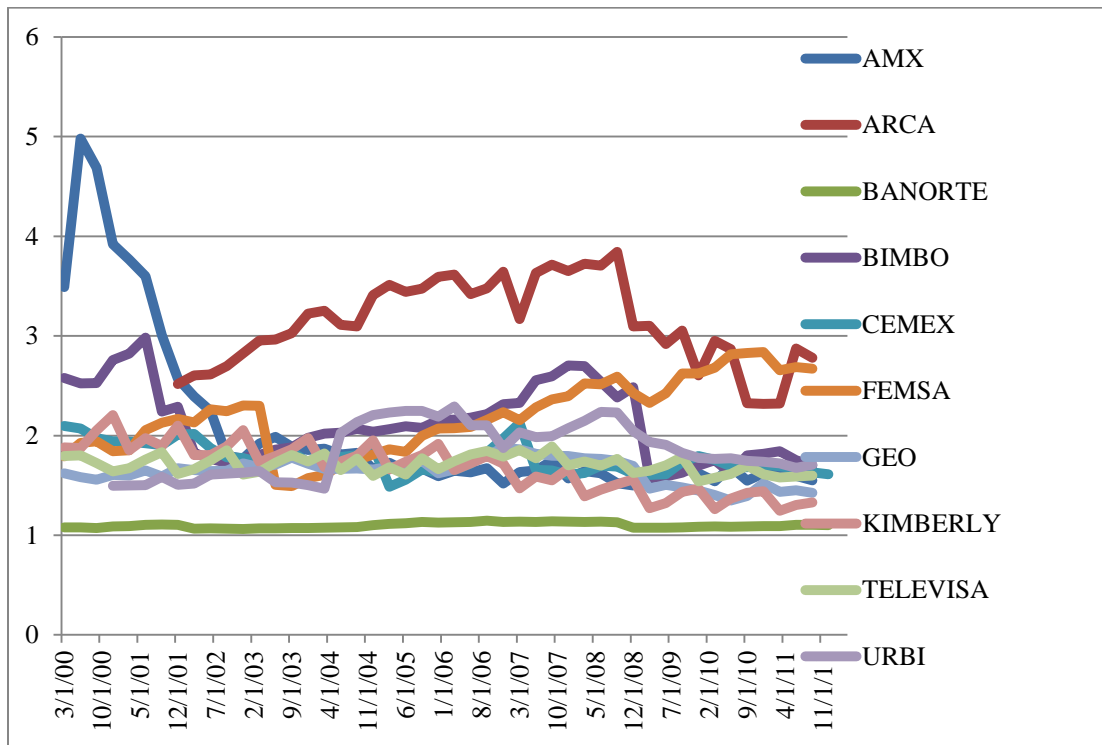
Las empresas con mayores razones de activo-pasivo son Arca, Femas y Bimbo, por lo que tienen un buen apalancamiento de deuda. Las empresas con razones de activo-pasivo menores son Televisa y Kimberly. La situación o percepción de las empresas ante los calificadores Standard & Poor's y Fitch, reportadas en Febrero del 2012¹⁵, reporta para las empresas ARCA, América Móvil, FEMSA, Kimberly y Televisa una capacidad de pago extremadamente fuerte y estable. Para el caso Bimbo y Banorte una capacidad de pago fuerte, pero la primera tiene tendencia negativa, lo que significa que la calificación de Bimbo puede bajar. Para el caso de

¹³ Dado que los rendimientos esperados no son observables, la estimación de la volatilidad necesita de métodos de optimización.

¹⁴ Cuando los rendimientos aumentan la volatilidad varía más que proporcionalmente.

¹⁵ No se incluyen calificaciones anteriores debido a que es complicado recabar información anterior.

GEO y Urbi una capacidad de pago fuerte y estable. Y por su parte Cemex muestra incertidumbre a circunstancias adversas y una tendencia negativa en su calificación.



Gráfica 6.1. Razón Activo-Pasivo. El eje vertical muestra el coeficiente obtenido
Elaboración propia con datos de los estados financieros de cada empresa tomados de Infosel y Económica.

6.4.1 Prueba de Retrospectiva

El análisis backtesting o prueba de retrospectiva toma lugar en el período de 2007 al 2010. La idea es posicionar el estudio en el año 2007 y analizar cual es la probabilidad de incumplimiento para cada empresa cada uno de los siguientes años (2008-2010).

En primer caso se realizó el cálculo de la probabilidad de incumplimiento con el modelo de Merton tradicional. Se calcula con los tres diferentes estimaciones de la volatilidad antes mencionados.

Probabilidad de Incumplimiento, Merton Tradicional					
2008					
r = 7.68%	V_t	D	Probabilidad de Incumplimiento		
			VOL HIST	ARMA	GARCH
AMX	\$ 332,443,292	\$ 5,355,027	0.0002%	0.0000%	0.0000%
ARCA	\$ 16,705,269	\$ 1,531,852	0.0470%	0.0004%	0.0000%
BIMBO	\$ 41,482,285	\$ 2,938,775	0.0014%	0.0021%	0.0000%
CEMEX	\$ 484,731,600	\$ 3,471,394	0.0000%	0.0000%	0.0000%
FEMSA	\$ 77,458,168	\$ 1,750,274	0.0000%	0.0046%	0.0000%
GEO	\$ 17,692,924	\$ 469,228	0.0000%	0.0000%	0.0000%
TELEVISA	\$ 77,803,320	\$ 51,033,572	20.6495%	18.3370%	4.6389%
URBI	\$ 18,295,891	\$ 826,423	0.0000%	0.0003%	0.0006%
2009					
r = 5.42%	V_t	D	Probabilidad de Incumplimiento		
			VOL HIST	ARMA	GARCH
AMX	\$ 332,443,292	\$ 2,467,624	1.1384%	0.0163%	0.0000%
BIMBO	\$ 41,482,285	\$ 1,953,727	3.5236%	0.9249%	0.0006%
CEMEX	\$ 484,731,600	\$ 5,224,279	0.0003%	0.0130%	0.0000%
FEMSA	\$ 77,458,168	\$ 1,870,706	0.0000%	10.0790%	0.0150%
GEO	\$ 17,692,924	\$ 686,598	0.0000%	0.0008%	0.0002%
2010					
r = 4.4%	V_t	D	Probabilidad de Incumplimiento		
			VOL HIST	ARMA	GARCH
AMX	\$ 332,443,292	\$ 4,392,365	7.2553%	0.5613%	0.0094%
ARCA	\$ 16,705,269	\$ 2,938,390	0.3394%	4.0904%	1.6296%
BANORTE	\$ 261,204,948	\$ 4,692,258	0.4305%	1.0951%	0.0000%
CEMEX	\$ 484,731,600	\$ 3,576,867	0.0002%	0.0695%	0.0000%
KIMBERLY	\$ 21,656,917	\$ 4,367,768	0.4142%	3.1477%	0.0000%
TELEVISA	\$ 77,803,320	\$ 13,144,321	1.4969%	2.8830%	0.0120%
URBI	\$ 18,295,891	\$ 2,663,611	0.3886%	3.1850%	4.0477%

Tabla 6.1-. Probabilidades de incumplimiento por modelo de Merton con Brownianos Tradicionales.

Elaboración propia. Se entiende por V_t el valor de la empresa o activos totales, D como la deuda de los bonos y r como la tasa de interés. Se obtiene la probabilidad de incumplimiento con la volatilidad histórica, el proceso ARMA y el proceso GARCH.

Antes de hablar sobre las probabilidades de incumplimiento es importante analizar que tan relevante es la deuda sobre el activo total de cada empresa. Se observa que solo la deuda de bono por parte de Televisa en el año 2008 representa un porcentaje alarmante con el 59% respecto a su activo total. Para el año 2009 ninguna es muy significativa. Y para el año 2010 Arca, Kimberly, Televisa y Urbi están entre el 12% y el 16%.

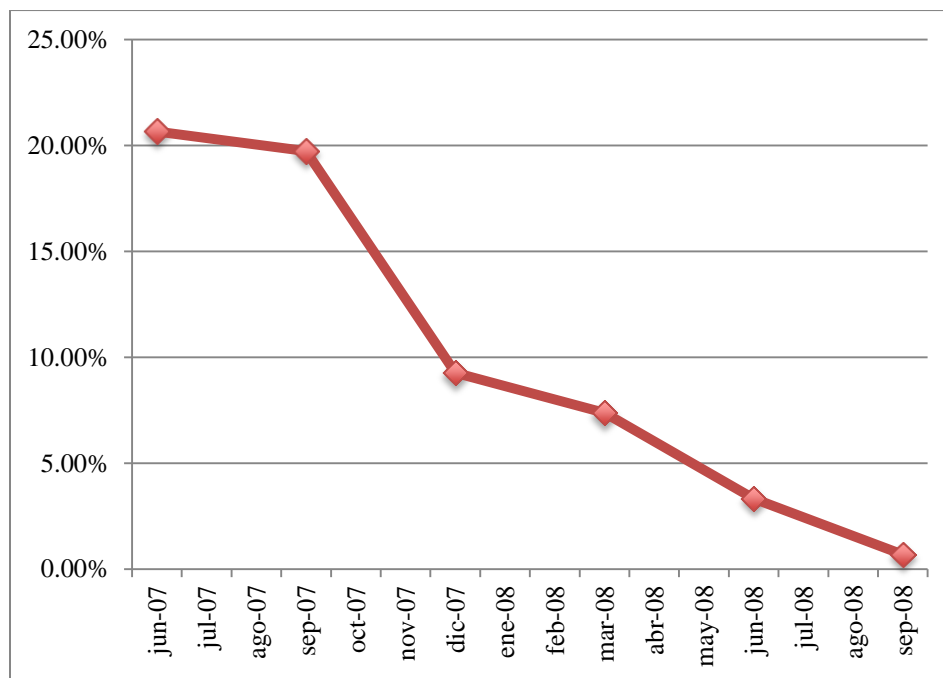
Podemos observar como con cada una de las diferentes volatilidades se generan diferentes probabilidades de incumplimiento, esto debido a que la fórmula del modelo de Merton es muy sensible a la varianza. Aunque se esperaba obtener una probabilidad de incumplimiento mayor para Cemex, esto no ocurre, se puede deber a que la deuda no representa una cantidad muy elevada sobre sus activos y además de que la empresa logró crear un plan de refinanciamiento¹⁶.

Se observa que, para la mayoría de las empresas, la probabilidad de incumplimiento es mayor para el caso del método de volatilidad histórica y del proceso ARMA. Es importante recordar que cada método de volatilidad se calcula de manera diferente; la volatilidad histórica asigna ponderaciones fijas, el proceso ARMA permite la estructura intertemporal y optimiza el efecto clustering, y el proceso GARCH modela diferentes características de la distribución de rendimientos.

Para el 2008 Televisa fue la empresa que arrojó una probabilidad de incumplimiento alarmante, para el 2009 todas tienen buenos resultados a excepción de FEMSA con el modelo ARMA. Y para el 2010 todas se mantienen por debajo del 8% de probabilidad de incumplimiento.

Para saber que tan sensible es el modelo de Merton respecto al tiempo se realiza un ejercicio con la deuda de Televisa con vencimiento en 2008, ya que este caso fue el que arrojó resultados de probabilidad de incumplimiento más altos. Se recorre el tiempo desde el 2007 hasta llegar al vencimiento de la deuda en 2008. Se utiliza la volatilidad histórica debido a que fue la más alta.

¹⁶ Cemex crea en diciembre del 2008 su plan de refinanciamiento, el cual logra con cinco bancos en enero del 2009, con la finalidad de extender su vencimiento un año más.



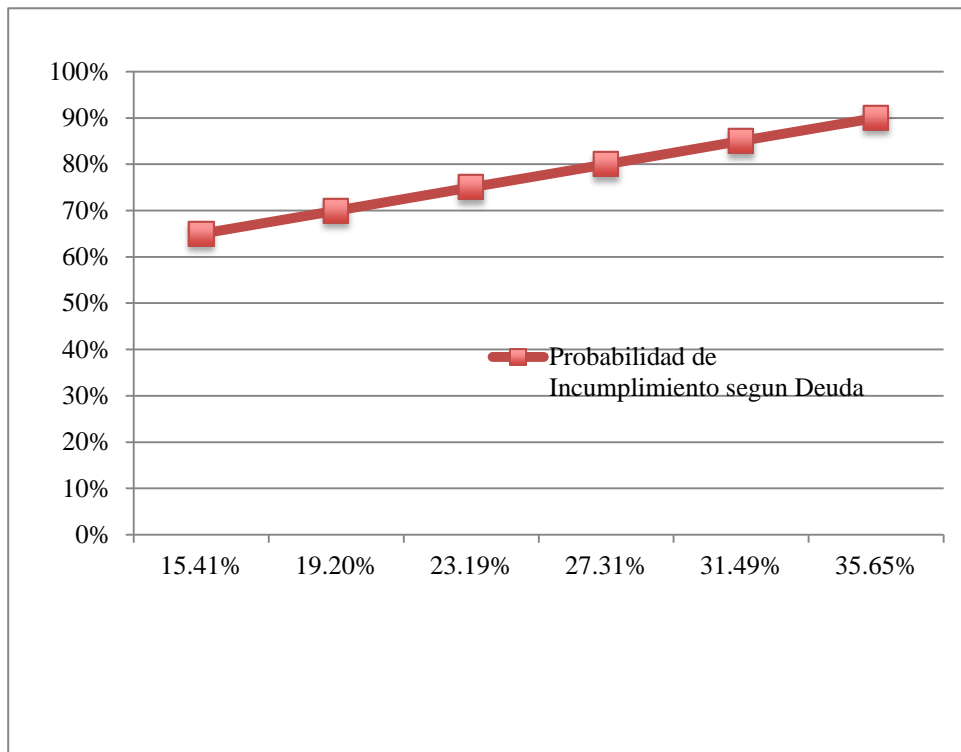
Gráfica 6.2.- Probabilidad de incumplimiento de Televisa en el año 2008.

Elaboración Propia.

Se tomó la deuda y se corre hasta el vencimiento en diciembre del 2008, el valor de la empresa fue cambiando según el valor en libros de activo total para cada trimestre. En este caso la probabilidad de incumplimiento fue bajando debido a que Televisa tuvo un crecimiento significativo en sus activos totales, lo cual hacía más posible cubrir su deuda. Si se deja el activo fijo, se obtiene un incremento en la probabilidad de incumplimiento debido a que el tiempo para cubrir la deuda es menor.

Para analizar la sensibilidad del modelo a la deuda se realiza un ejercicio parecido pero ahora se supone el caso en el que la deuda aumenta. Tomando de nuevo el caso de Televisa, se aumenta la deuda como porcentaje del activo total, se obtiene que cuanto mayor es la deuda, mayor es la probabilidad de incumplimiento dejando fijo el activo total.

En la gráfica 6.3 se puede observar como va aumentando la probabilidad de incumplimiento, el eje de las Y's es el porcentaje que representa la deuda sobre el activo total y el eje de las X's es la probabilidad de incumplimiento. Es importante tener en consideración que deudas muy pequeñas no son relevantes debido a que el modelo es muy sensible al monto de deuda.



Gráfica 6.3-. Probabilidad de Incumplimiento según porcentaje de Deuda sobre Activo Total.
Elaboración Propia.

6.4.2 Modelo de Merton con Brownianos Fraccionales

En esta sección se estima la probabilidad de incumplimiento por el modelo de Merton utilizando brownianos fraccionales. Antes de obtener la probabilidad es necesario obtener el coeficiente de Hurst para poder utilizar el modelo.

La tabla 2 muestra la prueba de hipótesis del coeficiente de Hurst para probar su significancia. Para obtener el coeficiente de Hurst fue necesario realizar el proceso de rango escalado.

	T	n	H	E(H)	DES STD (H)	H-E(H)/DSTD(H)	H ₀
AMX							
2008	30	6	0.6740	0.6762	0.1826	-0.0122	ACEPTO
2009	36	7	0.5409	0.6762	0.1667	-0.8119	ACEPTO
2010	40	6	0.5877	0.6762	0.1581	-0.5600	ACEPTO
ARCA							
2008	24	6	0.5848	0.6762	0.2041	-0.4479	ACEPTO
2010	30	6	0.5145	0.6762	0.1826	-0.8859	ACEPTO
BANORTE							
2010	42	6	0.6350	0.6762	0.1543	-0.2668	ACEPTO
BIMBO							
2008	30	6	0.6737	0.6762	0.1826	-0.0136	ACEPTO
2009	36	7	0.6846	0.6762	0.1667	0.0504	ACEPTO
CEMEX							
2008	30	6	0.6513	0.6762	0.1826	-0.1367	ACEPTO
2009	36	7	0.6135	0.6762	0.1667	-0.3764	ACEPTO
2010	42	6	0.6442	0.6762	0.1543	-0.2072	ACEPTO
FEMSA							
2008	30	6	0.5799	0.6762	0.1826	-0.5275	ACEPTO
2009	36	7	0.5772	0.6762	0.1667	-0.5939	ACEPTO
GEO							
2008	30	6	0.6637	0.6762	0.1826	-0.0684	ACEPTO
2009	36	7	0.5687	0.6762	0.1667	-0.6450	ACEPTO
KIMBERLY							
2010	42	6	0.5878	0.6762	0.1543	-0.573143746	ACEPTO
TELEVISA							
2008	30	6	0.5947	0.6762	0.1826	-0.446191227	ACEPTO
2010	42	6	0.5375	0.6762	0.1543	-0.899086119	ACEPTO
URBI							
2008	30	6	0.6044	0.6762	0.1826	-0.393122388	ACEPTO
2010	40	6	0.6117	0.6762	0.1581	-0.408136204	ACEPTO

Tabla 6.2-. Prueba de hipótesis para el coeficiente de Hurst.

Elaboración propia. Siendo T el número de observaciones de la serie, n las observaciones del rango escalado, H el coeficiente obtenido, E(H) el esperado de Hurst, DesStd(H) la desviación estándar, después $H-E(H)/DSTD(H)$ es el cálculo y H₀ (Browniano tradicional) señala si la hipótesis nula es aceptada o rechazada.

Al no poder rechazar H₀, aunque los coeficientes son mayores a 0.50, no se puede afirmar que las series tienen características de persistencia debido a que no superan la prueba de significancia. Este problema se debe a la muestra ya que son menos de 10 observaciones, se sugiere un número

mayor a 10 observaciones en el rango escalado para afirmar que los resultados son certeros. Este es un inconveniente con la información, pero como ya se obtuvo el coeficiente se calculara el modelo de Merton fraccional.

Ya que se obtiene el coeficiente de Hurst se ingresa a la fórmula del modelo de Merton fraccional y se obtiene la probabilidad que arrojaría si los coeficientes de Hurst llegarán a ser persistentes.

Probabilidad de Incumplimiento, Merton Fractal					
2008					
r = 7.68%	Vt	D	Probabilidad de Incumplimiento		
			VOL HIST	ARMA	GARCH
AMX	\$ 332,443,292	\$ 5,355,027	0.0008%	0.0000%	0.0000%
ARCA	\$ 16,705,269	\$ 1,531,852	0.0725%	0.0008%	0.0000%
BIMBO	\$ 41,482,285	\$ 2,938,775	0.0052%	0.0074%	0.0000%
CEMEX	\$ 484,731,600	\$ 3,471,394	0.0000%	0.0002%	0.0000%
FEMSA	\$ 77,458,168	\$ 1,750,274	0.0000%	0.0081%	0.0000%
GEO	\$ 17,692,924	\$ 469,228	0.0000%	0.0000%	0.0000%
TELEVISA	\$ 77,803,320	\$ 51,033,572	21.8874%	19.5552%	5.3581%
URBI	\$ 18,295,891	\$ 826,423	0.0000%	0.0008%	0.0014%
2009					
r = 5.42%	Vt	D	Probabilidad de Incumplimiento		
			VOL HIST	ARMA	GARCH
AMX	\$ 332,443,292	\$ 2,467,624	1.5305%	0.0291%	0.0000%
BIMBO	\$ 41,482,285	\$ 1,953,727	7.9018%	2.9076%	0.0142%
CEMEX	\$ 484,731,600	\$ 5,224,279	0.0027%	0.0607%	0.0000%
FEMSA	\$ 77,458,168	\$ 1,870,706	0.0000%	13.1403%	0.0425%
GEO	\$ 17,692,924	\$ 686,598	0.0000%	0.0028%	0.0006%
2010					
r = 4.4%	Vt	D	Probabilidad de Incumplimiento		
			VOL HIST	ARMA	GARCH
AMX	\$ 332,443,292	\$ 4,392,365	11.6613%	1.4103%	0.0499%
ARCA	\$ 16,705,269	\$ 2,938,390	0.3880%	4.3881%	1.7887%
BANORTE	\$ 261,204,948	\$ 4,692,258	1.7479%	3.4873%	0.0000%
CEMEX	\$ 484,731,600	\$ 3,576,867	0.0081%	0.5383%	0.0001%
KIMBERLY	\$ 21,656,917	\$ 4,367,768	0.9854%	5.2552%	0.0000%
TELEVISA	\$ 77,803,320	\$ 13,144,321	2.0037%	3.6672%	0.0239%
URBI	\$ 18,295,891	\$ 2,663,611	1.1747%	6.1089%	7.3865%

Tabla 6.3-. Probabilidades de incumplimiento por modelo de Merton con Brownianos Fraccionales.

Elaboración propia. Se entiende por V_t el valor de la empresa o activos totales, D como la deuda de los bonos y r como la tasa de interés. Se obtiene la probabilidad de incumplimiento con la volatilidad histórica, el proceso ARMA y el proceso GARCH.

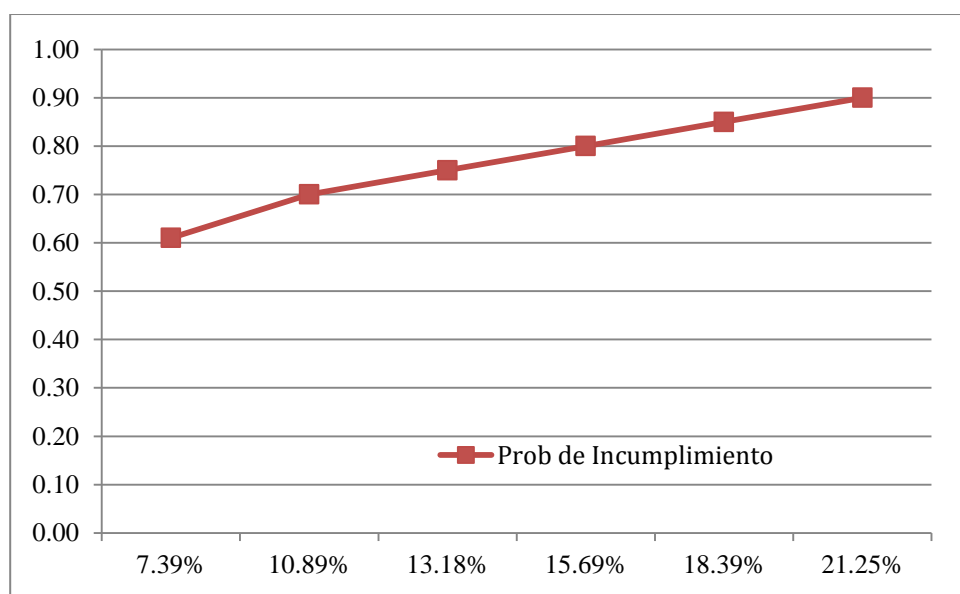
De nuevo se observa que la fórmula es sensible a la varianza y que por esto se obtienen probabilidades diferentes. Los coeficientes obtenidos no son mucho mayores a 0.50 por lo que la probabilidad de incumplimiento no cambia drásticamente entre el modelo tradicional y el fraccional, pero existe un aumento.

La persistencia en los activos significa que estos tienen memoria, se esperaba que los activos estudiados contaran con esta propiedad, aún cuando los resultados de los coeficientes de Hurst son mayores a 0.5 no fue posible garantizar este resultado.

Al comparar los resultados de la probabilidad de incumplimiento entre el modelo de Merton tradicional y el modelo de Merton fraccional, se observa un incremento en la probabilidad por parte del modelo fraccional. Esto podría indicar que el modelo tradicional subestima la probabilidad de incumplimiento y que el modelo de Merton fraccional logra una mejor estimación que hubiera sido útil en el período de crisis.

Se realiza un ejercicio para ver que tan sensible es el modelo de Merton Fraccional ante el coeficiente de Hurst. Se toma el caso de Urbi con su deuda en 2010, ya que es uno de los casos en los que el incremento de la probabilidad es mayor entre el modelo tradicional y el fraccional. La idea es ir aumentando el coeficiente de Hurst, la volatilidad utilizada para el cálculo es la de GARCH, debido a que arrojo mayor probabilidad.

La gráfica 6.4 muestra como al aumentar el coeficiente de Hurst, es decir, la medida de persistencia de la serie, la probabilidad de incumplimiento aumenta. Entre mayor sea la persistencia mayor será la probabilidad de incumplimiento. El eje de las Y's es el valor del coeficiente de Hurst y el eje de las X's es la probabilidad de incumplimiento obtenida.



Gráfica 6.4-. Probabilidad de Incumplimiento con cambios en el coeficiente de Hurst

Elaboración Propia.

6.4.3 Pronóstico de Incumplimiento para 2013

Se realiza el cálculo de la probabilidad de incumpliendo para la empresas que tienen bonos con vencimiento en 2013; América Móvil, Geo, Femsa y Kimberly.

Nuevamente se inicia con el cálculo de la probabilidad de incumplimiento por el modelo de Merton tradicional para cada una de estas empresas con los tres métodos de volatilidad utilizados, resultados en la siguiente tabla:

Probabilidad de Incumplimiento, Merton Tradicional					
2013					
r = 4.39%	Vt	D	Probabilidad de Incumplimiento		
			VOL HIST	ARMA	GARCH
AMX	\$ 255,796,491	\$ 7,275,528	5.8483%	0.1617%	0.0035%
GEO	\$ 42,070,785	\$ 301,465	0.0000%	0.0000%	0.0000%
FEMSA	\$ 255,796,491	\$ 5,527,609	0.0003%	3.3253%	0.0007%
KIMBERLY	\$ 29,017,440	\$ 2,996,145	0.0000%	0.0231%	0.0000%

Tabla 6.4-. Probabilidades de incumplimiento por modelo de Merton con Brownianos Tradicionales.

Elaboración propia. Se entiende por V_t el valor de la empresa o activos totales, D como la deuda de los bonos y r como la tasa de interés. Se obtiene la probabilidad de incumplimiento con la volatilidad histórica, el proceso ARMA y el proceso GARCH.

La deuda de bonos sobre los activos totales solo es significativa para el caso de Kimberly, tiene una deuda de bonos que representa el 10% de sus activos totales.

América Móvil obtiene la probabilidad de incumplimiento mayor bajo el método de volatilidad histórica, aunque ninguno de los resultados es alarmante.

6.4.4 Modelo de Merton con Brownianos Fraccionales

Para obtener la probabilidad de incumplimiento por el modelo fraccional se debe de obtener el coeficiente de Hurst y probar su significancia.

La tabla 5 muestra los resultados y la prueba de hipótesis:

	T	n	H	E(H)	DES STD (H)	H- E(H)/DSTD(H)	H₀
AMX							
2013	48	8	0.5425	0.6762	0.1443	-0.9264	ACEPTO
FEMSA							
2013	48	8	0.6633	0.6762	0.1443	-0.0892	ACEPTO
GEO							
2013	48	8	0.5935	0.6762	0.1443	-0.5728	ACEPTO
KIMBERLY							
2013	48	8	0.5135	0.6762	0.1443	-1.127318432	ACEPTO

Tabla 6.5-. Prueba de hipótesis para el coeficiente de Hurst para pronóstico.

Elaboración propia. Siendo T el número de observaciones de la serie, n las observaciones del rango escalado, H el coeficiente obtenido, E(H) el esperado de Hurst, DesStd(H) la desviación estándar, después $H-E(H)/DSTD(H)$ es el cálculo y H_0 (Browniano tradicional) señala si la hipótesis nula es aceptada o rechazada.

Aun cuando los coeficientes obtenidos son mayores a 0.5 no se puede probar su significancia, esto es un problema con los datos porque es recomendable una muestra de rango escalado mayor a 10 observaciones. Aunque no pasaron la prueba de hipótesis se utilizaran estos coeficiente obtenidos para el modelo fraccional.

Probabilidad de Incumplimiento, Merton Fractal					
2013					
r = 4.39%	Vt	D	Probabilidad de Incumplimiento		
			VOL HIST	ARMA	GARCH
AMX	\$ 255,796,491	\$ 7,275,528	6.5355%	0.2096%	0.0054%
GEO	\$ 42,070,785	\$ 301,465	0.0000%	0.0000%	0.0000%
FEMSA	\$ 255,796,491	\$ 5,527,609	0.0023%	5.4458%	0.0044%
KIMBERLY	\$ 29,017,440	\$ 2,996,145	0.0000%	0.0257%	0.0000%

Tabla 6.5-. Probabilidades de incumplimiento por modelo de Merton con Brownianos Fraccionales Elaboración propia. Se entiende por V_t el valor de la empresa o activos totales, D como la deuda de los bonos y r como la tasa de interés. Se obtiene la probabilidad de incumplimiento con la volatilidad histórica, el proceso ARMA y el proceso GARCH.

De nuevo tenemos que la probabilidad de incumplimiento varía para la misma empresa debido a las diferentes formas de calcular la volatilidad, esto por la sensibilidad que tiene la fórmula ante la volatilidad. Ninguna de estas 4 empresas tiene probabilidad de incumplimiento significativa para el año 2013, se vuelve a tener América Móvil con la probabilidad de incumplimiento más alta. Estos resultados coinciden con las calificaciones por parte de Standard&Poors y Fitch para las empresas publicadas en febrero del 2012.

Los resultados entre el modelo tradicional y modelo fraccional discrepan un poco, el modelo fraccional obtiene probabilidades ligeramente mayores. Que la diferencia entre modelos sea pequeña se puede deber a que los coeficientes de Hurst obtenidos no son muy grandes, recordando que $H=0.5$ es el movimiento browniano tradicional. Aunque la diferencia sea poca se podría afirmar que el modelo tradicional subestima la probabilidad de incumplimiento y esto indicaría que el modelo fraccional genera mejores estimaciones.

6.5 Conclusiones

Las crisis financieras anteriores y la especulación que se genera alrededor de las bolsas de valores ha puesto en evidencia la deficiencia en cuanto a la capacidad que tienen los modelos para

evaluar o predecir sobre el sistema financiero, es por ello que la meta principal de este trabajo era analizar el modelo de Merton tradicional y aportar una inclusión del movimiento browniano fractal al modelo, además de realizar un análisis empírico para evaluar ambos.

En el presente trabajo se presento el movimiento browniano tradicional y el fraccional, la diferencia entre ellos, es que el segundo incorpora características de persistencia propias de las series.

Para el análisis del modelo, se considera la volatilidad como la forma de considerar la varianza de los rendimientos de activos financieros. Se emplean tres métodos distintos para medirla utilizados en el modelo; la volatilidad histórica, el proceso ARMA y el proceso GARCH. La volatilidad histórica al ser una medida estadística del movimiento pasado con ponderaciones fijas se ve opacada por el proceso ARMA, el cual determina las ponderaciones mediante un mecanismo estadístico, con lo que logra disminuir el efecto clustering, además permite estimar la estructura intertemporal de la volatilidad. El proceso GARCH se caracteriza por la acumulación de la volatilidad, supone que la varianza cambia a través del tiempo. La diferencia entre ARMA y GARCH es que el primero estima con base a rendimientos observados y el segundo con rendimientos esperados, por lo que GARCH disminuyendo el efecto de apalancamiento.

Además se presenta el exponente de Hurst (H), el cual tiene como función distinguir entre series tradicionales y fractales, en otras palabras medir la persistencia en la serie. H es sensible al tamaño de la muestra y al momento de medición.

El modelo de Merton supone información perfecta en el mercado por parte de los inversionistas, con la aportación se introducen características de persistencia que buscaban mejorar las estimaciones. La aportación del modelo de Merton fractal se logra gracias al desarrollo de la fórmula de valuación de opciones Black-Scholes fractal por Sierra (2007). La adaptación fractal es un proceso más general ya que incluye al tradicional al asignar al exponente Hurst (H) un valor de 0.5.

La muestra se conformó por 10 de las 35 empresas mexicanas que conforman el IPC, las cuales emiten bonos corporativos no perpetuos con vencimiento fijo. El análisis empírico está dividido en dos períodos; el primero es una prueba de retrospectiva del 2007 al 2010; y el otro es el pronóstico del año 2013. Además se utilizó la volatilidad de los activo debido a que la volatilidad del precio de las acciones y la volatilidad de los activos no tiene una relación directa como tal, y la de los activos es más estable.

Los resultados muestran que en la prueba de retrospectiva se obtiene que las diferentes formas de medir la volatilidad generan diferentes probabilidades de incumplimiento, esto debido a que la fórmula del modelo de Merton es muy sensible a la varianza.

La sensibilidad del modelo al tiempo depende de la tendencia de los activos en el plazo del análisis, pero a menor plazo mayor probabilidad. La sensibilidad del modelo a la deuda indica que con montos muy pequeños de deuda el modelo es insensible. Los bonos contraídos por las empresas aquí evaluadas representaban un porcentaje pequeño del activo total, esto puede explicar porque no se presentan probabilidades altas de incumplimiento. De igual forma se analiza la sensibilidad del modelo ante el coeficiente de Hurst y se obtiene que entre mayor sea la persistencia de la serie mayor será la probabilidad de incumplimiento.

Uno de los problemas que se tuvo en el análisis fue que no se pudo comprobar la significancia del coeficiente de Hurst, esto debido a los datos, ya que se sugiere un número mayor a 10 observaciones en el proceso del rango escalado.

Al comparar los resultados de la probabilidad de incumplimiento entre el modelo de Merton tradicional y el modelo de Merton fraccional, se obtiene un incremento en la probabilidad de incumplimiento por parte del modelo fraccional, lo que podría indicar que el modelo tradicional subestima la probabilidad de incumplimiento y que el modelo de Merton fraccional logra una mejor estimación.

En general se obtiene que la modificación al modelo de Merton por medio de los brownianos fraccionales, la cual engloba las características de persistencia de las series, logró una buena estimación de la probabilidad de incumplimiento de las empresas.

En el pronóstico para el 2013 se encontró que las cuatro empresas analizadas no arrojan probabilidad de incumplimiento significativa, esto coincide con la valuación de Standard&Poors y Fitch de esta año (2012), por lo que se puede decir que el modelo de Merton genera buenas estimaciones.

La mayor limitación del modelo es la obtención de los datos, la información financiera de las empresas que cotizan en la BMV es pública, pero es difícil obtener la información de períodos pasados.

Como extensiones de esta investigación se podría realizar un análisis en dos períodos, es decir un bono con dos fechas de vencimiento, una de corto y otra de largo plazo; el utilizar datos constantes y por último analizar la volatilidad con el proceso EGARCH, el cual tiene como

objetivo modelar el efecto de apalancamiento e incorporar a la estimación la asimetría de la volatilidad en condiciones de alza o baja de los mercados. La volatilidad se estima mediante un modelo logarítmico por lo que no impone restricciones en los parámetros.

6.6 Bibliografía

Altman, Edward I. “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy”. *The Journal of Finance*, September 1968, Vol. 23, No. 4, pp. 589-609.

America Móvil. < <http://www.americamovil.com/amx/cm/investor/repQ.html>> Citado en Julio 2012.

Banco de México. “Reporte del sistema financiero” . Julio, 2009.

<http://www.banxico.org.mx/publicaciones-y-discursos/publicaciones/informes-periodicos/reporte-sf/{5286741D-A39E-9745-B393-AF3DF0A5AE85}.pdf>

Bank For International Settlements.< <http://www.bis.org/bcbs/about.htm>> Citado en Julio 2012.

Basilea. “Convenio de capital. Comité de Supervisión Bancaria de Basilea” . Suiza, 1988.

Basilea III y los retos de la banca.< http://www.ie.edu/IE/pdf/Informe_Basilea.pdf> Citado en Julio 2012.

Black, F and Scholes, M. “The pricing of options and corporate liabilities”. *Journal of Political Economy*. 1973.

BLOOMBERG <<http://www.bloomberg.com>> Citado en Julio 2012.

Bolsa Mexicana de Valores. <<http://www.bmv.com.mx>> Citado en Julio 2012.

Carlos, Vania. “*Aplicación empírica del modelo de Blanck & Scholes en México: 2004-2007*”. Tesis. Universidad de Guadalajara, 2010.

Cemex <<http://www.cemex.com>> Citado en Julio 2012.

Cerezo, E.; Bielsa, M, & Ramón, M. “Medición del riesgo de crédito mediante modelos estructurales: una aplicación al mercado colombiano”. *Cuadernos de Administración*, 2011, pp. 73-100.

Chen, C. & Panjer, H. “Unifying discrete structural models and reduced-form models in credit risk using a jump-diffusion process.” *Insurance: Mathematics and Economics*, 2003, 33 (2), pp. 357-380.

De Lara, H. A. Medición y control de riesgos financieros: Limusa Noriega Editores, 2003.

ECONOMÁTICA <<http://www.economica.com>> Citado en Julio 2012.

Eitman, David; Stonehill, Arthur & Moffett Michael. *Multinational Business Finance*: Prentice Hall, 2010.

Engle, R., “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”, *Econometrica*, 1982, vol. 50, núm. 1, pp. 987-1007.

Fitch Ratings. < <http://www.fitchmexico.com>> Citado en Julio 2012.

Geske, R. “The valuation of corporate liabilities as compound options”. *Journal of Financial and Quantitative Analysis*, 1977, 12 (4), pp. 541-552.

INFOSEL <<http://www.infosel.com.mx>> Citado en Julio 2012.

International Actuarial Association “*A global Framework for Insurer Solvency Assessment*”. Research Report of the Insurer Solvency Assessment Working Party, 2004.

Jackson, P.; Nickell, P. and Perraudin, W. “Credit risk modeling”. *Financial Stability Review*, pp. 94-121.

Jarrow, R. A. and Protter, P. “Structural versus reduced form models: A new information based perspective”. *Journal of Investment Management*, 2004, 2 (2), pp. 1-10.

JP Morgan and Company. “*CreditMetrics*”. Documento Técnico. JP Morgan, 1997.

Lu, Yuqian. “*Default Forecasting in KMV*”. Master thesis, University of Oxford, 2008.

Márquez, J. Una nueva visión del riesgo de crédito: Limusa - Noriega Editores, 2006.

Márquez Diez-Canedo, Javier. “Suficiencia de Capital y Riesgo de Crédito en Carteras de Préstamos Bancarios: Un modelo de impago que relaciona la suficiencia de capital con el riesgo de crédito, mide el riesgo de concentración, y determina límites individuales para los créditos”. Documento de Investigación, Banco de México, Abril 2002.

Mercado de Derivados <http://www.mexder.com.mx/MEX/vimex.html> Citado en Julio 2012.

Merton, R. “On the pricing of the corporate debt: the risk structure of interest rates”. *Journal of Finance*, 1974, 29 (2), pp. 449-470.

Moody’s Corporation. < <http://www.moodys.com/Pages/atc.aspx> > Citado en Julio 2012.

Moreno, Isaac “Aplicación del Modelo de Merton al Cálculo de la Distancia de Incumplimiento: Caso Bancomer, 1998-2003”. Tesina. Universidad Autónoma de México, 2007.

Peters, E. *Fractal Market Analysis*(Applying Chaos Theory to Investment and Economic): John Wiley and Sons. 1994.

Ruiz-Porras, Antonio. “Globalización, ciclos económicos y crisis global, 2007-2010.” *Expresión Económica*, Enero-Junio 2010, número 24.

Ruiz Porras, Antonio; Vásquez Quevedo Noemí & Núñez Mora, José Antonio. “Efectos de la globalización financiera en la administración y regulación de riesgos bancarios en México”. *Contaduría y Administración*, Universidad Nacional Autónoma de México, Agosto 2006, número 219, pp. 155-141.

Saavedra, María “Aplicación Empírica del modelo de Black y Scholes en México: 1991-2000”. *Contaduría y Administración*, Universidad Nacional Autónoma de México, Diciembre 2005, número 217, pp. 65-95.

Saavedra, Patricia, “Riesgo y los Acuerdos de Basilea II”, *Revista Laberintos e Infinitos*, Instituto Tecnológico Autónomo de México, 2004, núm. 10.

Sánchez, Carlos. Valor en riesgo y Otras Aproximaciones, *Valuación Análisis y Riesgo*: S.C, 2001.

Sierra, Guillermo. “Procesos Hurst y Movimiento Browniano Fraccional en Mercados Fractales”: *Valuación y Aplicaciones a los Derivados y Finanzas*, 2007.

Standard & Poors. <<http://www.standardandpoors.com/home/es/la>> Citado en Julio 2012.

VALMER <<http://www.valmer.com.mx>> Citado en Julio 2012.

Venegas, Francisco. *Riesgos financieros y económicos. Productos derivados y decisiones económicas bajo incertidumbre*: Cengage Learning, 2008.

Capítulo 7

Normas Internacionales de Información Financiera (IFRS) en la Administración de Riesgos

Noemí Vásquez Quevedo

Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Ciudad de México

Resumen

El capítulo presenta las Normas Internacionales de Información Financiera (IFRS) que regulan el tema de Instrumentos Financieros y por tanto, la administración de riesgos que se debe medir y revelar sobre dichos instrumentos. El objetivo principal del capítulo es mostrar un panorama de lo que estipulan las normas contables internacionales, para contextualizar en un entorno normativo, las repercusiones en los estados financieros de los diferentes modelos de valuación de riesgos. El capítulo presenta una descripción general de lo que cada una de las IFRS trata acerca del tema de riesgos financieros.

7.1 Introducción

Ante la creciente apertura de los mercados de capitales alrededor del mundo, el tema de la confiabilidad de la información financiera ha tomado cada vez mayor relevancia. Dicha confiabilidad está basada principalmente en la calidad de los reportes financieros, y para alcanzarla, el principal mecanismo es el regulatorio. De esta manera, las normas de contabilidad internacionales que permitan un lenguaje contable uniforme entre países y entidades, así como la

medición y revelación de los riesgos financieros, desempeñan hoy en día un papel muy importante en la conformación de información que sea de utilidad para los usuarios en su labor de toma de decisiones.

De acuerdo con Daske *et. al*, 2008, una mayor calidad de reportes financieros, con mejores revelaciones, se traducen en una reducción de problemas de selección adversa en mercados de capitales, así como en una disminución de los riesgos estimados.

En el año 2012 México se sumó a los más de 120 países que emiten sus reportes financieros bajo las Normas Internacionales de Información Financiera (IFRS, por sus siglas en inglés) (Joshi *et. al*, 2012). La obligatoriedad de implementar las IFRS en México fue anunciada por la Comisión Nacional Bancaria y de Valores en el año 2009 y es de aplicación general para las empresas que cotizan en la Bolsa Mexicana de Valores, excepto aquéllas que pertenezcan al sector financiero.

Las IFRS tienen implicaciones económicas importantes en las cifras contables de las empresas, así como en sus indicadores financieros; y principalmente en las revelaciones que deben hacer las entidades en las notas de sus reportes financieros. Uno de los principales impactos de las IFRS es el relacionado con la medición, presentación y divulgación de los Instrumentos Financieros. Con respecto a este rubro, está intrínsecamente relacionada la Administración de Riesgos que requieren evaluar y difundir las entidades. La alineación que tienen las IFRS con sus contrapartes regulatorias, tales como Basilea, se ha incrementado en mayor medida a través de las normas relativas a Instrumentos Financieros; principalmente, con la norma IFRS 7 que trata la revelación de dichos instrumentos, con un énfasis particular en lo que corresponde a la administración de riesgos.

El objetivo de este capítulo es mostrar la importancia de la normatividad contable internacional en la generación de confiabilidad en los reportes financieros, al regular la medición y gestión de los riesgos financieros en una organización. El capítulo presenta todas las normas internacionales relacionadas con los Instrumentos Financieros, sus objetivos, alcance y vigencia; así como una descripción de la parte en la que cada una de ellas trata el tema de riesgos. Finalmente, también se muestra la relación de la importante labor que realizan dos organismos regulatorios: el IASB (International Accounting Standards Board) y el BIS (Bank for International Settlements).

7.2 Normatividad contable para la administración de riesgos

Las Normas Internacionales de Información Financiera (IFRS) que contienen parámetros para revelar en las notas de los Estados Financieros, aspectos relacionados con la administración de riesgos de una entidad, son las siguientes:

Objetivo	Alcance
IAS 32 - Instrumentos Financieros: Presentación. Vigente desde el 1 de enero de 2005.	
Establecer los principios para presentar los instrumentos financieros (IF) como pasivos o patrimonio y para compensar activos y pasivos financieros.	Esta norma se aplica a todas las entidades y a toda clase de instrumentos financieros (IF), excepto a: <ul style="list-style-type: none"> - Participaciones en subsidiarias, asociadas y negocios conjuntos. - Planes de beneficios a empleados. - Contratos de seguro. - IF, contratos y obligaciones derivados de pagos basados en acciones.
IAS 39 - Instrumentos Financieros: Reconocimiento y medición. Vigente desde el 1 de enero de 2005.	
Establecer las reglas para el reconocimiento y la valuación de los pasivos financieros, así como de algunos contratos de compra o venta de partidas no financieras.	Esta norma tiene el mismo alcance de la IAS 32. Adicionalmente, se aplica a toda clase de instrumentos financieros, excepto a: <ul style="list-style-type: none"> - Derechos y obligaciones surgidos de contratos de arrendamiento. - IF con características de instrumento de patrimonio. - Contratos a término entre un adquirente y un accionista que darán lugar a una combinación de negocios en una fecha de adquisición futura. - Compromisos de préstamo diferentes a los que la entidad designe como pasivos financieros al valor razonable, o que puedan liquidarse por el neto, en efectivo o emitiendo otro IF.
IFRS 7 - Instrumentos Financieros: Información a revelar. Vigente desde el 1 de enero de 2007.	
Requerir a las entidades que en sus estados	Esta norma tiene el mismo alcance de las

<p>financieros, revelen información que permita a los usuarios evaluar:</p> <ul style="list-style-type: none"> - La importancia de los instrumentos financieros en la situación financiera y en el rendimiento de la entidad; y - La naturaleza y grado de los riesgos derivados de los instrumentos financieros a los que está expuesta la entidad y cómo administra la entidad esos riesgos. 	<p>normas IAS 32 y 39.</p>
<p>IFRS 9 - Instrumentos financieros. Vigente a partir del 1 de enero de 2013.</p>	
<p>Establecer las reglas para la información financiera sobre activos financieros, que permitan presentar información útil y relevante para los usuarios de los estados financieros, para la evaluación de los importes, calendario e incertidumbre de los flujos de efectivo futuros de la entidad.</p>	<p>Esta norma es aplicable a todos los activos contemplados en el alcance de la norma IAS 39.</p>
<p><i>IAS</i>: International Accounting Standards. También se le conoce por sus siglas en español como NIC (Norma Internacional de Contabilidad). <i>IFRS</i>: International Financial Reporting Standards. También se le conoce por sus siglas en español como NIIF (Norma Internacional de Información Financiera).</p>	

Por otro lado, también es importante mencionar las Normas de Información Financiera (NIF, normas contables mexicanas) que dictan las reglas referentes a los Instrumentos Financieros:

Boletín C-2: Instrumentos Financieros. Vigente desde enero de 2001.

Documento de adecuaciones al Boletín C-2. Vigente desde enero de 2005.

Boletín C-10: Instrumentos financieros derivados y operaciones de cobertura. Vigente desde enero de 2005.

Boletín C-12: Instrumentos financieros con características de pasivo, de capital o de ambos. Vigente desde enero de 2004.

Interpretación a las Normas de Información Financiera (INIF) 5: Reconocimiento de la contraprestación adicional pactada al inicio del instrumento financiero derivado para ajustarlo a su valor razonable. Vigente desde enero de 2008.

INIF 6: Oportunidad en la designación formal de la cobertura. Vigente desde enero de 2008.

INIF 7: Aplicación de la utilidad o pérdida integral generada por una cobertura de flujo de efectivo sobre una transacción pronosticada de compra de un activo no financiero. Vigente desde enero de 2008.

INIF 16: Transferencia de categoría de instrumentos financieros primarios con fines de negociación. Vigente desde enero de 2009.

Orientación a las Normas de Información Financiera (ONIF) 2: Recopilación de normas de presentación y de revelación sobre instrumentos financieros. Vigente desde enero de 2009.

De estas normas, las que contienen definiciones de riesgos, así como reglas para revelar en los reportes financieros, aspectos relacionados con la administración de los mismos, son el Boletín C-2 y C-10, así como la ONIF 2. Actualmente el Consejo Mexicano de Normas de Información Financiera (CINIF) se encuentra modificando estas normas y proponiendo la emisión de nuevas normas contables que se encuentren en convergencia con las normas internacionales.

A continuación se describe la parte de Administración de Riesgos que menciona cada una de las normas internacionales (IFRS).

7.2.1 IAS (NIC) 32 - Instrumentos Financieros: Presentación

Esta norma señala en su párrafo 42, la posibilidad que tiene una entidad de compensar un activo y un pasivo financiero, de tal forma que presente en su Estado de Situación Financiera el importe neto, siempre y cuando cumpla dos condiciones:

- Mantenga en el momento actual, el derecho, exigible legalmente, de compensar los importes reconocidos.
- Mantenga la intención de liquidar por el importe neto, o de realizar el activo y cancelar el pasivo simultáneamente.

El derecho de compensación se trata de una prerrogativa legal del deudor, adquirida a través de un contrato u otro medio distinto, para cancelar o eliminar, total o parcialmente, el importe de una cantidad debida al acreedor, aplicando con ese importe otro que el acreedor adeuda.

El párrafo 46, señala que al existir el derecho efectivo a compensar un activo financiero y un pasivo financiero, se afecta al conjunto de derechos y obligaciones asociados con los activos y los pasivos financieros correspondientes, pero también puede afectar el nivel de exposición de la entidad a los riesgos de crédito y de liquidez.

En este sentido, la norma menciona que ante la intención de ejercer el derecho y liquidar simultáneamente las dos posiciones, la presentación del activo y del pasivo en términos netos, refleja más adecuadamente los importes y el calendario de los flujos de efectivo esperados en el futuro, así como los riesgos a que están sujetos tales flujos.

Se puede presentar la situación en la cual una entidad tenga el derecho de compensar pero no la intención de liquidar en términos netos, o de realizar el activo y liquidar el pasivo de forma simultánea. En este caso se debe revelar el efecto del derecho sobre la exposición de la entidad al riesgo de crédito siguiendo las reglas de la IFRS 7.

Se considera que la realización de un activo financiero es simultánea con la liquidación de un pasivo financiero, sólo cuando las dos transacciones ocurren en el mismo momento. Bajo esta premisa, si dos instrumentos financieros se liquidan simultáneamente, los flujos de efectivo son equivalentes a una única cantidad neta, por lo que no existe exposición al riesgo de crédito o de liquidez. Sin embargo, si una entidad liquida dos instrumentos mediante cobros y pagos independientes, debe revelar su exposición al riesgo de crédito por el importe total del activo, o su exposición al riesgo de liquidez por el importe total del pasivo (párrafo 48).

7.2.2 IAS (NIC) 39 - Instrumentos Financieros: Reconocimiento y medición

Esta norma dedicada básicamente a aspectos de valuación (medición) de los instrumentos financieros, regula una de las principales transacciones que se realiza con tales instrumentos: la transferencia de un activo financiero.

El párrafo 18 de la norma estipula que una entidad transfiere un activo financiero cuando:

- Se han transferido los derechos contractuales a recibir los flujos de efectivo de un activo financiero; o
- Se retienen los derechos contractuales a recibir los flujos de efectivo del activo financiero, pero se asume la obligación contractual de pagarlos a uno o más perceptores, bajo ciertas condiciones.

La relación de este tipo de transacción con la administración de riesgos, surge porque cuando una entidad transfiere un activo financiero, debe evaluar en qué medida retiene los riesgos y los beneficios inherentes a la propiedad de dicho activo.

La transferencia de riesgos y ventajas se debe evaluar comparando la exposición de la entidad, antes y después de la transferencia, a la variación en los importes y en el calendario de los flujos de efectivo netos del activo transferido.

La retención de manera sustancial de todos los riesgos y ventajas inherentes a la propiedad de un activo financiero, se da cuando su exposición a la variación en el valor presente de los flujos de efectivo futuros netos de dicho activo, no varía significativamente como resultado de la transferencia.

Por su parte, la transferencia de manera sustancial de los riesgos y ventajas inherentes a la propiedad de un activo financiero, se presenta si su exposición a tal variabilidad deja de ser significativa en relación con la variabilidad total del valor presente de los flujos de efectivo futuros netos asociados con el activo financiero

.

7.2.3 IFRS (NIIF) 7 - Instrumentos Financieros: Información a revelar

El tema de Administración de Riesgos, se trata de un asunto básicamente de revelación en la normatividad contable internacional. La norma IFRS 7 reglamenta con detalle todo lo que debe una entidad revelar en las notas de sus reportes financieros con respecto a su gestión de riesgos. La Figura n.1 muestra el resumen de revelaciones tanto cualitativas como cuantitativas que se deben revelar acerca de los riesgos financieros de una entidad.

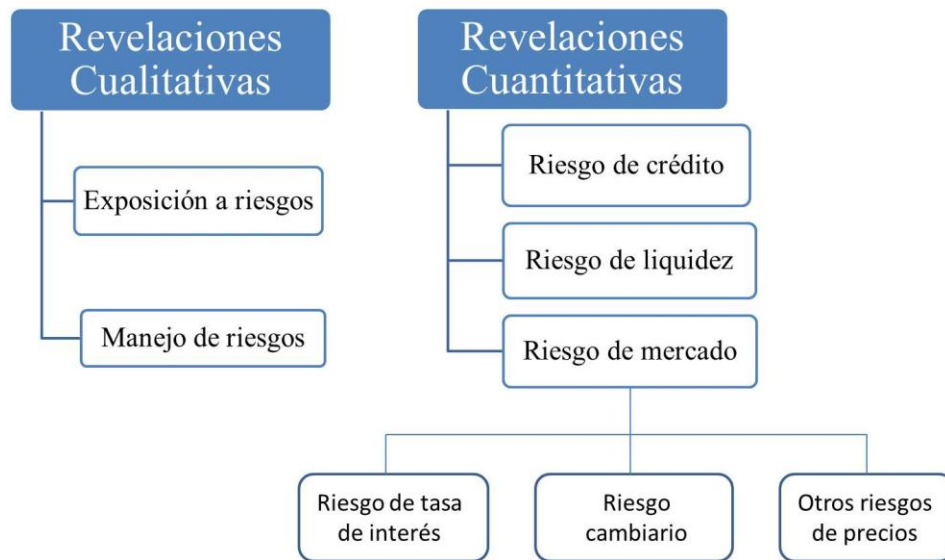


Figura 7.1

Fuente: IFRS 7

La información cualitativa que se debe presentar para cada tipo de riesgo que surja de los instrumentos financieros, debe contener: las exposiciones al riesgo y la forma en que éstas surgen; los objetivos, políticas y procesos para la administración del riesgo, así como los métodos utilizados para medirlo. Adicionalmente, se debe mencionar cualquier cambio que de lo anterior se haya suscitado desde el periodo anterior al que se esté reportando (IFRS 7, párrafo 33).

En cuanto a la información cuantitativa, para cada tipo de riesgo que emane de los instrumentos financieros, se debe declarar (IFRS 7, párrafo 34):

- Datos cuantitativos resumidos acerca de su exposición al riesgo al final del periodo sobre el que se informa.
- La información a revelar requerida para cada tipo de riesgo en particular.
- Las concentraciones de riesgo, si no resultan aparentes de los puntos anteriores.

7.2.3.1 Riesgo de Crédito

En cuanto al riesgo de crédito, la principal revelación que se debe presentar en los reportes financieros, es el importe que mejor represente el máximo nivel de exposición al riesgo de crédito

al final del periodo sobre el que se está informando. Dicho importe también debe ir acompañado de una descripción de las garantías tomadas y de cualquier otra mejora crediticia (por ejemplo avales).

Por otro lado, para cada tipo de activo financiero, se debe mostrar un análisis de la antigüedad de los activos financieros que al final del periodo sobre el que se esté informando, estén en mora, más no deteriorados. En adición, si al realizar un análisis individual de los activos financieros, se determina que al final del periodo algunos de ellos están deteriorados, se deberá hacer una descripción de los mismos que incluya los factores considerados para medir el deterioro (IFRS 7, párrafo 37).

La evidencia de que un activo o grupo de activos financieros están deteriorados, la genera eventos que causan una pérdida. Algunos ejemplos de tales eventos son (IAS 39, párrafo 59):

- Dificultades financieras significativas del emisor o del obligado.
- Infracciones de las cláusulas contractuales, tales como incumplimientos o moras en el pago de los intereses o el principal.
- Probabilidad que el prestatario entre en quiebra o en otra forma de reorganización financiera.
- Desaparición de un mercado activo para el activo financiero en cuestión, debido a dificultades financieras.

7.2.3.2 Riesgo de Liquidez

La revelación requerida en la normatividad contable consiste en mostrar un análisis de los vencimientos de los pasivos financieros que exhiba los plazos contractuales de vencimiento restantes, así como una descripción de cómo la entidad administra el riesgo de liquidez (IFRS 7, párrafo 39).

7.2.3.3 Riesgo de Mercado

Como lo muestra la Figura 7.2, la principal revelación en cuanto al riesgo de mercado, es la que corresponde al análisis de sensibilidad para cada tipo de riesgo de mercado al que la entidad esté expuesta al final del periodo sobre el que se informa.



Figura 7.2

Fuente: IFRS 7

Una entidad debe mostrar las repercusiones en el resultado del periodo y el capital contable originadas por cambios en la variable relevante de riesgo (razonablemente posibles). Además se deben dar a conocer los métodos e hipótesis que se hayan aplicado para elaborar el análisis de sensibilidad, así como los cambios que se hayan realizado en tales métodos e hipótesis con respecto al periodo anterior al que se esté informando (IFRS 7, párrafo 40).

La norma también señala que en vez del análisis de sensibilidad antes mencionado, una entidad puede utilizar como dicho análisis, el Valor en Riesgo, si éste manifiesta las interdependencias entre las variables de riesgo, tales como las tasas de interés y de cambio, y lo utilizase para administrar sus riesgos financieros (IFRS 7, párrafo 41).

La Figura 7.3 muestra un ejemplo de qué áreas de una entidad pondrían intervenir en el proceso de obtención de datos y elaboración del análisis de sensibilidad, así como la función general que cada una tendría en dicho proceso.

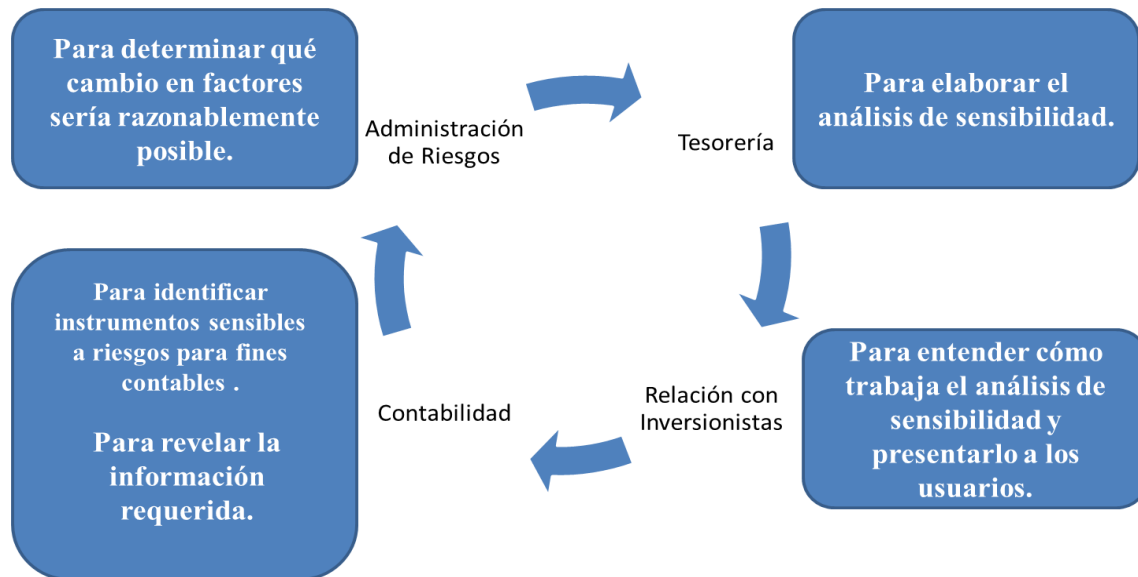


Figura 7.3

Fuente: Deloitte (2012-A)

7.2.4 IFRS (NIIF) 9 - Instrumentos financieros

En el Apéndice B (guía de aplicación) de esta norma, se estipula que en algunos tipos de transacciones, una entidad puede priorizar pagos a los tenedores de activos financieros haciendo uso de varios instrumentos vinculados contractualmente que generan concentraciones de riesgos de crédito (también llamados tramos). Cada tramo tiene una clasificación de prioridad que especifica el orden en el que los flujos de efectivo generados por el emisor se asignan al mismo. Bajo estas circunstancias, los tenedores de un determinado tramo tienen el derecho a pagos del principal e intereses sobre el importe del principal pendiente, solo si el emisor genera suficientes flujos de efectivo para satisfacer pagos a los tramos clasificados como de prioridad más alta (párrafo B4.20).

7.2.5 Análisis comparativo entre IFRS y NIF

Un rasgo característico de la normatividad contable internacional (IFRS), es el grado de detalle que solicita a las entidades revelar en sus estados financieros, en comparación con cualquier otra normatividad de algún país. La Tabla n.1 muestra un análisis comparativo de lo que solicitan las normas mexicanas (NIF) en relación con información relativa a riesgos, y lo que solicitan las

IFRS. Como se puede observar, todos los parámetros analizados son requeridos por la normatividad internacional, principalmente por la IFRS 7.

Información a revelar	C-2	C-10	ONIF	CNBV	IFRS
Exposición a riesgo de mercado.		✓	✓		✓
Exposición a riesgo de liquidez (incluyendo fuentes internas y externas de liquidez).		✓	✓	✓	✓
Exposición a riesgo de crédito, criterio y bases de reconocimiento de quebranto esperado por el deterioro en capacidad de pago del emisor.	✓	✓	✓		✓
Política de administración de riesgos.	✓		✓		✓
Características de instrumentos financieros y políticas de uso.	✓	✓		✓	✓
Técnicas de valuación, métodos y supuestos (incluyendo medición de efectividad).	✓	✓	✓	✓	✓
Cambios en exposición a riesgos y en su administración (incluyendo impacto en resultados y flujos de instrumentos derivados así como número y monto de posiciones cerradas en el periodo y llamadas de margen).				✓	✓
Análisis de sensibilidad (incluyendo identificación de riesgos, 3 escenarios: probables, posibles - variación 25% en precio de subyacente - y remotos - variación 50% en precio subyacente, y su impacto en resultados y flujos).				✓	✓
Información cuantitativa: tipo de derivado, designación, notional, valor del subyacente, valor razonable, vencimientos, colaterales, garantías, líneas de crédito etc.		✓	✓	✓	✓
Políticas contables y efectos en información financiera (cobertura, impacto en resultados, capital, etc.).	✓	✓	✓		✓

Tabla 7.1

Fuente: IASB (2011), Deloitte (2012-A)

7.3 IFRS y Basilea III

Si bien la normatividad contable internacional es de aplicación general para todas las entidades, las normas relativas con los instrumentos financieros tienen un sesgo particular hacia el sector financiero, en especial el bancario. Por tal motivo en esta sección se analizará la relación que guardan ambos cuerpos normativos, teniendo en consideración que las IFRS y Basilea III son dos enfoques acerca de un mismo tema de importancia análoga, aunque con objetivos distintos (contable y preventivo, respectivamente).

Como consecuencia de la crisis financiera internacional del año 2008, el órgano de gobierno del Comité de Basilea dio a conocer en septiembre de 2010, los nuevos requerimientos internacionales de Basilea III para la banca comercial. Dentro de los aspectos más significativos de esta reforma, se encuentran (BIS, 2010):

- Mejorar la calidad de capital.
- Elevar significativamente el nivel de capital.
- Aumento de ponderadores para activos sujetos a riesgo complejos.
- Mejora en las características de las obligaciones subordinadas para absorber pérdidas.
- Introducir estándares de liquidez, a través de un ratio de cobertura a corto plazo (2015) y de liquidez estructural a largo plazo (2018) para lograr una estructura de balance que no descansa excesivamente en financiación a corto plazo.

Para alcanzar lo anterior, en México se ha determinado que a partir de septiembre de 2012 se adopten anticipadamente algunas de las medidas de Basilea III, implementando inicialmente los siguientes cambios (CNBV, 2012; Deloitte, 2012-B):

- Nueva definición e integración del capital.
- Redefinición del marco de alertas tempranas.
- Criterios para la inclusión de obligaciones subordinadas en el capital.

Todos estos aspectos tienen una implicación relevante en la contabilidad. Si bien desde hace varios años los criterios de Basilea y del IASB están en mayor compatibilidad debido a que los organismos supervisores han sugerido que los cambios en las IFRS no afecten los criterios preventivos de capital, sean consistentes con sus estándares y no dupliquen requerimientos; aún se requiere mayor simetría entre ambas regulaciones.

De acuerdo con Espiga (2011), una de las principales necesidades de convergencia de IFRS con Basilea III, se encuentra en el tema de Provisiones, es decir, en el reconocimiento de la pérdida esperada de una cartera. En este sentido, se requiere concordancia en las siguientes definiciones:

- Definición de cartera buena y cartera mala
- Definición de pérdida esperada a lo largo de toda la vida de la cartera comercial y el registro contable de la provisión de dicha pérdida.

En gran parte, las IFRS son concebidas para dictar normas de revelación de información. En este sentido, para efectos de lo que ha requerido Basilea en cuanto a divulgación se refiere, las IFRS integran un buen complemento para la asimilación de la condición de riesgos que enfrentan las entidades bancarias.

7.4 Conclusiones

La creciente aceptación de las Normas Internacionales de Información Financiera (IFRS) alrededor del mundo, ha generado implicaciones de distinta índole. Por una parte, la evidencia empírica en algunos países muestra un cambio en la valuación de las acciones a partir de la adopción de dichas normas. Por otro lado, las implicaciones económicas en las entidades también han tenido una relevancia considerable. Los rubros tales como patrimonio, resultado integral y otros conceptos más que involucran la valuación a través del valor razonable (*fair value*), son algunos de los que mayor afectación han causado a las entidades, positiva y negativamente.

Al encontrarse México en un momento histórico en términos de regulación contable, al añadirse a los más de 120 países que han adoptado las IFRS para la elaboración de sus estados financieros, resulta interesante conocer cuáles son los lineamientos que estas normas señalan en el área de administración de riesgos. Es la norma IFRS 7 la que principalmente dicta los preceptos a seguir para revelar/divulgar aspectos cualitativos y cuantitativos de los distintos riesgos financieros.

Ante la necesidad de homologar regulaciones, la creciente aceptación internacional de recomendaciones del Comité de Supervisión Bancaria de Basilea (BCBS por sus siglas en inglés) en materia prudencial o preventiva, y del Consejo Internacional de Normas de Contabilidad (IASB), en materia contable, a través de la emisión de las IFRS; permite abatir fronteras en materia de comparabilidad y usar un lenguaje financiero en común.

7.5 Bibliografía

- Bank for International Settlements (BIS). (2010). *Basilea III: Marco regulador global para reforzar los bancos y sistemas bancarios*. Comité de Supervisión Bancaria de Basilea. Recuperado el 31 de octubre de 2012 de: http://www.bis.org/publ/bcbs189_es.pdf
- Comisión Nacional Bancaria y de Valores (CNBV). (2012). *Implementación de Basilea III en México*. Conferencia dictada el 21 de agosto, 2012. Recuperado el 31 de octubre de 2012 de: [www.cnbv.gob.mx/.../Basilea%203%20en%20México_presentacion....](http://www.cnbv.gob.mx/.../Basilea%203%20en%20México_presentacion...)

Consejo Mexicano de Normas de Información Financiera. (2012). *Normas de Información Financiera* (NIF). IMCP, México.

Daske, H., Hail, L., Leuz, C., Verdi, R. (2008). Mandatory IFRS Reporting around the World: Early Evidence on the Economic Consequences. *Journal of Accounting Research*, Vol. 46, No. 5.

Deloitte (2012-A). Material Diplomado IFRS. Tecnológico de Monterrey, Ciudad de México.

Deloitte (2012-B). Basilea III en México. Financial Daily. Mayo 2012. Recuperado el 31 de octubre de 2012 de: <http://www.deloitte.com/assets/Dcom-Mexico/Local%20Assets/Documents/mx>

Espiga, D. (2011). *Basilea III, las nuevas reglas*. X Jornada Anual de Riesgos. Club de Gestión de Riesgos de España. Conferencia dictada el 24 de febrero de 2011.

International Accounting Standard Board (IASB). (2011). *Normas Internacionales de Información Financiera*. Versión en español consultada en: www.ifrs.org

Joshi, S., Takara, R., Gremone, L., González, B. (2012). *América Latina está en el camino de las IFRS ¿Seguirá con la misma velocidad o la reducirá?* Standard & Poor's.

Capítulo 8

Aplicación de los modelos de valuación de bonos con incumplimiento a dos empresas mexicanas.

José Benito Díaz Hernández

Departamento de Matemáticas. UAM-Iztapalapa

Patricia Saavedra Barrera

Departamento de Matemáticas. UAM-Iztapalapa

Resumen

En este trabajo se presenta un estudio sobre la estimación del riesgo de incumplimiento que corre una empresa al emitir deuda a corto plazo, a través de los modelos estocásticos de tipo estructural, como el de Merton y Black-Cox, y los modelos de intensidad.

8.1 Introducción

El mercado de bonos corporativos en México no está tan desarrollado como el mercado accionario o la compra de deuda soberana. Durante el 2011 se invirtieron 225,000 millones de pesos en deuda corporativa, mientras que en 2012 se habían invertido 116,000 millones de pesos hasta fines de agosto. Posiblemente, algunas de las razones por las cuales no se invierte más en

este mercado son por un lado, la falta de oferta, la deuda se concentra en diez emisores, y por otro a que las metodologías tradicionales no nos permiten estimar en forma dinámica el riesgo de incumplimiento que conlleva la deuda emitida por las empresas. La falta de oferta se explica porque sólo las empresas que cotizan en bolsa están obligadas a informar en forma sistemática sobre sus estados financieros.

En cuanto a la metodología, la forma tradicional de estimar si un bono es buena inversión es buscar su valor en el mercado y comparar el *yield* con la tasa de Cetes y de esta manera estimar el *spread* o diferencial de rendimiento. Mientras más alto sea el diferencial, mayor riesgo conlleva el bono. Para calificar la deuda a corto y largo plazo, se toman en cuenta el análisis de resultados, las ventas, el EBITDA¹ y la utilidad neta de la empresa para determinar su nivel de apalancamiento y de esa forma decidir si puede o no hacer frente a sus obligaciones. Este es un esquema estático que se usa para inferir el riesgo de incumplimiento de la empresa es sus pagos futuros. Otra posibilidad es utilizar directamente la calificación anual que otorgan las agencias calificadoras estadounidenses a las empresas que cotizan en bolsa. Este mecanismo no ha dejado de ser puesto en duda desde la crisis del 2008, al comprobar que estas agencias están lejos de ser objetivas en el proceso de calificación².

El objetivo de este trabajo es aplicar los modelos dinámicos estocásticos de tipo estructural y de intensidad para analizar la capacidad que tiene una empresa al inicio del año para hacer frente a las obligaciones contraídas al emitir deuda a corto plazo, menor a un año. Para ello, se representará toda la deuda a corto plazo, como un bono cupón cero con fecha de vencimiento a un año. En este contexto definimos riesgo de incumplimiento como la posibilidad de no recibir el importe íntegro prometido o de que se produzca un retraso en el pago del mismo.

El planteamiento básico de los modelos estructurales fue desarrollado por Merton (1974), y se basa en el modelo de Black-Scholes (1973). Se asume que los activos de una empresa consisten, para todo tiempo, en la suma del valor de sus acciones más el valor de su deuda. De esta forma, el valor de las acciones de una empresa puede tratarse como una posición larga en una opción de compra cuyo subyacente es el valor de sus activos y el precio de ejercicio, es el valor nominal de la deuda al vencimiento. De esta forma, aplicando el modelo de Black-Scholes

¹ Earnings before interest, taxes, depreciation & amortization.

² Periódico El País 21 de octubre del 2012.

(1973), es posible valorar las acciones y la deuda al vencimiento y estimar la probabilidad de incumplimiento. Sin embargo, la simplicidad de este razonamiento no permite describir la situación real de las empresas y de los mercados financieros al suponer una tasa de interés libre de riesgo constante y que toda la estructura de la deuda de la empresa esté representada por un único bono cupón cero con vencimiento T , lo que implica que la insolvencia sólo puede producirse al tiempo T . Con el fin de superar las limitaciones del modelo anterior, Black y Cox (1975), proponen un modelo en el que el emisor puede incumplir sus obligaciones en cualquier momento durante la vigencia del bono, introduciendo una barrera para los activos de la empresa. En el momento que el valor de los activos cruza la barrera, se produce el incumplimiento. Posteriormente, Brennan y Schwartz modifican el modelo de Black y Cox en 1980 al considerar la tasa de interés estocástica. Finalmente, Zhou (1997) introduce un modelo con saltos con el fin de que se produzcan incumplimientos impredecibles.

Ante las imposibilidad de reproducir el diferencial de rendimiento en los bonos corporativos, surgió a mediados de la década de los noventa los modelos de intensidad o de forma reducida. Estos modelos utilizan los modelos de tasa de interés estocástico para valorar el bono corporativo al considerar que la tasa de descuento consiste de dos tasas: la tasa de referencia y otra que describe el comportamiento del diferencial de rendimiento. De esta forma, la insolvencia del emisor se produce como el primer tiempo de paro de un proceso estocástico que modela el incumplimiento. Jarrow y Turnbull en 1995 son de los primeros en utilizar estos modelos con el supuesto de que las probabilidades de incumplimiento son constantes. Lando en 1997 amplía el anterior modelo permitiendo que haya correlación entre la tasa de interés y el proceso que describe la insolvencia. Estos modelos también han recibido críticas debido a que la probabilidad de incumplimiento se define en forma exógena, sin tomar en cuenta los estados financieros de las empresas. En años recientes, han aparecido en la literatura métodos híbridos que buscan combinar las ventajas de ambos modelos, Lando (2004).

La estructura de este artículo es la siguiente: en la segunda sección se presentan los modelos estructurales, en la tercera los de intensidad y en la cuarta su aplicación en la estimación del riesgo de incumplimiento de la deuda de dos empresas mexicanas: Bimbo y Cemex. Por último, en la quinta sección se presentan algunas conclusiones.

8.2 Modelos estructurales

El primer supuesto del modelo de Black-Scholes-Merton es acerca del proceso estocástico para el precio subyacente del contrato. En el artículo original de Black & Scholes (1973) se asume que el activo con riesgo no paga dividendos, la tasa de interés libre de riesgo es una constante conocida y la dinámica de los precios sigue un Movimiento Browniano Geométrico (MBG).

Considerando una empresa que posee activos riesgosos con valor V que se financian a través de la emisión de acciones con valor S y deuda con valor de mercado D ; y asumiendo que la deuda de la empresa se representa como un único bono cupón cero con fecha de vencimiento T y valor nominal $F > 0$, se tiene que para todo $t \leq T$,

$$V_t = S_t + D_t,$$

es decir, el valor de la empresa es igual a la suma de sus obligaciones, ver Lando (2004). En este modelo, los acreedores no pueden forzar a la empresa a declararse en quiebra antes del tiempo de madurez T . La empresa incumple si, al tiempo T el valor de la empresa es menor que su deuda. En caso de incumplimiento el pago al tenedor del bono al tiempo de vencimiento es el $\min(V_T, F)$ que es el valor de mercado de la deuda de la empresa al tiempo T .

Al suponer que V_t , el valor de los activos con riesgo, sigue un movimiento browniano geométrico³ y bajo los supuestos del modelo de Black & Scholes, los valores del bono riesgoso D y de las acciones S al tiempo $t = 0$ están dados por:

$$\begin{aligned} D_0 &= F e^{-rT} \varphi(d_2) - V_0 \varphi(d_1), \\ S_0 &= V_0 \varphi(d_1) - F e^{-rT} \varphi(d_2), \end{aligned}$$

con

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy,$$

³ V_t sigue un movimiento browniano si $dV_t = rV_t dt + \sigma_V V_t dW_t$, donde $(W_t), t \in [0, \infty)$, es un movimiento browniano estándar bajo la medida neutral al riesgo.

$$d_1 = \frac{\ln\left(\frac{V_0}{F}\right) + \left(r + \frac{\sigma_V^2}{2}\right)T}{\sigma_V\sqrt{T}},$$

$$d_2 = \frac{\ln\left(\frac{V_0}{F}\right) + \left(r - \frac{\sigma_V^2}{2}\right)T}{\sigma_V\sqrt{T}}.$$

Así mismo, la probabilidad de incumplimiento de una empresa al tiempo T puede ser caracterizada como $\varphi(-d_2)$, donde d_2 esta dada anteriormente. Sin embargo, para calcular esta probabilidad se necesita conocer el valor de mercado de todos los activos de la empresa así como la volatilidad de los mismos; desafortunadamente estos valores no son observables directamente en ninguno de los reportes que emiten las empresas y tampoco se pueden desprender de alguno de los indicadores de mercado. Pero si la empresa que se está analizando cotiza en algún mercado, se puede obtener tanto el valor de mercado del capital accionario, multiplicando el número de acciones en circulación por el precio de mercado de éstas, así como la volatilidad del rendimiento del precio de las acciones. Se puede demostrar que existe una relación directa entre la volatilidad de las acciones σ_S y la de los activos σ_V , dada por:

$$\sigma_S = \frac{V_t}{S_t} \varphi(d_1) \sigma_V.$$

Con este resultado y con el valor del bono riesgoso, D_0 , al tiempo $t = 0$ se tiene un sistema no lineal, por medio del cual se puede obtener el valor nominal del bono, F , y el valor de la volatilidad de los activos con riesgo, σ_V .

Una vez calculadas estas variables, definiendo $B(t, T)$ el precio al tiempo t de un bono descontado, es decir, un bono cupón cero que paga una unidad monetaria al tiempo T con tasa continua $R(t, T)$, se tiene que el diferencial (spread) de rentabilidad o lo que es lo mismo, la diferencia entre el rendimiento a vencimiento de un instrumento riesgoso y una tasa libre de riesgo bajo las mismas condiciones, se obtiene por

$$Spread = R(t, T) - r = -\frac{1}{T} \ln\left(\varphi(d_2) - \frac{V_0 \varphi(d_1)}{F e^{-rT}}\right).$$

Como se puede observar de las fórmulas anteriores, bajo este modelo, el spread de crédito únicamente depende de la razón de la deuda al valor de mercado de la empresa, de la madurez del instrumento de deuda y de la volatilidad de los activos.

8.2.1 Modelo de Merton con tasa de interés estocástica

Un primer trabajo que aparece en el estudio de la tasa corta es el modelo de Merton (1973), quien es el primero en proponer la modelación de tasas de interés por medio de procesos estocásticos. Los modelos de tasa de interés de un sólo factor más utilizados para modelar la tasa spot son el modelo de Vasicek y el modelo Cox, Ingersoll, Ross (CIR). El proceso de estimación de parámetros de un modelo estocástico permite conocer características adicionales que la estructura por sí misma no es capaz de captar; entre las técnicas de estimación de parámetros de modelos de tasa de interés se encuentran: el método de momentos, mínimos cuadrados ordinarios y máxima verosimilitud, ver Ball, C (1996).

8.2.1.1 Modelo de Merton bajo el modelo Vasicek

Asumiendo que la tasa de interés sigue el modelo de Vasicek⁴, el problema de los precios en el modelo de Merton con un sólo instrumento de deuda es simplemente una aplicación del cambio de numeraria, ver Lando (2004) y Björk (2004). Bajo la medida martingala, P^* , la dinámica del valor de los activos de la empresa y la tasa corta están dadas por:

$$\begin{aligned}dV_t &= r_t V_t dt + \sigma_V V_t (\rho dW_t^1 + \sqrt{1 - \rho^2} dW_t^2), \\dr_t &= k(\theta - r_t) dt + \sigma_r dW_t^2,\end{aligned}$$

donde W_t^1 y W_t^2 son movimientos brownianos independientes y σ_r la volatilidad de la tasa corta. Por lo tanto, el valor del bono riesgoso D y de las acciones S al tiempo $t = 0$ están dados por:

$$\begin{aligned}S_0 &= V_0 \varphi(d_1) - FB(0, T) \varphi(d_2), \\D_0 &= V_0 - S_0,\end{aligned}$$

con

⁴ El precio del bono cupón cero bajo el modelo Vasicek está dado por $B(t, r_t, T) = e^{a(t, T) - r_t b(t, T)}$, donde $b(t, T) = \frac{1 - e^{-k(T-t)}}{k}$ y $a(t, T) = \frac{1}{k^2} (b(t, T) - T + t) \left(k^2 \theta - \frac{1}{2} \sigma_r^2 \right) - \frac{\sigma_r^2 b^2(t, T)}{4k}$.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy,$$

$$d_1 = \frac{\ln\left(\frac{V_0}{FB(0,T)}\right) + \frac{1}{2}(\Sigma_{V,T}(T))^2}{\sqrt{(\Sigma_{V,T}(T))^2}},$$

$$d_2 = d_1 - \sqrt{(\Sigma_{V,T}(T))^2},$$

$$(\Sigma_{V,T}(T))^2 = \int_0^T (2\rho\sigma_V\sigma_r b(T-t) + \sigma_r^2 b^2(T-t) + \sigma_V^2) dt.$$

Para evitar la dependencia entre la variable estocástica que se usa para descontar y la función de pago de la opción, supóngase que $\rho \equiv 0$. Entonces, la probabilidad de incumplimiento está dada por $\varphi(-d_2)$, donde d_2 esta dada anteriormente. Al igual que el modelo estructural de Merton, si las empresas que se están analizando cotizan en algún mercado, se puede calcular el valor de mercado del capital accionario, la volatilidad del rendimiento, así como la correlación entre los dos movimientos brownianos. En esta ocasión, la relación entre la volatilidad de los activos, σ_V , y la volatilidad de acciones, σ_S , esta dada por:

$$\sigma_S = \frac{V_t \varphi(d_1) \sigma_V \sqrt{1 - \rho^2} + \varphi(d_2) \sigma_r}{S_t \sqrt{1 - \rho^2}}.$$

Con estos resultados, se tiene un sistema completo de ecuaciones no lineal, por medio del cual se obtienen los valores de F , σ_V y ρ .

8.2.1.2 Modelo de Merton bajo el modelo CIR

Cuando se asume que la tasa de interés es estocástica, la expresión para la valuación de una opción es en general difícil de obtener, excepto para los procesos gaussianos. Kim y Kunitomo (1999) desarrollaron una simple expresión para aproximar el valor de una opción cuando la tasa de interés sigue un proceso de Itô. Por su parte en Yong (2001) se presenta una adaptación en la aproximación para el caso en donde la tasa de interés no es gaussiana.

Al suponer los supuestos del modelo de Merton con tasa de interés estocástica bajo el modelo Vasicek, se tiene que el cambio del precio de bono $B(t, r_t, T)$ bajo la medida neutral al riesgo está dado por:

$$dB(t, r_t, T) = r_t B(t, r_t, T) dt - b(t, T) b(t, r_t, T) \sigma_r \sqrt{r_t} dW_t.$$

Al usar que bajo el modelo CIR⁵, el precio del bono cupón cero está dada por $B(t, r_t, T) = a(t, T) e^{-r_t b(t, T)}$ se tiene que

$$\begin{aligned} \frac{dB(t, r_t, T)}{B(t, r_t, T)} &= r_t dt - \sigma_r b(t, T) \sqrt{r_t} dW_t \\ &= \frac{1}{b(t, T)} \ln \left(\frac{a(t, T)}{B(t, T)} \right) dt - \sigma_r \sqrt{b(t, T) \ln \left(\frac{a(t, T)}{B(t, r_t, T)} \right)} dW_t. \end{aligned}$$

Observe que el porcentaje de volatilidad del precio del bono es una función no determinista, pues depende de los niveles actuales del precio del bono. Por lo tanto, para obtener el valor de mercado de la empresa S_t al tiempo t se calcula:

$$S_t = E^* \left[e^{-\int_t^T r_s ds} \max(V_T - F, 0) | \mathfrak{S}_t \right].$$

Se sugiere usar la aproximación propuesta por Yong (2001), la cual se muestra a continuación:

Proposición. Sea S el valor de las acciones de una empresa, supóngase que la tasa de interés a corto plazo tiene la siguiente dinámica $dr_t = k(\theta - r_t)dt + \sigma_r \sqrt{r_t} dW_t$, entonces el valor de las acciones de mercado S_t al tiempo $t = 0$ esta dado por:

$$\begin{aligned} S_0 &= \left[V_0 \varphi(d_1) - F e^{-\int_0^T r_t^* dt} \varphi(d_2) \right] + \\ &+ \sigma_r C_0 \left[V_0 \vartheta(d_1) - F e^{-\int_0^T r_t^* dt} \{ \vartheta(d_2) - \sigma_r \sqrt{T} \varphi(d_2) \} \right] \\ &+ \sigma_r C_1 \left[d_2 V_0 \vartheta(d_1) - d_1 F e^{-\int_0^T r_t^* dt} \vartheta(d_2) \right] + o(\delta), \end{aligned}$$

donde $\varphi(\cdot)$ es la función de distribución acumulada de una variable aleatoria normal estándar y $\vartheta(\cdot)$ es su densidad. Por su parte, r_t^* es una versión determinista de la tasa corta dada por:

$$r_t^* = r_0 e^{-kt} + \theta(1 - e^{-kt})$$

⁵ El precio del bono cupón cero bajo el modelo CIR está dado por $B(t, r_t, T) = a(t, T) e^{-r_t b(t, T)}$, donde $b(t, T) = \frac{2(e^{\gamma T} - 1)}{(\gamma + k)(e^{\gamma T} - 1) + 2\gamma}$, $a(t, T) = \left(\frac{2\gamma e^{(k+\gamma)\frac{T}{2}}}{(\gamma+k)(e^{\gamma T} - 1) + 2\gamma} \right)^{2k\theta/\sigma^2}$ y $\gamma = \sqrt{k^2 + 2\sigma^2}$.

y así la versión determinista del factor de descuento tiene la forma:

$$e^{-\int_t^T r_t^* dt} = e^{\frac{r_0 - \theta}{k}(1 - e^{-kT}) - \theta T}.$$

Las variables d_1 y d_2 están dadas por:

$$d_1 = \frac{1}{\sigma_V \sqrt{T}} \left[\ln \left(\frac{V_0}{F} \right) + \frac{r_0 - \theta}{k} (1 - e^{-kT}) + \theta T + \frac{\sigma_V^2}{2} T \right],$$

$$d_2 = d_1 - \sigma_V \sqrt{T}.$$

Por otro lado C_0 está representado por:

$$C_0 = \frac{1}{\sigma_V \sqrt{T}} \left[\frac{\varepsilon(r_0 - \theta)}{k} \left(\frac{1 - e^{-kT}}{k} - T e^{-kT} \right) + \frac{\varepsilon \theta T}{k} \left(1 - \frac{1 - e^{-kT}}{k} \right) \right].$$

Finalmente, $C_1 = -\frac{\rho}{\sigma_V T} C_{11}$, con

$$C_{11} = \frac{2\sqrt{\theta} \left((1 + 2e^{kT})\sqrt{r_0} - 3e^{\frac{kT}{2}}\sqrt{r_0 - \theta(1 - e^{kT})} \right) + (\theta(1 + 2e^{kT}) - r_0)\omega}{2e^{kT} k^2 \sqrt{\theta}}$$

y

$$\omega = \ln \left(\frac{\theta(2e^{kT} - 1) + r_0 + 2e^{\frac{kT}{2}} \sqrt{\theta^2(e^{kT} - 1) + \theta r_0}}{(\sqrt{r_0} + \sqrt{\theta})^2} \right).$$

Con esta proposición, se tiene un sistema completo de ecuaciones no lineal, por medio del cual se obtienen los valores de F , σ_V y ρ . Para poder analizar el diferencial de rentabilidad en estos modelos como función de los parámetros, se debe centrar el análisis en el efecto que tiene la tasa de interés cuando no existe correlación y el efecto de la correlación en determinados niveles de volatilidad.

8.2.2 Modelo de Black y Cox

Ahora se considera otra extensión del modelo de Merton implementada por Black y Cox en 1976. La idea principal es permitir que el incumplimiento ocurra antes de la maduración del bono. Es decir, el incumplimiento ocurrirá cuando el nivel del valor de los activos toque la cota inferior, modelado como función del tiempo. En general la cota de incumplimiento está dada por:

$$L_1(t) = L e^{-\alpha(T-t)}, \gamma \geq 0.$$

Al suponer que V_t sigue un movimiento browniano, la barrera es plana ($\alpha = 0$) y que el bono emitido por la empresa tiene un principal F , tal que $L < F$, se obtiene que el tiempo de incumplimiento τ está dado por el primer tiempo que un movimiento browniano con cierta tendencia alcanza la barrera:

$$\begin{aligned}\tau &= \inf\left\{0 \leq t \leq T: \ln V_0 + \left(r - \frac{1}{2}\sigma^2\right)t + \sigma W_t = \ln L - \alpha(T - t)\right\}, \\ &= \inf\left\{0 \leq t \leq T: \left(r - \frac{1}{2}\sigma^2 - \alpha\right)t + \sigma W_t = \ln L - \alpha(T) - \ln V_0\right\}.\end{aligned}$$

En el modelo de Black y Cox la función de pago de los tenedores del bono con madurez T esta dada por:

$$B(V_T, T) = \min(V_T, F)I_{\{\tau > T\}}, \quad (2.1)$$

cuya expresión coincide con la función de pago usual, cuando la cota no se alcanza en $[0, T]$. Denotando $B^m(V, T, F, L, \alpha)$ como el valor al tiempo $t = 0$ de la función de pago de un bono con madurez T , valor nominal F y cota de incumplimiento $L_1(t) = L$, se tiene que

$$B^m(V, T, F, L, \alpha) = E^*[e^{-rT} \min(V_T, F)I_{\{\tau > T\}}].$$

Si la barrera se toca antes de la maduración del bono, entonces la función de pago de los tenedores del bono tiene la forma:

$$B(V_\tau, \tau) = L_1(\tau)I_{\{\tau \leq T\}}$$

y así, el valor del bono cuando se alcanza la cota antes de la fecha de maduración es,

$$B^b(V, T, F, L, \alpha) = E^*[e^{-rT} L_1(\tau)I_{\{\tau \leq T\}}].$$

Haciendo una analogía con el modelo de Merton, la expresión de la función de pago del bono se expresa como la diferencia entre un bono libre de riesgo y put europeo, notemos que (2.1) se puede reescribir como:

$$\begin{aligned}B(V_T, T) &= \min(V_T, F)I_{\{\tau > T\}} = FI_{\{\tau > T\}} - \max(F - V_T, 0)I_{\{\tau > T\}} \\ &= V_T I_{\{\tau > T\}} - \max(V_T - F, 0)I_{\{\tau > T\}}.\end{aligned}$$

Es un contrato de pago del valor de los activos a la madurez si la cota no se alcanza y el precio de un *call* con barrera. Sea $H(V, T, L)$ ⁶ el valor de un contrato que paga una unidad a la madurez si el valor de los activos está por encima de L , por lo tanto si denotamos $C^{BS}(V, K, T)$ como una opción tipo *call* europea, se tiene que el valor al tiempo $t = 0$ de la función de pago de un bono con madurez T está dada por:

$$B^m(V, T, F, L, \alpha) = BL(Ve^{-\alpha T}, T, Fe^{-\alpha T}, Le^{-\alpha T}),$$

con

$$\begin{aligned} BL(V, T, F, L) &= LH(V, T, L) - L \left(\frac{L}{V} \right)^{\left(\frac{2r}{\sigma^2} - 1 \right)} H \left(\frac{L^2}{V}, T, L \right) \\ &\quad + C^{BS}(V, L, T) - \left(\frac{L}{V} \right)^{\left(\frac{2r}{\sigma^2} - 1 \right)} C^{BS} \left(\frac{L^2}{V}, L, T \right) \\ &\quad - C^{BS}(V, F, T) + \left(\frac{L}{V} \right)^{\left(\frac{2r}{\sigma^2} - 1 \right)} C^{BS} \left(\frac{L^2}{V}, F, T \right). \end{aligned}$$

Mientras que el valor al tiempo $t = 0$ de la función de pago de un bono con madurez T si el incumplimiento ocurre antes de la fecha de maduración es:

$$B^b(V, T, L, \alpha) = L \frac{e^{b\mu}}{e^{b\beta}} \left(\varphi \left(\frac{b - \beta T}{\sqrt{T}} \right) + e^{2\beta b} \varphi \left(\frac{b + \beta T}{\sqrt{T}} \right) \right).$$

Se observa que el valor de las acciones en el modelo Black y Cox sin dividendos es una opción *call* del tipo *down-and-out*, entonces para determinar el valor B^b se utiliza la diferencia entre el valor de los activos y el valor de las acciones, ver Lando (2004). En este modelo, la probabilidad de incumplimiento está dada por:

$$P[\tau \leq T | \tau > t] = \varphi(d_1) + e^{2 \left(r - \frac{\sigma^2}{2} \right) \ln \left(\frac{L}{V_t} \right) \frac{1}{\sigma^2}} \varphi(d_2),$$

donde

⁶ $H(V, T, L) = e^{-rT} \varphi \left(\frac{\ln \left(\frac{V}{L} \right) + \left(r - \frac{1}{2} \sigma^2 \right) T}{\sigma \sqrt{T}} \right), V > L.$

$$d_1 = \frac{\ln\left(\frac{L}{V_t}\right) - \left(r - \frac{\sigma_V^2}{2}\right)(T - t)}{\sigma_V \sqrt{T - t}},$$

$$d_2 = d_1 - \sigma_V \sqrt{T - t}.$$

8.3 Modelos de Intensidad

Los modelos de intensidad o reducidos representan el tipo más extendido de modelos de forma reducida. Estos modelos no toman en cuenta la relación entre el incumplimiento y el valor de la empresa de manera explícita. En contraste con los modelos estructurales, el momento de incumplimiento en los modelos de intensidad o reducidos no se determina a través del valor de la empresa, sino es el primer salto de un proceso dado exógenamente, ver Embrechts (2005). Otra diferencia es que las tasas de recuperación, es decir, la recuperación de la empresa como porcentaje del valor de mercado se especifica exógenamente.

De manera general, los modelos de intensidad suponen una relación entre el tiempo de incumplimiento (condicionado a la filtración \mathfrak{C}^7) y la filtración o información del mercado \mathfrak{F} , ver Grasselli (2010). Esencialmente se asume que el hecho de que se presente un incumplimiento o no al tiempo t no tiene impacto en la evolución de la filtración del mercado más allá de t .

Considérese a X_t un proceso indicador de incumplimiento dado por $X_t = I_{\{\tau \leq t\}}$ para todo $t \geq 0$. Si el evento de incumplimiento no influye en la tasa de incumplimiento, se puede imaginar que el proceso X es un proceso continuo después del incumplimiento a la misma tasa λ_t . Por lo tanto, $X_t = N_{t \wedge \tau}$ donde N_t es un proceso de conteo no decreciente con valores en los enteros tal que $N_0 = 0$. Entonces el tiempo de incumplimiento está definido por:

$$\tau = \inf\{t | N_t > 0\}.$$

El modelo de intensidad más sencillo es cuando se considera una tasa de incumplimiento, $\lambda > 0$, constante. Por lo tanto es natural modelarlo con un proceso de Poisson⁸ con parámetro $\lambda > 0$. En particular, definiendo el tiempo de incumplimiento como el tiempo de arribo del primer salto del

⁷ \mathfrak{C}_t es la filtración generada por el proceso H , es decir, $\mathfrak{C}_t = \sigma\{X_s, 0 \leq s \leq t\}$.

⁸ Un proceso Poisson con parámetro $\lambda > 0$ es un proceso no decreciente con valores en los enteros cuyo valor inicial es $N_0 = 0$.

proceso N_t , entonces la esperanza del tiempo de incumplimiento es $1/\lambda$ y la probabilidad de sobrevivencia después de t años es,

$$P[\tau > t] = E^*[N_t = 0] = e^{-\lambda t}.$$

Por lo tanto, la probabilidad de sobrevivencia se puede interpretar como un factor de descuento a cierta tasa λ .

Un modelo más general resulta de considerar una tasa de incumplimiento estocástica, éste modelo es conocido como proceso Cox⁹ o proceso doblemente estocástico. Dado que la tasa de incumplimiento tiene un comportamiento similar a la tasa de interés a corto plazo, algunos autores como Embrechts (2005) y Jorda (2010) sugieren modelar a la tasa de incumplimiento como un proceso similar a la tasa de interés a corto plazo. En estos modelos se tiene que la probabilidad de sobrevivencia tiene la forma:

$$P^*[\tau > t] = E^* \left[e^{-\int_0^t \lambda_s ds} \right].$$

Considerando un bono cupón cero con maduración T , en donde el tiempo de incumplimiento τ tiene una intensidad λ_s , y asumiendo que existe un proceso de tasa corta r_s tal que el precio del bono cupón cero sin incumplimiento está dado por:

$$B(t, T) = E^* \left[e^{-\int_t^T r_s ds} | \mathfrak{F}_t \right].$$

Si se supone que no hay tasa de recuperación, el precio del bono con riesgo de incumplimiento en el tiempo cero está dado por:

$$\mathbb{B}(0, T) = E^* \left[e^{-\int_0^T r_s ds} I_{\{\tau > T\}} \right] = E^* \left[e^{-\int_0^T (r_s + \lambda_s) ds} \right],$$

dado que $E^* \left[E^* \left[e^{-\int_0^T r_s ds} I_{\{\tau > T\}} | \mathfrak{F}_T \right] \right] = E^* \left[e^{-\int_0^T r_s ds} E^* \left[I_{\{\tau > T\}} | \mathfrak{F}_T \right] \right]$.

La fórmula anterior es general y se puede observar el impacto en la valuación del bono si r_s y λ_s son dependientes. Por otro lado, si se asume independencia entre las dos variables se tiene que

⁹ El proceso N_t es un proceso Cox, si condicionado a la información disponible \mathfrak{F}_t al tiempo t , N_t es un proceso Poisson no homogéneo con intensidad variable en el tiempo $\lambda(s) = \lambda_s$ para $0 \leq s \leq t$.

$$B(0, T) = E^* \left[e^{-\int_0^T (r_s + \lambda_s) ds} \right] = B(t, T) P^*[\tau > T].$$

Un bono cupón cero con incumplimiento es igual a un bono cupón cero sin incumplimiento por su probabilidad de sobrevivencia bajo la medida neutral al riesgo P^* . En general el supuesto de independencia resulta poco probable, por lo tanto la dependencia entre los dos procesos se puede modelar a través de la correlación entre ambas tasas. De esta manera el diferencial de rentabilidad para los modelos de intensidad tiene la siguiente forma:

$$Spread = \frac{1}{T} \ln \left(E^* \left[e^{-\int_0^T (r_s + \lambda_s) ds} \right] \right) - r = \frac{1}{T} \ln \left(\frac{e^{-\int_0^T (r_s + \lambda_s) ds}}{e^{rT}} \right).$$

Otra alternativa para valuar un bono corporativo es a través de la tasa de recuperación. En cuanto a la parametrización de la tasa de recuperación, se han propuesto tres procedimientos diferentes en la literatura, ver Embrechts (2005). La primera de ellas considera que la tasa de recuperación es una fracción exógena del valor nominal del bono con incumplimiento (*Recovery of Face Value, RFV*). Jarrow y Turnbull consideran la tasa de recuperación como una fracción exógena del valor de un equivalente al bono libre de incumplimiento (*Recovery of Treasury, RT*). Por último, Duffie y Singleton fijan una tasa de recuperación igual a una fracción exógena del valor de mercado de los bonos, justo antes del incumplimiento (*Recovery of Market Value, RMV*).

8.4 Aplicación a dos empresas mexicanas

Con ayuda de los modelos estructurales y los modelos de intensidad se pretende modelar el comportamiento a un año para Cemex y Bimbo de la deuda a corto plazo, entendiéndose por ello, la deuda emitida con duración menor o igual a un año, como estimar su probabilidad de incumplimiento. Cabe aclarar que los resultados que se muestran en este estudio son presentados en unidades monetarias de México, aunque un análisis similar se podría llevar a cabo en dólares utilizando, por ejemplo, la T-bill como tasa de interés libre de riesgo. Uno de los retos que presentó este trabajo fue poder calibrar los modelos con la información que dispone cualquier inversionista en el mercado mexicano. Para los modelos estructurales, la información se obtuvo de Bloomberg.

Bajo los supuestos del modelo de Merton se requiere conocer el valor de sus activos y el de sus acciones para el tiempo inicial $t = 0$, así como estimar la tasa de interés libre de riesgo del mercado y el horizonte de tiempo. Además hay que estimar la volatilidad de sus acciones, a través de modelos tipo GARCH o usando datos históricos.

En la siguiente tabla se presenta la información financiera básica, obtenida de Bloomberg, de las dos empresas para la implementación del modelo de Merton: el valor de los activos V_0 , el valor de las acciones S_0 y el valor de sus pasivos (corto plazo) D_0 para diciembre de 2011. Como las empresas que se están analizando cotizan en el mercado mexicano, se pudo obtener el valor de mercado del capital accionario, multiplicando el número de acciones en circulación por el precio de mercado de éstas. Para obtener el valor de la deuda, se propone considerar solamente la deuda a corto plazo obtenida del estado de resultados de cada empresa. También se presenta la estimación¹⁰ de la volatilidad del rendimiento de los precios σ_S y el rendimiento del activo libre de riesgo¹¹. El horizonte de tiempo para el cual se realizaron los cálculos fue $T = 1$, que corresponde a un año.

Variable	CEMEX	BIMBO
V_0	231,958,518	64,785,160
S_0	175,288,070	38,294,997
D_0	56,670,448	26,490,163
σ_0	0.60445331	0.28176025

Tabla 8.1. Información de las dos empresas para diciembre de 2011 (en miles de pesos). Fuente: Bloomberg.

Para el modelo de Merton con tasa de interés estocástica, se propone modelar la tasa de interés a corto plazo con el modelo de Vasicek dado que existe una forma cerrada para estimar el valor del bono corporativo como una opción del tipo europea. Sin embargo, con objeto de comparar los resultados del valor de la deuda bajo el modelo de Vasicek, se presenta la aproximación propuesta por Yong (2001).

¹⁰ La estimación de la volatilidad fue implementada utilizando el modelo $GARCH(1,1)$.

¹¹ Cetes a 28 días según la subasta del último día de cotización del 2011, $r=4.31$.

Por último, para los modelos estructurales se presentan los resultados del modelo Black y Cox, en el cuál se asume que el incumplimiento ocurre cuando el valor de los activos alcanza una cota inferior con valor L . Dicho valor fue simulado al 1% de confianza¹², asumiendo que V_t sigue un movimiento browniano bajo la medida neutral al riesgo y tomando como base la condición $V > L$. Al igual que en el modelo de Merton, en éste modelo se asume que no hay pago de dividendos, por lo tanto se muestra el valor de la deuda cuando el incumplimiento no ocurrió en el tiempo de maduración el bono. El valor del bono cuando el incumplimiento ocurre antes de la fecha de maduración es simplemente la diferencia entre el valor de mercado de los activos y el valor de mercado de las acciones, ver Lando (2004).

A continuación se muestran las estimaciones del valor de la deuda de Cemex y Bimbo bajo los modelos estructurales antes mencionados. Así como estimación de la probabilidad de incumplimiento (PI) y el spread o diferencial de rentabilidad anual (DC) para el año 2012.

Variable	Merton	Merton (Vasicek)	Black-Cox
F	59.145,120	59,129,950	55,842,460
σ_V	0.4569894	0.456779	0.4569894
PI	0.00215295	0.00375448	0.1871314
DC	0.03681717	0.00231523	0.000677748

Tabla 8.2. Resultados asociados a Cemex utilizando Newton-Raphson.

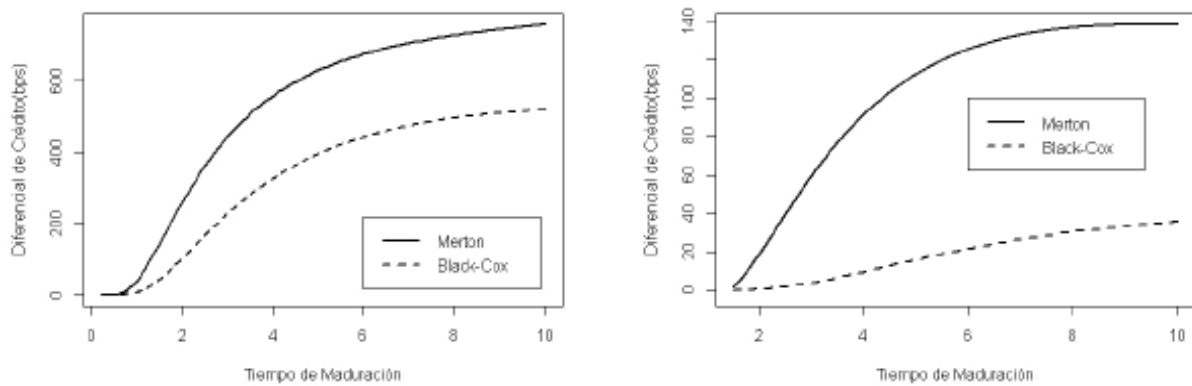
Variable	Merton	Merton (Vasicek)	Black-Cox
F	27,639,840	27,639,980	28,741,333
σ_V	0.166550	0.176503	0.4569894
PI	$6.2 e^{-6}$	$8.3 e^{-5}$	0.1447
DC	$1.07 e^{-5}$	$3.3 e^{-5}$	$5.6 e^{-6}$

Tabla 8.3. Resultados asociados a Bimbo utilizando Newton-Raphson.

¹² Se realizaron 100,000 simulaciones de V_t . El nivel de confianza es el porcentaje de trayectorias simuladas que al tiempo $t = T$ estuvieron por debajo de la cota L .

Adicionalmente con el fin de comparar la estimación propuesta en Yong (2001) se estimó el valor de la deuda bajo el modelo CIR, obteniendo un valor de \$59,129,990 para Cemex y \$27,639,840 para Bimbo. Estos resultados muestran que ambos modelos dan resultados similares.

Con los resultados obtenidos en los tres modelos estructurales antes estudiados, se puede concluir que Cemex tiene mayores diferenciales de rentabilidad que Bimbo, lo cual modela bien el hecho que la primera empresa tiene mayor apalancamiento que la segunda. Los diferenciales se traducen en un mayor exceso de rentabilidad del bono de Cemex sobre el bono de Bimbo, ambos respecto a la tasa de Cetes. Comparando estos resultados con la calificación para deuda a corto plazo que otorgó Standard & Poor's se tiene que para Cemex es mxB, mientras que para Bimbo es de mxAA+, por lo que se espera que el diferencial de rendimiento sea mayor en Cemex que en Bimbo. En el caso del modelo de Cox las probabilidades de incumplimiento no reflejan las diferencias en el nivel de la calificación. Una comparación importante entre los modelos de valuación de bonos corporativos es el diferencial de rentabilidad, pues, cuando no hay pago de dividendos, el modelo de Black-Cox muestra menores diferenciales en comparación con el modelo clásico de Merton¹³. En la Figura 8.1 se puede observar que el diferencial de rentabilidad, tanto para Cemex como para Bimbo, muestra menores diferenciales o spreads en el modelo Black-Cox que en el modelo de Merton.



a) Diferencial de rentabilidad de Cemex b) Diferencial de rentabilidad de Bimbo.

Figura 8.1. Comparativo del diferencial de rentabilidad para el modelo de Merton y Black-Cox.

¹³ Ver Lando (2004).

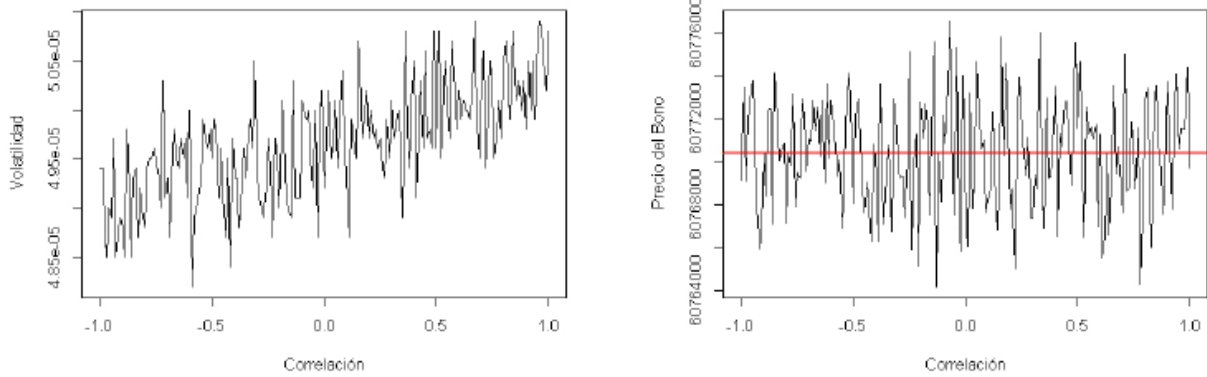
En los modelos de intensidad, una de las principales dificultades es una correcta estimación de los parámetros de la tasa de interés y de la tasa de incumplimiento. Debido a que en el mercado no se tiene un bono corporativo similar al utilizado en este estudio, no fue posible estimar la probabilidad de incumplimiento de manera implícita. Por lo tanto, algunos autores sugieren utilizar la tasa publicada por las agencias calificadoras referentes al incumplimiento.

En este trabajo se usó la información publicada anualmente por Standard & Poor's en su *Estudio anual de incumplimiento y análisis de transición de calificaciones de los sectores corporativos, de servicios financieros y finanzas públicas*¹⁴.

La valuación del bono corporativo se realizó bajo el supuesto de dependencia entre el tiempo de incumplimiento τ y a tasa de interés a corto plazo r_t . Dicha dependencia puede ser resultado de la dependencia que existe entre la intensidad estocástica λ_t y la tasa de interés r_t . Dado que la tasa de incumplimiento se comporta de manera similar a la tasa de interés, se propone utilizar el modelo Vasicek para modelar ambas tasas. En éste modelo se tuvo que hacer un ajuste en el periodo de análisis, debido a que la tasa de incumplimiento disponible se presenta en forma anual.

El precio del bono corporativo para Cemex y Bimbo depende de la correlación entre los dos movimiento brownianos estándar. La correlación, definida por $\rho dt = dW_t^1 dW_t^2$, modela la dependencia entre la tasa de interés a corto plazo r_t y la tasa de incumplimiento λ_t , con $\rho \in [-1,1]$. En Ekstrom y Tysk (2008) se muestran ciertas propiedades de los parámetros de algunos procesos de difusión, de manera particular se muestra que el precio del bono es creciente respecto a la volatilidad del proceso subyacente. La evolución del precio del bono de Cemex y Bimbo se presenta en las siguientes gráficas, intentando mostrar que el precio del bono es una función creciente como función de la correlación entre los dos movimientos brownianos.

¹⁴Página electrónica: www.standardandpoors.com.



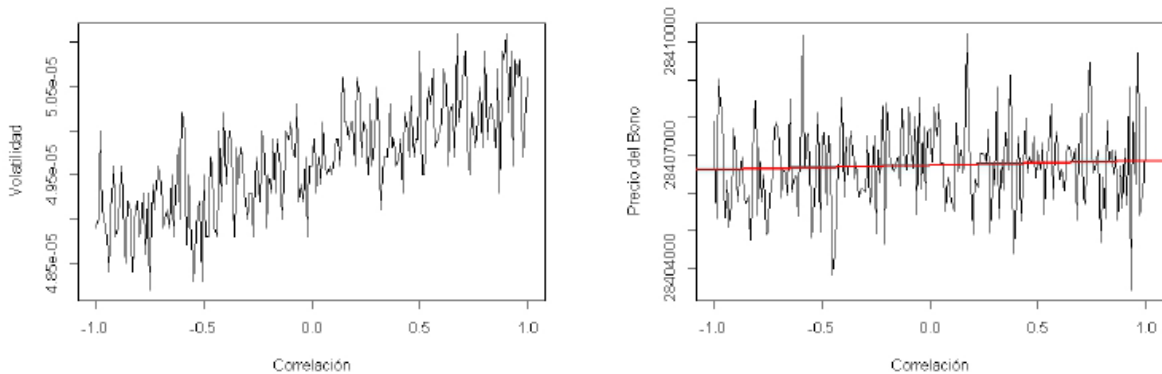
a) $Var(\int_0^1 (r_s + \lambda_s) ds)$

b) Precio del bono

Figura 8.2. Evolución del precio del bono corporativo para Cemex como función de la correlación.

En la Figura 8.2 a), se observa que una correlación negativa entre r_t y λ_t disminuye la volatilidad de la suma, mientras que una correlación positiva aumenta la volatilidad, y por lo tanto la volatilidad es creciente cuando la correlación también crece. El incremento en el precio del bono se debe al incremento en la volatilidad de la suma de los procesos. Sin embargo, en la Figura 8.2 b) no se puede observar de manera clara la tendencia positiva del precio del bono como función de la correlación.

Por su parte para la valuación del bono corporativo de Bimbo se obtuvieron resultados similares, los cuales se muestran en la siguiente figura:



a) $Var(\int_0^1 (r_s + \lambda_s) ds)$

b) Precio del bono.

Figura 8.3. Evolución del precio del bono corporativo para Bimbo como función de la correlación.

Para el caso de Bimbo, se observa una ligera tendencia positiva en la evolución del precio del bono como función de la correlación, a diferencia del bono de Cemex. Por lo tanto, se pudo corroborar una de las afirmaciones de Ekstrom y Tysk (2008) para el precio de un bono.

Para estimar la correlación (ρ) entre las dos tasas, se propone discretizar el modelo de Vasicek para ambas tasa y dado que la correlación es lineal bajo transformaciones, se tiene que los residuales de r_t y λ_t , y los dos movimientos brownianos tiene una correlación de $\rho = .83$. El precio de los bonos de Cemex y Bimbo es:

Precio	CEMEX	BIMBO
Bono	60,770,322	28,405,271

Tabla 8.4. Estimación del precio del bono en base a la correlación de W_t^1 y W_t^2 .

En la tabla anterior se muestra que el precio del bono con incumplimiento para Cemex, bajo el modelo de intensidad es del \$60,770,322; mientras que el precio del bono de Bimbo con incumplimiento es de \$28,405,271. En general, el modelo de intensidad generó valores más grandes para los bonos que los modelos estructurales. Por lo tanto, la estimación de la deuda para finales del 2012 para ambas empresas es mayor bajo los modelos de intensidad.

8.5 Conclusiones

- I. Los modelos estructurales establecen un vínculo entre la calidad crediticia de una empresa y sus condiciones económicas y financieras y así, el incumplimiento se genera de forma endógena. Las condiciones del mercado sólo se toman en cuenta a través de la tasa de interés de referencia, siempre y cuando ésta se modele estocásticamente. Se observó que en este caso, el modelo de Vasicek y CIR dan resultados similares.
- II. En el modelo de Merton la probabilidad de incumplimiento es cercana a cero, dado que el incumplimiento sólo puede ocurrir al tiempo T . Para corregir esta deficiencia se introduce en el modelo de Black y Cox una barrera. Si ésta se supone constante, como en este trabajo, también puede predecirse el momento del incumplimiento. Para evitar este

problema se debe definir una barrera estocástica que, en cierta forma, es lo que los métodos de intensidad hacen.

- III. Un problema que surgió en este trabajo fue cómo determinar la barrera para el modelo de Black y Cox. La manera natural es relacionar la barrera con la tasa de recuperación. Sin embargo, no se cuenta en México con la información necesaria para establecer dicho porcentaje; el cual dependerá del punto de vista de los interesados, ya que no será el mismo para los accionistas o para los tenedores de bonos. A falta de información se utilizó un argumento estadístico. Los resultados obtenidos muestran que el valor de la deuda es muy sensible al valor de la barrera.
- IV. Para ambas empresas, el diferencial de rendimiento es menor en el modelo de Black y Cox que el de Merton lo que está acorde con lo observado, ver Lando. Esta diferencia puede explicarse por el nivel asignado a la barrera.
- V. La principal dificultad que surge al implementar los modelos de intensidad es la estimación de los parámetros del proceso que modela la tasa de incumplimiento. En este trabajo se utilizó la información que publica la agencia Standard & Poor's. Los datos que se incluyen en esta publicación son escasos y de tipo anual por lo que sería conveniente en México llevar un registro sistemático mensual de la tasa de incumplimiento de las empresas mexicanas. Otra posibilidad de estimar los parámetros es de forma implícita utilizando la información del mercado. En México no contamos con instrumentos de deuda tipo Swaps de crédito. Vale la pena implementar en otro estudio otras alternativas de estimación que se han propuesto para el mercado mexicano, ver Reynoso (2012).
- VI. Los resultados obtenidos con el modelo de intensidad son para ambas empresas mayores, aunque cercanos, a los que se obtienen con el Modelo de Merton con tasa estocástica, a pesar que las dos metodologías son totalmente distintas.

- VII. Cualitativamente los resultados obtenidos están acordes con las calificaciones que recibieron a lo largo de este año la deuda a corto plazo de ambas empresas.

8.6 Bibliografía

- Ball, C., Torous, W. (1996). Unit Roots and the estimation of Interest Rate Dynamics, *Journal of Empirical Finance*, 3:215-238.
- Björk, T. (2004). Arbitrage theory in continuous time, Segunda Edición, Oxford University Press.
- Black, F., Cox, J. (1975). Valuating Corporate Securities: Some Effects of Bond Indenture Provisions, *Journal of Finance*, 31, 2, 351-367.
- Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 3, 673-654.
- Brennan, M., E. Schwartz, (1980). Analyzing Convertible Bonds, *Journal of Financial and Quantitative Analysis*, 15, 907-929.
- Brigo, D., Mercurio, F. (2006). Interest Rate Models-Theory and Practice, Segunda Edición, Springer.
- Cox, J.C., Ingersoll, J.E., Ross, S.A. (1985). A Theory of Term Structure of Interest Rates, *Econometrica*, 53, 385-407.
- Das, S., P. Tufano, (1995). Pricing Credit Sensitive Debt When Interest Rates, Credit Ratings, and Credit Spreads Are Stochastic, *Journal of Financial Engineering*, 5, 161-198.
- Ekström, E., Tysk, J. (2008). Convexity theory for the term structure equation, *Finance Stochastic*, 12, 117-147.
- Elizalde, A. (2012a). Credit Risk Models I: Default Correlation in Intensity Models, *Frank Fabozzis Encyclopedia of Financial Models*, Chp. 22.
- Elizalde, A. (2012b). Credit Risk Models I: Structural Models, *Frank Fabozzis Encyclopedia of Financial Models*, Chp. 23.
- Embrechts, P., A.J. McNeil, R. Frey. (2005). Quantitative Risk Management, Princeton, *Series in Finance*, USA.

- Feller, W. (1951). Two Singular Diffusion Problems, *Annals of Mathematics*, Vol. 54, No.1.
- Grasselli, M. R., Hurd, T. R. (2010). Credit Risk Modeling, *working paper*, University Hamilton.
- Harrison, J. (1990). Brownian motion and stochastic flow systems, Malabar, FL: Robert E. Krieger.
- Heath, D., Jarrow, R., Morton, A. (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation, *Econometrica*, Vol. 60, 77-105.
- Jarrow, R., and S. Turnbull. (1995). Pricing Derivatives on Financial Securities Subject to Default Risk, *Journal of Finance*, 50, 53-86.
- Jarrow, R., D. Lando, S. Turnbull. (1997). A Markov Model for the Term Structure of Credit Spreads. *Review of Financial Studies*, 10, 481-523.
- Jorda, P.G. (2010). Modeling Credit Risk through Intensity Models, *Working paper*.
- Karatzas, I., Shreve, S. (1988). Brownian Motion and Stochastic Calculus, Springer-Verlag N.Y.
- Karatzas, I., Shreve, S. (1991). Stochastic Calculus for Finance II, *Springer-Verlag* N.Y.
- Kim, Y.-J. y Kunitomo, N. 1999. Pricing options under stochastic interest rates: a new approach. *Asia-Pacific Financial Markets* 6, 49-70.
- Lando D. (2004). Credit risk modeling, Theory and Applications, Princeton University Press.
- Lamberton, D., Lapeyre, B. (2008). Introduction to Stochastic Calculus Applied to Finance, Chapman & Hall, 2a edición.
- Longstaff, F., E. Schwartz. (1995). Valuing Risky Debt: A New Approach. *Journal of Finance*, 50, 789-820.
- Mendez M. (2010). Aplicaciones del cálculo estocástico al estudio de la estructura temporal de tasa de interés, Tesis Maestría, UAM.
- Merton, R.C. (1974). On the pricing of Corporate Debt: The Risk Structure of Interest Rates, *The Journal of Finance*, 29.
- Sukhomlin, N., Santana, L. (2010). Problema de calibración de mercado y estructura implícita del modelo de bonos de Black-Cox, *Revista de métodos cuantitativos para la economía y la empresa* (10). 73-98.

- Téllez, C. (2010). Medición del riesgo en crédito: Implementación y cálculo del VaR y el CVaR en tres modelos de incumplimiento, Tesis Maestría, UAM.
- Reynoso, V. J., Venegas, M., F., Rodríguez, N., A. (2012). Avances recientes en valuación de activos y administración de riesgos, *Universidad Panamericana*, vol. 3, 113-141.
- Yong-Jin, Kim. (2001). Option Pricing under Stochastic Interest Rates: An Empirical Investigation, *Asia-Pacific Financial Markets*. 9, 23-44.
- Zhou, C. (1997). A Jump-Diffusion approach to Modeling Credit Risk and Valuing Defaultable Securities, *Board of Governors of the Federal Reserve System Working Paper* 97-15.

Capítulo 9

Modelo de valuación de opciones con volatilidad estocástica mediante simulación Monte Carlo

Adrián Hernández-del-Valle

Sección de Estudios de Posgrado e Investigación, Escuela Superior de Economía del I.P.N.

Claudia I. Martínez García

Sección de Estudios de Posgrado e Investigación, Escuela Superior de Economía del I.P.N.

9.1 Introducción

Proponemos un modelo de opciones de Monte Carlo con volatilidad estocástica para la valuación de bonos comprables. Un bono comprable, también conocido como bono redimible, le da al colocador---el deudor---el derecho a retirar el bono antes del vencimiento, generalmente a cambio de una prima que se paga al tenedor del bono---el originador o acreedor---. La causa principal del ejercicio de la opción de compra---retirar el bono antes del vencimiento---es una baja en las tasas de interés que permite al deudor refinanciar a tasa menor. Por ejemplo, una hipoteca se puede pensar como un bono comprable, un bono respaldado por hipoteca BORHI. En caso de una baja en las tasas hipotecarias, el deudor puede obtener una segunda hipoteca a tasa menor y prepagar la hipoteca original, i.e. si bajan las tasas, ejerce la opción de comprar su deuda. El acreedor u originador puede castigar el prepago---e.g. imponiendo una prima por pago anticipado---, o no. Los esquemas de prepago han migrado de esquemas que penalizaban el prepago a aquellos que

no lo hacen. El modelo que proponemos tiene gran capacidad para incorporar todo tipo de riesgos:

- riesgo de movimientos en la tasa de referencia y en la curva de rendimientos;
- riesgo de impago (*default*);
- riesgo de prepago total o parcial (en inglés, *prepayment* y *curtailment* respectivamente);
- riesgo de cambios en el precio del subyacente; y
- riesgo de cambios en el precio del subyacente del subyacente---componente fundamental del bono respaldado por activos, BORA---un BORHI es un caso particular de un BORA---, como veremos mas adelante.

La idea fundamental es que necesitamos tres árboles binomiales: uno para estimar las tasas spot. En la estimación de estas tasas de descuento necesitamos considerar la volatilidad y escenarios. Estas tasas son muy importantes porque permiten descontar de manera “justa” prácticamente cualquier instrumento en la economía con base en el supuesto de riesgo-neutralidad.

El segundo árbol sirve para valorar el bono completo. Es decir, la canasta de préstamos sin riesgo de prepago ni de incumplimiento.

El tercer y último árbol binomial sirve para valorar el bono comprable. Recordemos que el comprable le da al deudor el derecho a retirar el bono antes del vencimiento; y que la causa principal para ejercer la opción es una baja en las tasas, e.g. cuando el deudor refinancia a una tasa menor. Es decir, el árbol binomial que represente el comportamiento del comprable debe servir para simular los diferentes riesgos a los que se encuentra expuesto el tenedor del bono: incumplimiento y prepago.

Bajo condiciones normales, precisamente una baja en la tasa de referencia generaría una revaluación del bono al alza, porque significa que el tenedor---el inversionista que está largo el bono comprable---"amarró" una tasa atractiva en un escenario de tasas a la baja. Pero resulta que la opción incrustada en este bono es ejecutable precisamente cuando el bono debería empezar a ganar valor. Esta es la razón por la que el prepago es un riesgo. El modelo de Kalotay, Williams y Fabozzi (1992), KWF, proporciona una solución intuitiva para la representación del riesgo de prepago: en cada nodo donde se puede ejercer la opción de compra, y donde el valor estimado recursivamente por el modelo, V_T , supera el precio de ejercicio, C , simplemente se sustituye V_T por C . Esto permite valorar la opción como parte integral del activo subyacente (la canasta de hipotecas).

Cabe mencionar que se aplica exactamente la misma técnica al caso del bono vendible, sólo que los valores que se sustituyen son aquellos de la opción de venta, i.e. en los nodos superiores del árbol binomial donde el valor estimado, V_T , es menor al valor de ejercicio P : tachar V_T sustituir por P .

Una vez que se tiene el valor del bono completo y del comprable, el valor de la opción se calcula usando la ecuación

(0): Valor del bono comprable = valor del bono completo - valor de la opción de compra sobre el bono completo.

Ahora bien, es importante notar que KWF presuponen agentes que pueden refinanciar de manera instantánea y sin fricciones. Sin embargo, en la realidad, esto es sólo cierto de operadores de valores y en mercados desarrollados. En la realidad, el prepago es poco sensible a variaciones en la tasa de interés. Presuponer prepago óptimo implica que el precio que genera nuestro modelo es mínimo posible.

Dos defectos del KWF que nosotros solucionamos son: i) presuponen una volatilidad constante; y ii) fallan en incorporar el riesgo de incumplimiento.

Las probabilidades de incumplimiento son independientes de las tasas de interés. Nosotros incorporamos tasas de incumplimiento dependiendo de las características de los deudores. La idea es que en México, la mayor parte de los créditos son a tasa fija, entonces la gente incumple por razones no necesariamente correlacionadas con los cambios en las tasas de interés. En México, las carteras de créditos que respaldan los bonos respaldados por hipotecas, BORHI's, muestran patrones homogéneos por segmento de mercado, y esto impacta las condiciones de colocación de los instrumentos colateralizados. Por ejemplo, la cartera de Su Casita tiende a incumplir mucho---tiene un alto riesgo de crédito---, esto merma las perspectivas de precio e incrementa los cupones; la cartera de BBVA-Bancomer tiene expectativas de incumplir poco y prepagar poco---poco riesgo de crédito y riesgo de prepago---, i.e. los flujos son menos volátiles por ende los instrumentos colateralizados que deriven de estas carteras tendrán precios mayores y cupones menores. Finalmente, la cartera de Metrofinanciera tiende a incumplir poco pero a prepagar mucho---alto riesgo de prepago---. El riesgo de prepago merma el precio e incrementa las tasas de los instrumentos. Recordemos que el prepago es un riesgo porque los deudores prepagan a la par, i.e. sólo el saldo del principal, pero el emisor de los instrumentos (el fideicomiso o special purpose vehicle SPV) está obligado a pagar capital más intereses. Un modelo de valuación de BORA's en general debe ser capaz de incorporar este tipo de características.

El supuesto respecto al comportamiento de los deudores que constituyen la garantía asume una correlación de 1 entre el comportamiento de las tasas de interés y el riesgo de prepago: el prepago es perfectamente elástico a la tasa de interés. El incumplimiento está correlacionado con la tasa pero la correlación es irrelevante. Las garantías incumplen con base en patrones establecidos y conocidos de pago, y dependiendo del bono que estemos valuando. Por ejemplo, en la Figura 9.1 se presupone que el bono titular A está libre de riesgo de incumplimiento por tanto la correlación con movimientos de la tasa de referencia es 0. Esto no es cierto en el caso de los bonos subordinados.

Figura 9.1. Resumen del portafolio de casas habitación del banco XYZ

Cantidad de principal pendiente	US\$500,000
Número de préstamos	19,000
Tasa cupón	9.00% a 23.00%
WACC	13.11%
Seasoning	1 a 2 meses
Seasoning ponderado	promedio 40 meses
Madurez restante	1 a 400 meses
WAM	200 meses
Balance corriente	\$2.00 a 80,000.00
Balance promedio	\$20,000.00

Figura 9.2. Cartera de préstamos a bursatilizar

Activos	Pasivos
Préstamos convertibles \$458.50	Bono titular, A, calificación AA, \$286.56 (62.5%)
	Bono subordinado, B, sin calificación \$85.97 (18.75%)
	Bono súper-subordinado, C, \$85.97 (18.75%)
	Exceso de servicio \$1.50 (aprox. 35 puntos base, bps, del principal, \$500.00, a 15% de tasa de descuento)

Nuestro supuesto respecto al prepago es muy fuerte, la implicación fundamental es que el resultado de nuestra valuación es una cota inferior del valor del comprable (un *peor escenario posible*).

Ahora bien, el cálculo del precio requiere 3 árboles:

i) uno que simule el comportamiento de la tasa de referencia: la tasa de interés a un año; este primer árbol es crucial porque de aquí salen las tasas spot que usaremos para descontar los otros instrumentos bajo el supuesto de riesgo-neutralidad. Este árbol emplea volatilidad estocástica.

ii) Un segundo árbol se usa para valorar el precio de las garantías sin prepago ni default, el bono *completo* o bono valor-en-libros; y iii) un último árbol para valorar el bono con opciones de compra incrustadas, i.e. el bono *comprable*. Este último simula el comportamiento del deudor: cuando bajan las tasas crecen los prepagos. Por otra parte, aunque los créditos sean a tasa fija, esto no exenta al deudor del riesgo de tasa de interés porque es probable que tenga otros créditos que pagar, e.g. tarjetas de crédito. Por lo anterior, cuando suben las tasas presuponemos que aumentan los incumplimientos dependiendo del tipo de bono: titular o subordinado; y también dependiendo del vendedor de la cartera. Recordemos que, por ejemplo, los deudores de Su Casita exhiben patrones de comportamiento distintos a los de Bancomer.

Finalmente, decimos que nuestro modelo es de Monte Carlo porque podemos simular el comportamiento de los componentes estocásticos, en particular, la volatilidad.

Con base en lo anterior, lo que resta del trabajo está organizado de la manera siguiente: primero describimos la construcción del árbol binomial de tasas de spot; después el bono completo y por último el comprable. La sección final presenta las especificaciones para realizar la simulación Monte Carlo, así como la reestimación del bono comprable y la postura con base en 100 escenarios.

9.2 El árbol binomial de la tasa de interés a un año

La Figura 9.3 muestra el ejemplo de un árbol binomial de tasas de interés. En este árbol, los nodos están espaciados en intervalos de 1 año de izquierda a derecha. Cada nodo N tiene un subíndice que indica el camino que ha seguido la tasa a un año para alcanzar ese nodo. La tasa de interés a un año puede moverse en una de dos direcciones al año siguiente: "arriba", u , o "abajo", d . Es importante mencionar que las tasas no necesariamente son más altas o bajas en el tiempo $t+1$ que en el tiempo anterior t : u significa que de las dos alternativas en el tiempo $t+1$, esta es la mayor y d es la menor. Con esto en mente, N_{uu} significa que la tasa a un año siguió el camino de arriba en los primeros dos años, tiempos $t=1$ y $t=2$. El nodo N es la base del árbol y r_0 es la tasa a un año en el presente.

El siguiente paso consiste en tomar un supuesto respecto a la probabilidad de que la tasa suba o baje: u ó d . Por ejemplo, Cox, Ross y Rubinstein (1979), CRR, asumen que u es función de la variancia del logaritmo del precio del subyacente σ^{2t} , i.e.

$$u = \exp\left(\sigma\sqrt{t}\right),$$

y d es la inversa

$$d = \exp\left(-\sigma\sqrt{t}\right) = \frac{1}{u}.$$

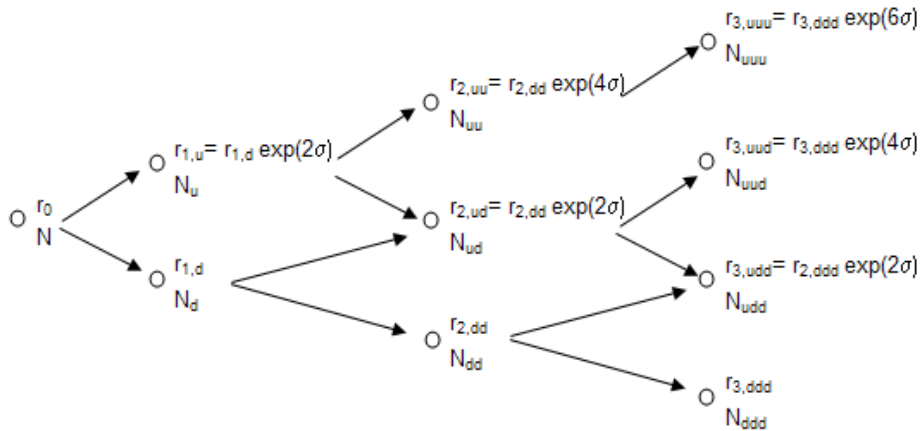
Por su parte, KWF asumen que la probabilidad es la misma, i.e. $u = d = 1/2$.

Nosotros tomamos la convención KWF. Esto nos permite aseverar que todas las tasas de los nodos superiores del árbol son función de la inferior, e.g.

$$r_{1,u} = r_{1,d} e^{2\sigma}, \tag{9.1}$$

donde $r_{1,d}$ y $r_{1,u}$ son las tasas forward a un año si las tasas bajan o suben respectivamente. La Figura 9.3 generaliza esta idea para un bono con $T=3$ años:

Figura 9.3. Tasas spot



El árbol de la Figura 9.3 representa una distribución lognormal en el tiempo porque el proceso estocástico r_t es multiplicativo. Si el proceso estocástico fuera aditivo, entonces la distribución de la Figura 9.3 sería la normal. Es convención usar la distribución lognormal para activos financieros porque sus precios son estrictamente positivos.

Ahora bien, la volatilidad de la Figura 9.3 es constante. Típicamente, $\sigma_P = \sigma_{\text{anual}} \sqrt{P}$.

Donde σ_P es la desviación estándar por periodo, e.g. mensual o trimestral; σ_{anual} es la desviación estándar anual de los rendimientos logarítmicos del instrumento; y P es la inversa de los periodos por año. Por ejemplo, la volatilidad mensual sería igual a:

$$\sigma_{12} = \sigma_{\text{anual}} \sqrt{1/12}.$$

Por otra parte, la volatilidad que presuponemos nosotros es estocástica. En particular, presuponemos que la dinámica de la volatilidad de tasa a un año está representada por un proceso de heteroscedasticidad condicional generalizado, GARCH de la manera siguiente:

$$dv_t = \theta(\omega - v_t)dt + \xi v_t dB_t, \quad dB_t \sim N(0, 1), \tag{9.2}$$

donde v_t es la volatilidad de la tasa de interés a un año observada en el tiempo t : la *volatilidad estocástica*; ω es la volatilidad media de largo plazo; θ es la tasa a la que v_t se revierte a su comportamiento promedio de largo plazo ω ; ξ es la volatilidad de v_t , i.e. la volatilidad de la volatilidad; y dB_t es un proceso normal estándar. Otra propiedad interesante es que se asume que la aleatoriedad de la tasa está correlacionada, ρ , con el proceso del precio subyacente.

Otros dos modelos de volatilidad estocástica populares son el de Heston y el 3/2. La diferencia básica entre los tres es el segundo término de (9.2): el modelo Heston toma $\xi\sqrt{v_t}dB_t$; y el modelo 3/2 asume $\xi v_t^{3/2}dB_t$.

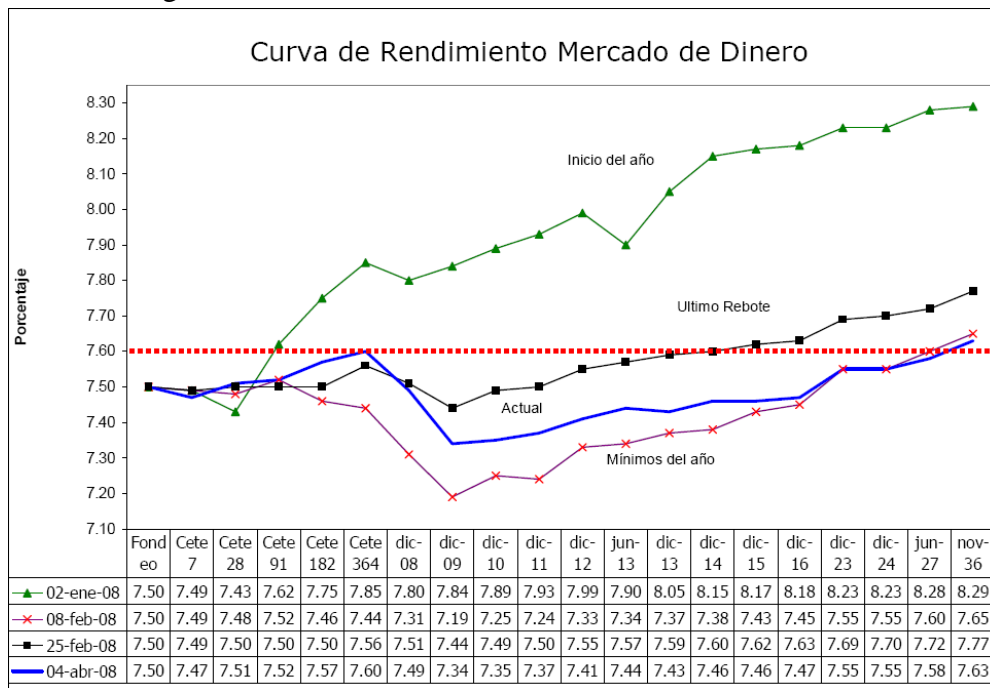
Preferimos el GARCH porque opinamos que el primero tiende a sobreestimar la volatilidad de r_t y el segundo a subestimarla.

Por otra parte, para determinar el valor del bono en un nodo, V , primero deben calcularse el valor de los dos nodos siguientes a la derecha, e.g. el valor del bono en N_u --- $V_{1,u}$ ---depende del valor en N_{uu} y en N_{ud} --- $V_{2,uu}$ y $V_{2,ud}$ respectivamente---. Ahora bien, el valor del bono en cada nodo se puede dividir en dos componentes: 1) el cupón; y 2) el valor del bono en el nodo futuro.

Los cupones futuros se toman de la curva de rendimientos---*yield curve*--- y varían según las condiciones del mercado de dinero. Por ejemplo, en los EE.UU. la convención es tomar las tasas LIBOR para la parte corta de la curva: plazos ≤ 3 meses; futuros para la parte media de la curva: $3m < t \leq 15m$; y swaps de tasas de interés para la parte larga de la curva: plazos de 1 año a 60 años.

En México, típicamente se usan Cetes a diferentes plazos para representar las partes corta y media de la curva; y bonos M para la parte larga como muestra la Figura 9.4:

Figura 9.4. Curva de rendimientos del mercado mexicano



Fuente: SIF-ICAP

Por su parte, V_t en el tiempo t es el promedio ponderado de ambos valores futuros $V_{t+1,u}$ y $V_{t+1,d}$:

$$V_t = u \left(\frac{V_{t+1,u} + C_{t+1}}{1 + r_t} \right) + d \left(\frac{V_{t+1,d} + C_{t+1}}{1 + r_t} \right). \quad (9.3)$$

Como mencionamos arriba, existen diferentes convenciones para estimar las ponderaciones, u y d , y nosotros tomamos la KWF, i.e. asumimos que ambos escenarios son igualmente probables por tanto $u=d=1/2$.

Ahora bien, (9.3) presenta un problema, tiene demasiadas incógnitas: desconocemos los valores en $t+1$, $V_{t+1, \cdot}$, pero también la tasa de descuento r_t ---las r_t , para toda t , son las *tasas spot* que buscamos---. Para resolver este problema usaremos el siguiente truco:

[1.] Tomaremos los primeros tres periodos $t=0,1,2$ e ignoraremos el resto.

[2.] El paso anterior implica que $t+1=2=T$ es el vencimiento.

Este hecho es importante porque por convención, toda deuda paga al vencimiento el último cupón más el principal. En consecuencia V_T es el valor nominal---*face value*---del instrumento, conocido también en $t=0$, es más: $V_0 = V_T$. Por ejemplo, los Certificados de la Tesorería mexicana, Cetes, tienen un valor nominal de $V_0 = V_T = \$10$ Mx---aunque cotizan a descuento, i.e. no pagan cupón.

Por su parte, C_T se obtiene de la curva de rendimientos.

[3.] Para evitar duplicar el pago del principal, (9.3) se reduce a:

$$V_{1,d} = \frac{V_T + C_2}{1 + r_{1,d}}$$

y

$$\begin{aligned} V_{1,u} &= \frac{V_T + C_2}{1 + r_{1,u}} \\ &= \frac{V_T + C_2}{1 + r_{1,d}e^{2v_1}} \end{aligned}$$

para $V_{1,d}$ y $V_{1,u}$. En el resultado anterior también usamos (9.1) con volatilidad estocástica.

[4.] Con base en los pasos anteriores tenemos. i) Por (9.3) y nuestro supuesto KWF:

$$\begin{aligned}
 V_0 &= \frac{1}{2} \left(\frac{V_{1,u} + C_1}{1 + r_0} \right) + \frac{1}{2} \left(\frac{V_{1,d} + C_1}{1 + r_0} \right) \\
 &= \frac{V_{1,u} + V_{1,d} + 2C_1}{2(1 + r_0)}
 \end{aligned}
 \tag{9.4}$$

ii) Sustituyendo $V_{1,d}$ y $V_{1,u}$ del paso 3 en (9.4):

$$\begin{aligned}
 V_0 (2(1 + r_0)) - 2C_1 &= V_{1,u} + V_{1,d} \\
 &= \frac{V_0 + C_2}{1 + r_{1,d}e^{2v_1}} + \frac{V_0 + C_2}{1 + r_{1,d}}
 \end{aligned}
 \tag{9.5}$$

La tasa spot en el tiempo $t=1$ se encuentra solucionando (9.5) con respecto a $r_{1,d}$. La solución tiene dos raíces: $r_{1,d}^+$ es la que nos interesa.

[5.] La tasa spot en el tiempo $t=2$ se encuentra desplazando el vencimiento a $t=3$ y repitiendo el procedimiento anterior. Con T en $t=3$, (5) toma la forma siguiente:

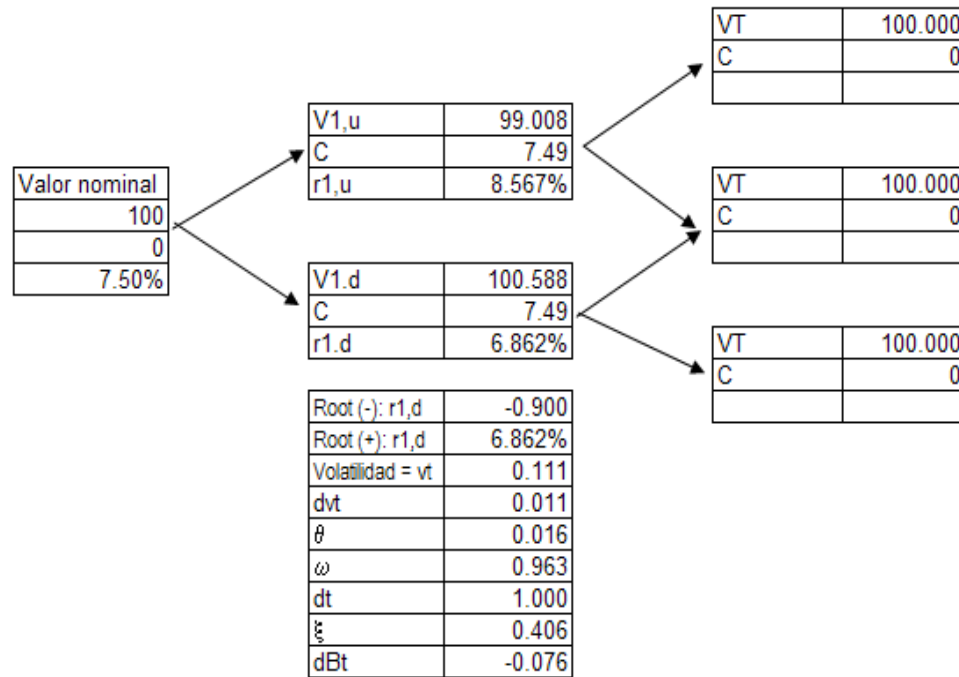
$$V_1 (2(1 + r_{1,d})) - 2C_2 = \frac{V_T + C_3}{1 + r_{2,dd}e^{2v_2}} + \frac{V_T + C_3}{1 + r_{2,dd}}.
 \tag{9.6}$$

Como en (9.5), la tasa spot se encuentra solucionando (9.6) con respecto a $r_{2,dd}$ y tomando la raíz positiva: $r_{2,dd}^+$.

[6.] Toda la serie de tasas spot se encuentra repitiendo los pasos 1 al 5 hasta el vencimiento.

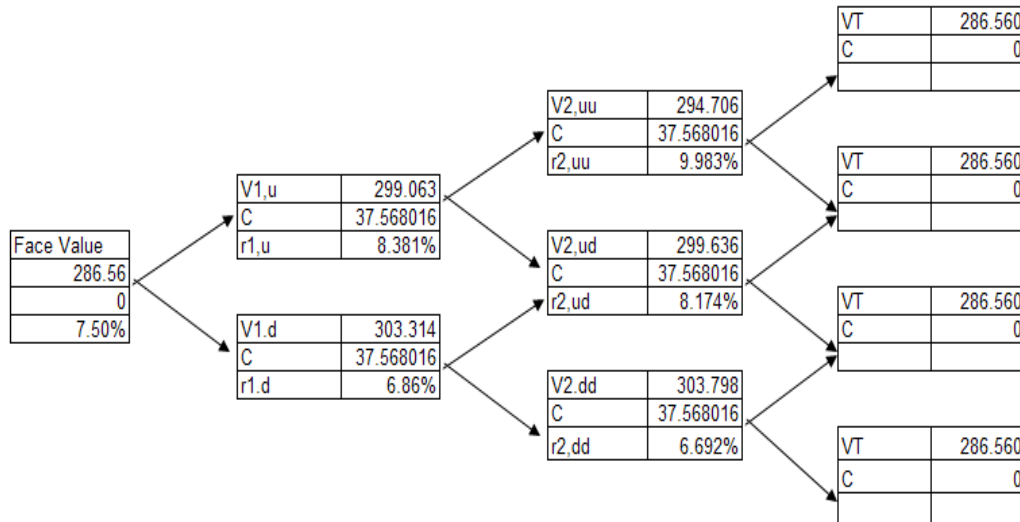
Por ejemplo, supongamos un bono que tiene valor nominal de \$100 Mx; y que los cupones que nos interesan son los observados el 4 de abril de 2008 en la Figura 9.4. Los cupones correspondientes a $t=0,1,2$ son respectivamente: la tasa de fondeo, 7.5%; el bono a diciembre 2008, 7.49%; y el bono a diciembre de 2009, 7.34%. La Figura 9.5 se obtiene siguiendo los pasos del 1 al 4:

Figura 9.5. Valor del bono de referencia a 2 años



La Figura 9.4 desplaza el vencimiento a $t = 3$ para encontrar la tasa spot en $t = 2$, como indica el paso 5:

Figura 9.6. Valor del bono de referencia a 3 años



9.3 Bono completo, bono comprable y valor del call

El bono completo se encuentra descontando cada flujo proyectado a las tasas spot, i.e. se sustituyen las r_t por las tasas spot en la ecuación (9.3). Se dice "completo" porque su valuación

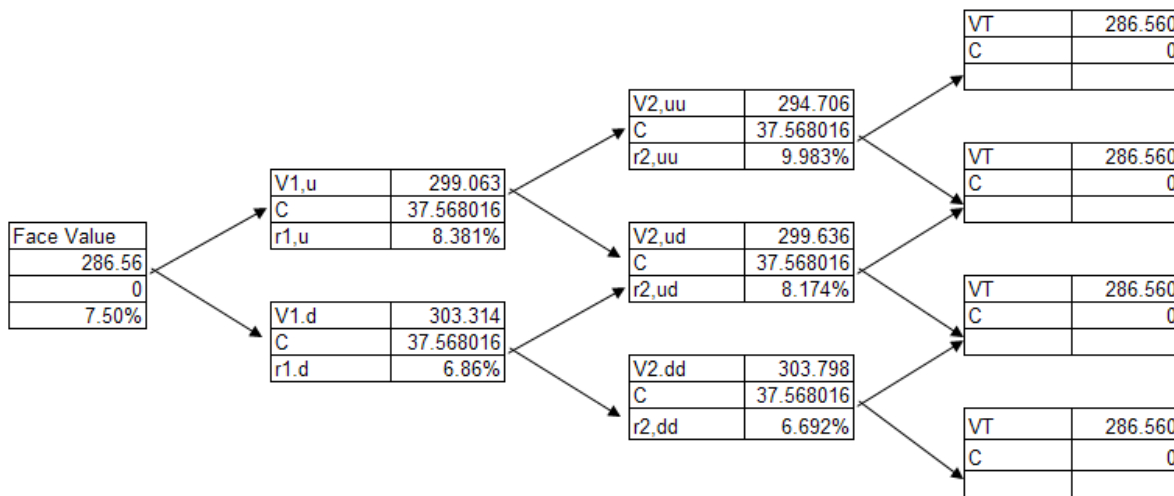
omite probabilidades de default y de prepago. Como tasa cupón se emplea el WAC y el WAM representa su madurez. WAC es la *tasa cupón promedio ponderado*, y WAM es la *madurez promedio ponderada* (por sus siglas en inglés *weighted average coupon* y *weighted average maturity* respectivamente). Las ponderaciones del WAC y del WAM están dadas por los balances pendientes de los proyectos. Por ejemplo, supongamos los datos siguientes:

Préstamo	Balance	Tasa del préstamo	Meses restantes	Ponderaciones
1	\$100,000	12.00%	300	11.11%
2	\$500,000	7.25%	260	55.56%
3	\$300,000	9.50%	190	33.33%

En este caso, el WAC = 8.53% y WAM = 241 meses.

La Figura 9.7 muestra el valor del bono titular A, con valor nominal \$286.56, WAC=13.11% y madurez $T=3$. Las tasas de descuento son las tasas spot que estimamos en la Figura 9.5.

Figura 9.7. Valor del bono titular A a 3 años, completo



El valor del bono comprable es el valor "justo" del bono dadas las características de las garantías y calificación subyacentes, y las condiciones del mercado: tasa de referencia, tasas spot y volatilidades. Este bono paga el cupón con base en la calificación que asigne la empresa calificadora al bono titular, i.e. el cupón del comprable es igual a la tasa de referencia más la *sobretasa---spread---*por calificación crediticia. El bono completo paga WAC con los intereses completos. La diferencia entre los cupones del bono completo y los del bono comprable son excedente para el pago de costos, cuotas, y demás primas.

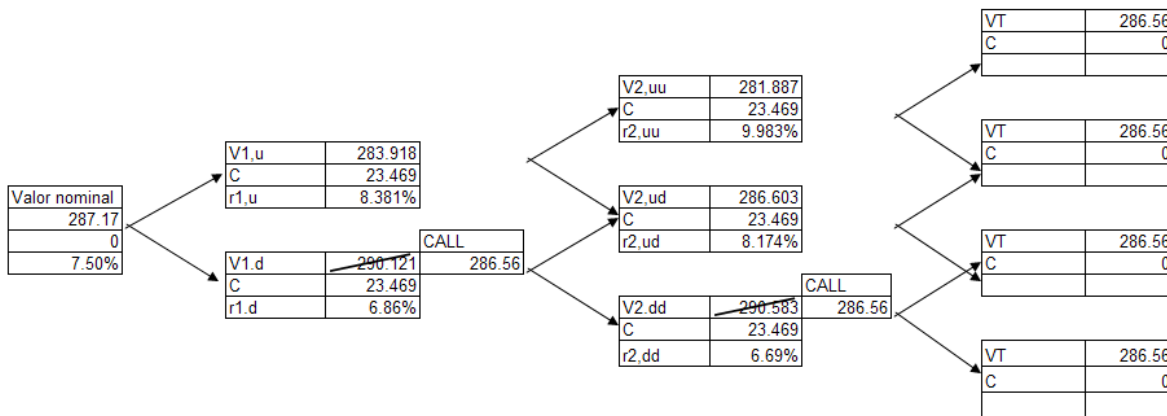
La valuación del bono comprable se hace de la misma manera que el bono completo con dos excepciones. Cuando la opción de compra se puede ejercer, el bono del valor en el nodo debe

cambiar para reflejar la pérdida de valor y se sustituye por el valor nominal o precio de ejercicio de la opción.

Como mencionamos arriba, Downing, Stanton y Wallace (2005) señalan que es importante incorporar el precio del subyacente del subyacente. Nosotros omitimos esta característica porque, como señalan Johnson y Cardarelli (2008) en su análisis de la importancia del precio de los hogares---el precio del subyacente del subyacente en el caso del mercado hipotecario---en el mecanismo de transmisión de la política monetaria, es sólo importante en el caso de mercados hipotecarios desarrollados; y prácticamente sólo aplica a ciertos mercados en EE.UU.---por ejemplo, en abril de 2008, la crisis hipotecaria que empezó en 2006 no había afectado al estado de Texas.

Con estas consideraciones, el valor del bono comprable correspondiente al bono titular A está representado en la Figura 9.8.

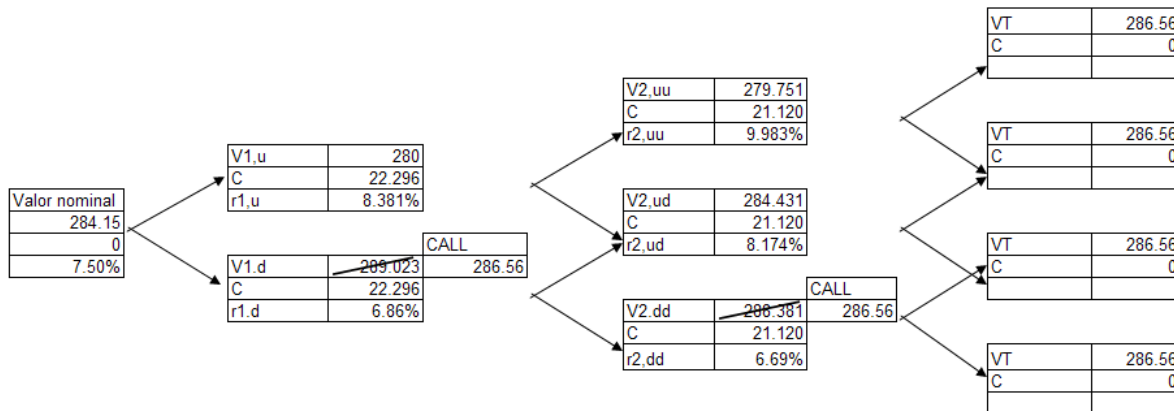
Figura 9.8. Valor del bono comprable a 3 años



Obsérvese que el cupón del bono titular A es 8.2%: una tasa de referencia en 7.49% más un *spread* por calificación crediticia igual a 0.7%.

El valor anterior es sin probabilidades de incumplimiento. Las probabilidades de incumplimiento reducen el valor del capital y por ende de los cupones que se pagan sobre el capital. La Figura 9 muestra el valor con probabilidades de default de 5% el primer año y 10% el segundo.

Figura 9.9. Valor del bono comprable a 3 años con probabilidades de default



9.4 Simulación Monte Carlo

El último componente de nuestro modelo está proporcionado por la capacidad para simular escenarios, i.e. la parte Monte Carlo del modelo. Ya hemos descrito distintos componentes aleatorios del modelo, en particular la volatilidad estocástica. Ahora describimos el comportamiento de la tasa de interés.

Presuponemos que la dinámica de la tasa de interés es del tipo Cox, Ingersoll y Ross (1985), CIR,

$$dr_t = (\kappa(\theta_r - r_t) - \eta r_t) dt + \phi_r \sqrt{r_t} dW_{r,t}, \quad dW_{r,t} \sim N(0, dt) \quad (9.7)$$

donde κ es la tasa de reversión a la tasa de interés promedio de largo plazo θ_r , η es el precio del riesgo de la tasa de interés y ϕ_r es la volatilidad de la tasa de interés; $W_{r,t}$ es un proceso de Wiener estándar. La tasa en el tiempo $t+1$ es igual a

$$r_t = r_{t-1} \cdot \exp(dr_t \cdot dt),$$

y como nuestros pagos son anuales, entonces $dt = 1$.

Estimamos los parámetros por máxima verosimilitud con datos de la tasa de fondeo entre enero de 1999 y marzo de 2008. Encontramos cambios estructurales en la serie usando la prueba CUSUM y CUSUM de cuadrados. La serie libre de cambios estructurales corresponde al periodo marzo de 2006 a marzo de 2008. También probamos aplicando términos rezagados a (7), y encontramos que usándolo en el segundo término minimiza el Criterio de Información Akaike, AIC. Con base en lo anterior, los parámetros estimados son:

$$\Delta r_t = (-0.5018(0.0708 - r_t) - 1.0001r_{t-1}) + 0.2662\sqrt{r_t}\Delta\varepsilon_{r,t}, \quad \Delta\varepsilon_{r,t} \sim N(0, \Delta t = 1).$$

Una vez estimados los parámetros, los mantenemos constantes, aunque en la realidad, conviene recalibrar cada semana o cada dos semanas dependiendo de la volatilidad en el mercado. La Figura 9.10 presenta propiedades estadísticas adicionales:

Figura 9.10. Coeficientes CIR

Variable dependiente: dr_t		
	Coeficiente	p-value
κ	-0.5018	0.0000
θ	0.0708	
η	1.0001	0.0000
ϕ	0.2661	0.0000
dt	1	

R^2 Ajustado = 0.9999
AIC=-20.4013

Obsérvese que todos los parámetros son significativos al 95% de confianza estadística.

Por su parte, para la estimación de la ecuación de la volatilidad estocástica anual usamos el mismo periodo sin cambios estructurales y también probamos incorporando rezagos. Usar rezago en el segundo término es el que minimiza AIC en este caso. La ecuación estimada por máxima verosimilitud es:

$$\Delta v_t = -0.1023(0.0039 - v_t) - 0.0769v_{t-1}\Delta\epsilon_{v,t}, \quad \Delta\epsilon_{v,t} \sim \text{Normal}(0, 1).$$

Los parámetros para la simulación son:

Figura 9.11. Coeficientes GARCH

Variable dependiente: dv_t		
	Coeficiente	p-value
θ	-0.1023	0.0919
ω	0.0039	
ξ	-0.0769	0.0529
dt	1	

R^2 Ajustado = 0.1277
AIC=-11.3799

Al igual que en el caso de la tasa,

$$v_t = v_{t-1} \cdot \exp(dv_t \cdot dt).$$

Con base en lo anterior, ahora reestimamos el valor justo de nuestro bono titular para encontrar escenarios mínimo y máximo con un número predeterminado de escenarios. La Figura 9.12 muestra los resultados con 100 escenarios:

Figura 9.12. Precios del comprable y posturas con 100 escenarios

	Bono comprable A	Postura
Máx	\$220.97	\$66.29
Mín	\$207.98	\$62.40
Promedio	\$214.32	\$64.30
Desviación estándar	\$3.99	\$1.20

Tanto la serie de precios del bono A como las posturas se distribuyen de manera Normal: la probabilidad de la prueba de Normalidad Jarque-Bera de la primera es 0.8555 y la probabilidad de la serie de posturas es 0.8539. Esto implica que asintóticamente el valor "justo" del comprable es el promedio \$214.32.

A manera de comparación mencionamos que el valor actual neto, VAN, de la canasta de hipotecas de la Figura 9.1, estimada usando la ecuación del VAN

$$VAN = -k + \sum_{t=0}^n \frac{f_t}{(1+i)^t}$$

es:

$$\begin{aligned} VAN &= -\$500 + \sum_{t=0}^{14} \frac{\$65.55}{(1+7.50\%)^t} + \frac{\$65.55 + \$500}{(1+7.50\%)^{15}} \\ &= -\$500 + \$622.01 + \$191.14 \\ &= -\$500 + \$813.15 \\ &= \$313.15. \end{aligned}$$

Donde $k=\$500$ millones de USDs es la cantidad de principal pendiente; la tasa de descuento de 7.5% es la tasa de fondeo observada el 4 de abril de 2008; tomamos el vencimiento como $n=15$ años, un aproximado del WAM para que este cálculo sea comparable con el de nuestras estimaciones anteriores; y los intereses mensuales de \$65.55 son el cupón---WAC=13.11%---sobre la deuda sobresaliente.

Es decir, el criterio del VAN sugeriría un valor de \$313.15 por toda la cartera. Ahora, es importante reflexionar la veracidad del VAN que acabamos de estimar: i) k , nuestra inversión inicial, es el equivalente de la postura. Empleando el criterio anterior, debería ser 30% del precio: $30\%(\$813.15)=\243.95 , en lugar de los \$500. ii) Los intereses \$65.55 no consideran el hecho de que no toda la cartera de préstamos es productiva, están sobreestimados. Además son constantes: qué sucede con el prepago. iii) Empleamos una tasa de descuento constante para toda la vida de los préstamos: ¿cuál es la probabilidad de que las tasas permanezcan en ese nivel en los próximos meses o años? Si la tasa de descuento sube el VAN también estará sobreestimado. iv) ¿Qué hay del riesgo de cambios en el precio del subyacente? A la luz de estas preguntas, que tan "justo" es el VAN. Obviamente, el VAN está totalmente superado por la naturaleza del mercado y resulta inútil para valuar este tipo de instrumentos, por eso es tan importante un modelo como el que proponemos.

Por último, una deficiencia de nuestro método es que estimar los 100 escenarios es relativamente lento. Otras posibilidades consistirían en aleatorizar las probabilidades de incumplimiento, aunque esto haría aún más lento el proceso de simulación. La capacidad de simulación incrementa la precisión del modelo.

9.5 Conclusiones

Como mencionamos en la Introducción, el objetivo de este trabajo fue proponer un modelo de valuación de alta precisión que además tiene gran flexibilidad para incluir características particulares de cada activo subyacente. El resultado es el modelo de valuación de opciones de Monte Carlo con volatilidad estocástica que tiene gran capacidad para incorporar todo tipo de riesgos:

- riesgo de movimientos en la tasa de referencia y en la curva de rendimientos;
- riesgo de impago (*default*);
- riesgo de prepago total o parcial (en inglés, *prepayment* y *curtailment* respectivamente);
- riesgo de cambios en el precio del subyacente; y
- riesgo de cambios en el precio del subyacente del subyacente---componente fundamental del bono respaldado por activos, BORA---un BORHI es un caso particular de un BORA--..

La idea fundamental es que necesitamos tres árboles binomiales: uno para estimar las tasas spot. El segundo árbol sirve para valorar el bono completo. El tercer y último árbol binomial sirve para valorar el bono comprable. Explicamos cada árbol de manera detallada.

Por último deseamos resaltar que el modelo permite variaciones importantes, por ejemplo, aleatorizar las probabilidades de default. Sin embargo, toda ganancia en precisión tiene un costo de implementación.

9.6 Bibliografía

Cox, J.C., J.E. Ingersoll, Jr. and S.A. Ross. (1985). "A theory of the term structure of interest rates." *Econometrica*, 53(2), pp. 385-407.

Cox, J.C., Ross, S.A. y M. Rubinstein. (1979). "Option pricing: a simplified approach." *Journal of Financial Economics*, 7(3), pp. 229-263.

Downing, C.T., R. Stanton y N. Wallace. (2005). "An empirical test of two-factor mortgage valuation model: how much do house prices matter?" *Real Estate Economics*, 33(4), pp. 681–710.

Johnson, S. y R. Cardarelli. (2008). "World economic outlook: housing and the business cycle." International Monetary Fund. Abril.

Kalotay, A.J., G.O. Williams y F.J. Fabozzi. (1993). "A model for valuing bonds with embedded options." *Financial Analysts Journal*. Mayo/Junio, pp. 35-46.

Capítulo 10

Análisis de la variación de precios mediante caminatas aleatorias: Weierstrass y fluctuaciones libres de escala

Francisco Javier Almaguer

Universidad Autónoma de Nuevo León, FCFM

Homero de la Fuente

Universidad Autónoma de Nuevo León, FCFM

Javier Morales

Universidad Autónoma de Nuevo León, FIME

Jonás Velasco

Universidad Autónoma de Nuevo León, FIME

Roberto Soto

Universidad Autónoma de Nuevo León, FCFM

Resumen

En este capítulo se da un breve repaso a la teoría de Black-Scholes del movimiento Browniano geométrico, la primera aplicación exitosa de la teoría del movimiento Browniano fuera del ámbito de la física. Se presentan algunos resultados ya conocidos de la caminata de Weierstrass y se incorpora, como propuesta, al movimiento Browniano geométrico para modelar situaciones de turbulencia en mercados financieros. Se espera que este nuevo modelo, además de proporcionar estimaciones de las fluctuaciones en las variaciones de los precios de acciones en escenarios ordinarios, “normales”, pueda incluso llegar a dar estimaciones cualitativas, globales, de la dinámica de precios durante etapas turbulentas de gran volatilidad. Se presentan y discuten algunos resultados de las simulaciones de caminatas aleatorias de Weierstrass, específicamente, la

existencia de una transición de fase orden-desorden en las cercanías de la curva crítica, $z^2 = \beta$, en el espacio de parámetros de la caminata. Cuando se presenta la transición, el desplazamiento cuadrático medio diverge y la distribución de probabilidad para el desplazamiento tiene un decaimiento en ley de potencias típico de las distribuciones estables con invariancia de escala. La hipótesis de trabajo subyacente, el hilo conductor de todo el capítulo, es considerar que la emergencia macroscópica de un fenómeno global, como el de las variaciones en los precios de acciones y opciones, es resultado de los constantes flujos e intercambios de información y toma de decisión de una gran cantidad de agentes económicos involucrados, y que dicho fenómeno colectivo puede ser modelado mediante procesos aleatorios discretos que evolucionan en el tiempo. Utilizando como punto de partida una aproximación de campo medio, los cambios discretos en la variación de los precios son modelados mediante caminatas aleatorias markovianas.

10.1 Introducción

En esta primera sección se da un breve resumen de las propiedades más importantes de los procesos estocásticos de Wiener o movimiento Browniano, del movimiento Browniano geométrico y de su conexión con un modelo simple de caminata aleatoria. En la segunda sección se presenta la metodología empleada en el trabajo y se introducen las caminatas de Weierstrass en el estudio de la valuación de opciones y las caídas repentinas de los índices de precios en los mercados financieros. En la última sección se presenta un análisis de los resultados y las conclusiones generales del material expuesto en el capítulo.

En un artículo ya clásico en el tema (Black *et al.*, 1973) Black y Scholes aplican de manera exitosa al ámbito de la economía las ideas de Bachelier (Cortault, 2000) sobre las variaciones de los precios de acciones y el marco teórico conceptual del movimiento Browniano de Einstein de 1905 (Einstein, 1905). Desde entonces, estos trabajos han inspirado una extensa literatura y debate sobre la aplicación de la teoría de sistemas complejos y fenómenos colectivos (Sornette, 2004), (Mantenga, *et al.*, 2007), de las teorías de campo medio de la física estadística (Bouchad, *et al.*, 2009) (Mahnke, 2009) y modelos de caminatas aleatorias (Wiersema, Ubbo F. 2008), (Mikosch, 1998) al estudio de las fluctuaciones de los precios de acciones y valuación de opciones, (Osborne, 1962), (Sottinen, 2001), (Laloux, 1999) y (Gerber, 2004). Sin embargo, la literatura publicada sobre las propiedades de la caminata de Weierstrass, u otros procesos aleatorios similares, y sus potenciales aplicaciones en economía es menos extensa. Un par de trabajos importantes que han inspirado la preparación del presente capítulo es el libro de Paul y Baschnagel, (Paul *et al.*, 1999) y los trabajos de Sornette sobre la predicción de derrumbes financieros (Sornette, *et al.*, 1999).

10.1.1 Movimiento Browniano

El movimiento Browniano, en cuanto modelo aleatorio inspirado en las propiedades físicas

globales del movimiento térmico de partículas suspendidas en un líquido, es un proceso estocástico $\{B(t); t \geq 0\}$ que satisface las siguientes condiciones (Wiersema, 1999):

- Para cualquier realización del proceso, al tiempo inicial cero el valor de la variable aleatoria es $B(0)=0$; los incrementos de B son continuos, esto es, $\Delta B(s)=B(t+s)-B(t)$ es un número real para cualesquier $t, s \geq 0$.
- Dada una secuencia ordenada de tiempos $t_1 < t_2 \leq t_3 \leq t_4$, los incrementos $\Delta B(t_2-t_1)$ y $\Delta B(t_4-t_3)$ son variables aleatorias independientes. Los valores del incremento aleatorio $\Delta B(s)$, se distribuyen normalmente con media cero, $E[\Delta B(s)]=0$ y varianza $Var(\Delta B(s))=s$, esto es, la probabilidad del evento $F(a)=P\{\Delta B(s) \leq a\}$ es

$$F(l) = \frac{1}{\sqrt{2\pi s}} \int_{-\infty}^l e^{-\frac{1}{2}(\frac{l}{\sqrt{s}})^2} dl \quad (10.1)$$

10.1.2 Movimiento Browniano geométrico

Partiendo del movimiento Browniano se define un proceso estocástico para el valor de una acción al tiempo t , $\{S(t); t \geq 0\}$, conocido como movimiento Browniano aritmético, mediante la ecuación de evolución diferencial estocástica

$$dS(t) = \mu dt + \sigma dB(t) \quad (10.2)$$

Es decir, durante el incremento diferencial de tiempo dt , el valor correspondiente del incremento diferencial de la acción se considera que es resultado de una superposición de un término determinista lineal, con tasa constante μ , y un término de *ruido* que representa el campo medio efectivo de los agentes económicos mediante fluctuaciones Brownianas de intensidad σ .

La solución de la ecuación anterior se escribe como

$$S(t) = S_0 + \mu t + \sigma B(t) \quad (10.3)$$

donde S_0 es el valor de $S(t)$ al tiempo inicial cero; los coeficientes constantes de “arrastre” μ y de “ruido” (difusión) $\sigma > 0$ están relacionados con las tasas del crecimiento lineal del valor esperado $E[S(t)]=S_0 + \mu t$ y la varianza $Var(S(t))=\sigma^2 t$ del proceso aritmético. Ver las Fig. 10.1 y 10.2 para una simulación del movimiento aritmético.

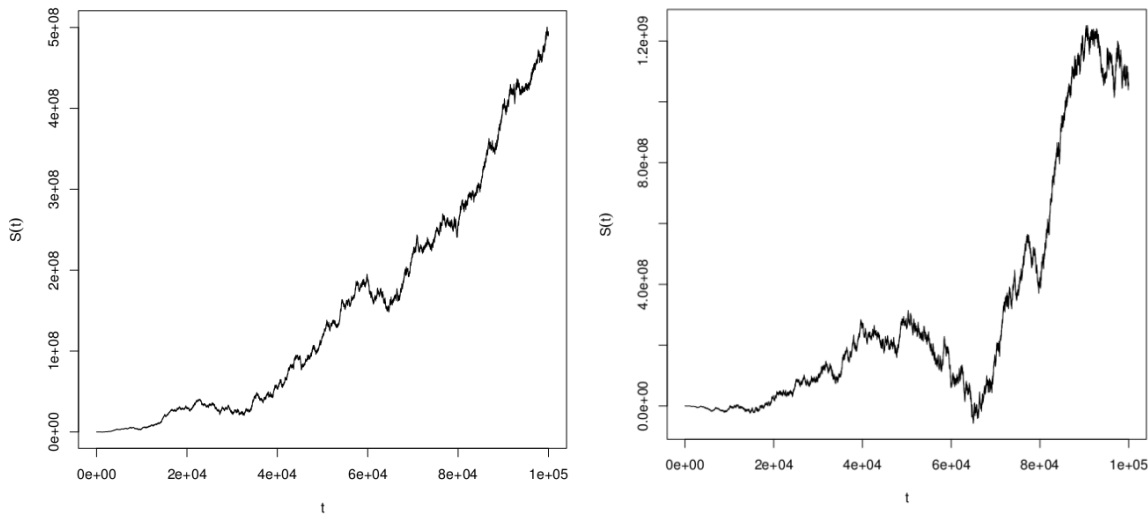


Fig. 10.1 y 10.2: Una simulación del movimiento Browniano aritmético para dos escenarios de volatilidad distintos, con $S_0 = 1$, $\mu = 0.1$ y, en ambos casos, $t = 10^5$ unidades de tiempo o “pasos”. En ellas se puede observar que cuanto mayor sea la “intensidad del ruido” σ mayor es la posibilidad de que $S(t)$ tome valores negativos. En la Fig. 10.1, $\sigma = 5$, mientras que en la Fig. 10.2, $\sigma = 10$.

Dado que el valor de la acción $S(t)$ en el modelo aritmético puede tomar valores negativos, debido a la presencia del movimiento Browniano $B(t)$, un modelo más realista para estudiar las variaciones de los precios de acciones es el movimiento Browniano geométrico (Black y Scholes, 1973): un movimiento Browniano aritmético para el cambio relativo $\frac{dS(t)}{S(t)}$ del precio de una acción definido por

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dB(t) \tag{10.4}$$

Usando la teoría de Itô del cálculo estocástico (Mikosch, 1998) la solución de la ecuación diferencial estocástica para el movimiento Browniano geométrico (10.1.2.3) se escribe como

$$S(t) = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B(t)} \tag{10.5}$$

donde el valor de la acción $S(t)$ es ahora estrictamente positivo. La media y la varianza en este caso están dadas por (Mahnke, R. *et al.*, 2009) $E[S(t)] = S_0 e^{\mu t}$ y $\sigma^2(s) = S_0^2 e^{\mu t} (e^{\sigma^2 t} - 1)$. Comparando las soluciones para el movimiento Browniano aritmético y geométrico, se observa que $\ln(S(t)/S_0)$ representa, esencialmente, el mismo proceso aritmético (10.1.2.1), de tal manera que los valores de $Y = \ln(S(t))$ siguen una distribución normal con media $E[Y(t)] = \ln(S_0) + (\mu - \frac{\sigma^2}{2})t$ y varianza $Var(Y(t)) = \sigma^2 t$.

La acumulada de $Y(t)$, $P\{Y(t) \leq \ell\}$ es dada por (10.1.1.1) al reemplazar $\ell = y - E[Y(t)]$ y $s = \sigma^2 t$, ver las figuras 10.3 y 10.4.

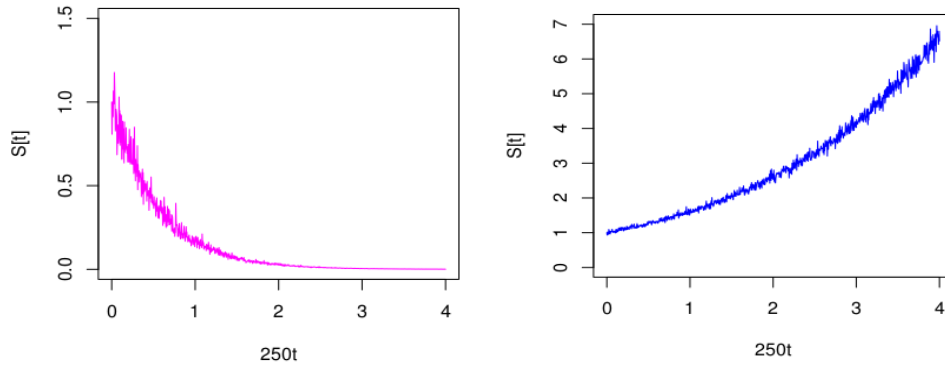


Fig. 10.3 y 10.4: Dos realizaciones del movimiento Browniano geométrico cuando $\mu < \sigma^2/2$ y $\mu > \sigma^2/2$ para 10^3 pasos o unidades de tiempo. En la Fig. 10.1.2.3, $\mu=0.2$ y $\sigma=2$, mientras que en la Fig. 10.4, $\mu=0.6$ y $\sigma=0.5$, para ver una simulación del Movimiento Browniano geométrico.

Si bien el movimiento Browniano geométrico es un modelo estándar relativamente exitoso para estudiar las variaciones en los precios de opciones e índices de precios, no es del todo apropiado para modelar situaciones turbulentas con valores divergentes o abruptamente divergentes en la volatilidad, cuando es altamente probable que se presente un derrumbe financiero. Para abordar este tipo de situaciones, una alternativa es reemplazar las fluctuaciones Brownianas del movimiento Browniano geométrico por un proceso estocástico que permita incorporar valores divergentes en el segundo momento de $S(t)$. La propuesta es que un movimiento geométrico que pueda incorporar las propiedades fractales de autosimilaridad, de una búsqueda aleatoria libre de escala como las que presenta la caminata de Weierstrass, debería modelar mucho mejor, al menos cualitativamente, situaciones de mayor inestabilidad y con ello lograr mejores predicciones durante la presencia de cambios bruscos en la dinámica de precios; dichos cambios repentinos pueden también abordarse desde el punto de vista teórico-experimental de las transiciones de fase orden-desorden de la física estadística.

10.1.3 Caminata aleatoria simple

Existe una caminata aleatoria simple discreta unidimensional la cual, en cierto límite, presenta todas las propiedades del movimiento Browniano continuo. Esto ocurre cuando el número de pasos de la caminata es lo suficientemente grande, $n \gg 1$, o, lo que es lo mismo, cuando haya transcurrido el tiempo suficiente, desde el inicio del proceso, para que los efectos aleatorios se “estabilicen” y haga su aparición el atractor Gaussiano, un resultado esperado por la ley de los grandes números y el teorema de límite central. Se tiene entonces que la distribución binomial de la caminata simple discreta se aproxima a la distribución normal cuando el número de pasos (ensayos) independientes $n \rightarrow \infty$. En este llamado límite difusivo la ecuación maestra de la evolución de la caminata aleatoria simple se transforma en la ecuación de difusión, cuya solución es la densidad de probabilidad Gaussiana, la derivada de la acumulada (10.1).

En forma más específica, considere una partícula que visita los sitios de una red unidimensional. Los sitios están igualmente espaciados una distancia unitaria. En cada paso o unidad de tiempo, partiendo del origen al tiempo cero, la partícula con probabilidad a se desplaza al sitio de la derecha o bien al de la izquierda con probabilidad $b=1-a$. Si $a \neq b$, el comportamiento global que predice el teorema de limite central da como resultado un “transporte” o flujo efectivo de partículas en una dirección preferencial (a la derecha si $a > b$ o a la izquierda en caso contrario) con un término de “arrastré” como el que aparece en el movimiento Browniano geométrico. En un juego no equilibrado, dicho flujo efectivo representa el incremento en la ganancia neta del jugador favorecido. Cuando $a = b = 1/2$ se dice que la caminata aleatoria es simétrica, el flujo efectivo de partículas es cero (una ganancia efectiva de cero en un juego equilibrado), no hay termino determinista o de “arrastré” en la ecuación diferencial macroscópica, y se tiene una dinámica meramente ruidosa o difusiva. En la Fig. 10.5 se muestra una caminata aleatoria simple simétrica con $a = b = 1/2$.

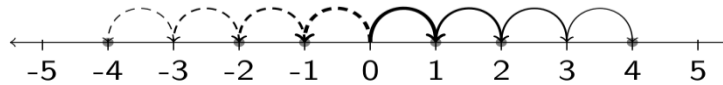


Fig. 10.5. En cada paso la partícula solo va al sitio vecino más cercano con igual probabilidad. Después de n pasos la probabilidad de localizar a la partícula en la posición k es

$$P_n(k) = \sum_{r=0}^n \frac{n!}{r!(n-r)!} a^r b^{n-r} \delta[k - (2r - n)] \tag{10.6}$$

donde $\delta(\cdot)$ representa una delta de Dirac, y r el número de pasos a la derecha. Los estados de posición k accesibles a la partícula, compatibles con el numero de pasos n , siempre satisfacen la relación $r=(k+n)/2$. Así, para una cantidad de pasos dada n y un valor de r fijo, habrá sitios prohibidos donde la partícula no podrá ser localizada. La Fig. 10.6 muestra los perfiles típicos difusivos de una simulación de 10^2 caminatas aleatorias con $n=10^4$ pasos cada una. Observe como el centro de la distribución de las trayectorias se localiza en cero, $\mu=0$, y que el grado de dispersión promedio con respecto al centro es del orden de $\sigma \sim n^{(1/2)}=10^2$.

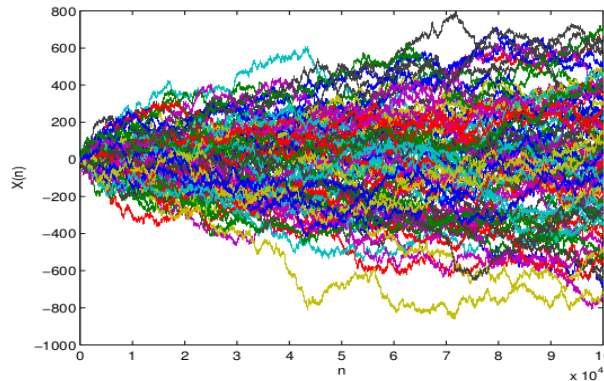


Fig. 10.6. Perfiles típicos difusivos de una simulación de 10^2 caminatas aleatorias con $n=10^4$ pasos cada una.

Observe como el centro de la distribución de las trayectorias se localiza en cero, $\mu=0$, y que el grado de promedio con respecto al centro es del orden de $\sigma \sim n^{1/2} = 10^2$.

10.2 Marco metodológico

En esta sección se presenta la caminata de Weierstrass, sus propiedades más interesantes en relación con la valuación de acciones y la posibilidad de utilizarla en situaciones de alta volatilidad. Se comenta brevemente también la manera de generar números aleatorios de Weierstrass para la visualización gráfica de exploraciones aleatorias en el plano.

10.2.1 Caminata de Weierstrass

A diferencia de la caminata simple, la caminata de Weierstrass conecta cualquier sitio de la red con todos los demás mediante un parámetro $z > 1$ que está relacionado con el tamaño de los saltos independientes. La probabilidad de dar esos saltos a su vez es regulada por medio del parámetro “térmico” $\beta > 1$. Concretamente, la probabilidad de hacer el desplazamiento $\ell = \pm z^k$ en un solo paso para $k=0,1,2,\dots$, es dada por la distribución de probabilidad (Paul *et al.*, 1999)

$$P(\ell) \equiv P(k; \beta) = \frac{a}{\beta^k} \quad (10.7)$$

siendo a la probabilidad de ir a los sitios más cercanos, el de la derecha o el de la izquierda, ($k=0$), ver la Fig. 10.7

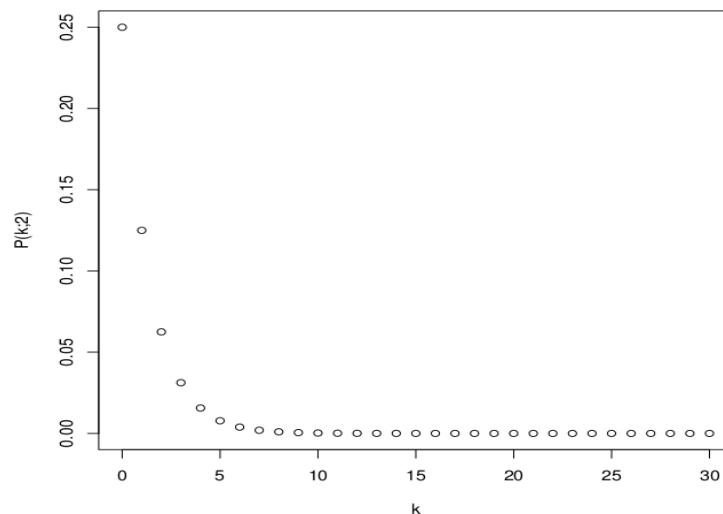


Fig. 10.7. El decaimiento de la probabilidad $P(k; \beta) = a/\beta^k$ de ir al nodo k de la red para $a=1/4$ y $\beta=2$.

Observe el rápido decaimiento de la probabilidad cuando $k > 1$.

La distribución de Weierstrass para un desplazamiento arbitrario $\ell = z^j$ también se puede escribir como, ver la Fig. 10.8,

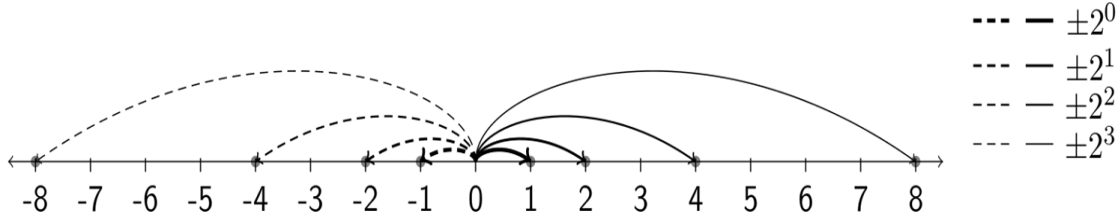


Fig. 10.8: Una caminata aleatoria de Weierstrass con $z = 2$ y $k=0, 1, 2, 3$.

$$P(l) = \sum_{j=0}^{\infty} \frac{a}{\beta^j} [\delta(l - z^j) + \delta(l + z^j)] \quad (10.8)$$

La normalización de la distribución de Weierstrass permite obtener el valor de a como función del parámetro β

$$a(\beta) = \frac{1}{2} \left(1 - \frac{1}{\beta} \right) \quad (10.9)$$

Según la fórmula anterior, cuando $\beta \rightarrow \infty$, la probabilidad de transición al nodo más cercano se aproxima al valor $a \rightarrow 1/2$ de tal manera que el término que más contribuye a la suma en la distribución de Weierstrass es el que corresponde a $j = 0$ y la probabilidad se puede aproximar mediante

$$P(l) \approx a[\delta(l - 1) + \delta(l + 1)] \quad (10.10)$$

esto es, la caminata de Weierstrass se transforma, en $z = 1$, en una caminata aleatoria simple simétrica $a = b = 1/2$ con distribución del tamaño de pasos dada por la distribución binomial (10.6) para $n = 1$.

Utilizando (10.8), el valor esperado del desplazamiento es $\langle l \rangle = 0$, un reflejo de la simetría de la caminata. Por otro lado, la magnitud de las fluctuaciones en el desplazamiento es dada directamente por el segundo momento de (10.6)

$$\langle l^2 \rangle = \left(1 - \frac{1}{\beta} \right) \sum_{j=0}^{\infty} \left(\frac{z^2}{\beta} \right)^j \quad (10.11)$$

Observe que si $z^2 < \beta$, el valor finito de la raíz cuadrática media es

$$\lambda = \sqrt{\langle l^2 \rangle} = \sqrt{\frac{1 - \frac{1}{\beta}}{1 - \frac{z^2}{\beta}}} \quad (10.12)$$

Así, cuando $\beta \gg 1$ y $z^2/\beta \ll 1$, la longitud característica de la caminata es del orden de la distancia entre los nodos de la red, $\lambda \approx 1$, resultado que está en acuerdo con una caminata simple simétrica para la cual $E[x] = 0$ y $E[x^2] = 1$, siendo x el desplazamiento en un solo paso (± 1). Un resultado similar se obtiene aún para $z^2 < \beta$ con $z \approx \beta^{(1/2)} \approx 1$, justo en las vecindades de la curva crítica $z^2 = \beta$ del espacio de parámetros de la caminata.

Una de las principales motivaciones para estudiar la caminata de Weierstrass es su versatilidad en la exploración del espacio de búsqueda. Si se hace un barrido del espacio de parámetros, la caminata tiene el poder de realizar transiciones de fase y ubicarse en distintos regímenes alrededor de la curva crítica $z^2 = \beta$, ver las figuras 10.9, 10.10 y 10.11.

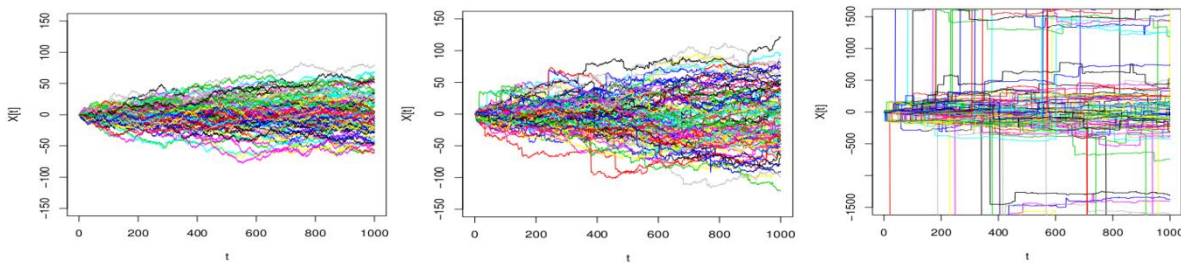
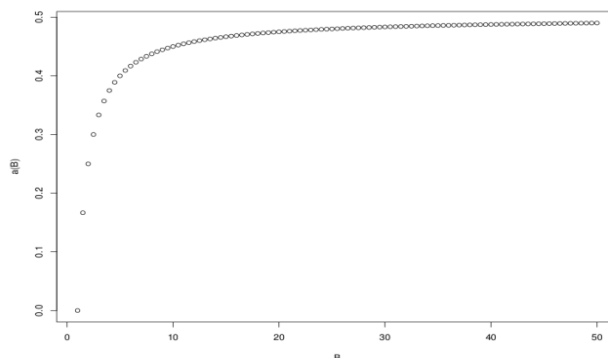


Fig. 10.9, 10.10 y 10.11: Una transición de fase orden-desorden desde el régimen cuasi-Browniano, al régimen libre de escala: Fig. 10.9, $z^2 < \beta$, con $\beta = 10$ y $z^2 = 1.0201$; Fig. 10.10, $z^2 = \beta = 9$; Fig. 10.11, $z^2 > \beta$, con $\beta = 10$ y $z^2 = 144$.

Las gráficas anteriores se obtuvieron utilizando un generador de números aleatorios basado en la función de distribución acumulada de Weierstrass.

En las Fig. 10.12 y 10.13 se han graficado la constante de normalización $a = (1-1/\beta)/2$ como función de β y también como función del parámetro térmico $T = 1/\beta$, respectivamente (ver la discusión de más abajo).



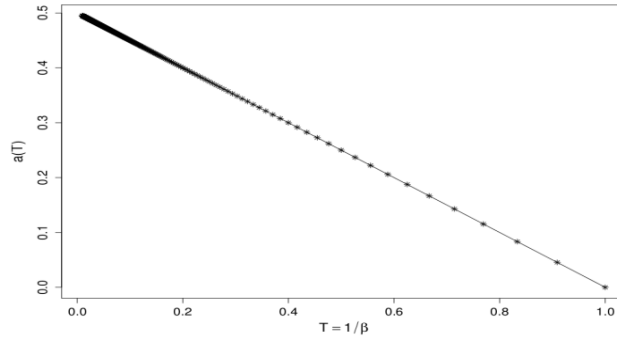


Fig. 10.12 y 10.13: Probabilidad de transición a los vecinos más cercanos como función de β , Fig. 10.2.1.6, y como función del parámetro térmico $a=(1-T)/2$, Fig. 10.13

Por otro lado, la fórmula que se obtiene al reemplazar la constante de normalización en la distribución de Weierstrass, $P(\ell) \equiv P(k; \beta) = (1-1/\beta)/(2\beta^k)$ proporciona información cualitativa valiosa acerca de la existencia de la “transición de fase”, ya observada en las simulaciones, subyacente en la caminata de Weierstrass.

En el límite asintótico $\beta \rightarrow \infty$ se castiga mucho la posibilidad de dar saltos grandes y la probabilidad de dar un salto de tamaño $\ell = z^k$ decae como una ley de potencias, $P(k; \beta) \sim \beta^{-(k/2)}$. La probabilidad de ir a una posición fija, determinada por un valor de k , es máxima para el valor crítico $\beta_c = 1+1/k$, esto es, la manera más eficiente de alcanzar nodos muy alejados de la red es con un valor de β próximo a uno. La probabilidad máxima de alcanzar esos nodos decae como $P_m(k; \beta_c) \approx e^{-k}/2k$, para $k \gg 1$. En la Fig. 10.14 se muestra el comportamiento de la probabilidad de transición al nodo k como función de β .

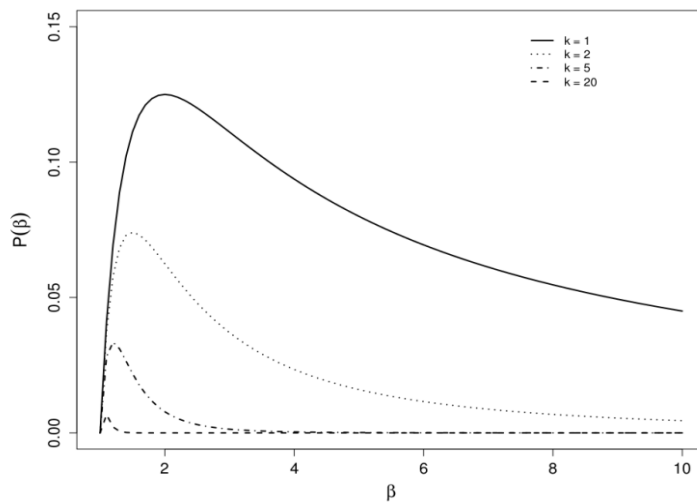


Fig. 10.14. Observe como el máximo de la probabilidad, $P_m = e^{-2k}/2k$ se desplaza hacia $\beta_c=1$ para nodos cada vez más alejados del origen, $k \gg 1$.

La transición de fase orden-desorden en la exploración del espacio de la caminata no es explícita en la distribución $P(k; \beta)$. La naturaleza fractal y el origen de la invarianza de escala de la caminata de Weierstrass ha de buscarse en el análisis del segundo momento. Un estudio detallado de la expresión para el segundo momento (10.11) muestra la existencia de dos regiones en el espacio de parámetros, separadas por la curva crítica $z^2 = \beta$; una cuasi-Browniana, $z^2 < \beta$ y otra superdifusiva $z^2 > \beta$. En cada región, las propiedades de las variaciones de los incrementos de la variable son radicalmente diferentes.

A continuación se da un breve resumen de las dinámicas cuasi-Brownianas, y superdifusivas de la caminata de Weierstrass.

- **Búsqueda cuasi-Browniano, $z^2 < \beta$:** En este caso la caminata lleva a cabo una exploración “ordenada” con una raíz cuadrática media finita dada por $\lambda \approx \sqrt{(1 - 1/\beta)}$, ver la ecuación (10.12). Si también se cumple que $\beta \gg 1$, la caminata es presa de un “potencial Browniano” intenso y explora su espacio de manera aproximadamente difusiva, con un patrón de búsqueda muy semejante al de la Fig. 10.6. Con probabilidad casi uno solo los primeros vecinos son visitados, de ahí que los valores de la acumulada discreta sean dados prácticamente por las probabilidades asociadas con los nodos vecinos más cercanos, $k=0$. Por otro lado, cuando $\beta \approx 1$ se puede escribir $\beta = 1+e$, con $e \ll 1$. En este caso, en el “campo cercano” cuando $ke \ll 1$, la probabilidad es aproximadamente constante $P(k; e) \sim e/2$ y, en el “campo lejano” cuando $ke \gg 1$, decae como $P(k; e) \sim 1/(2k)$. Las dos expresiones coinciden cuando $e = 1/k$. En ambos casos el desplazamiento cuadrático medio es $\lambda = (e/(1+e-z^2))^{(1/2)}$ el cual permanece finito siempre y cuando $z^2 < 1+e$, precisamente la condición para un régimen con escala finita de exploración. Dado que también $z^2 > 1$, se puede elegir, por ejemplo, $z^2 = 1 + e/\sigma$ con la restricción de que $\sigma > 1$. Con la elección anterior se tiene que $\lambda = 1/(1-1/\sigma)^{(1/2)}$, la cual crece sin límite cuando $\sigma \rightarrow 1$, esto nos traslada al régimen siguiente.
- **Curva crítica $z^2 = \beta$:** En este régimen la caminata, al carecer de una escala determinada, puesto que $\lambda \rightarrow \infty$, lleva a cabo una exploración “desordenada”. Sin embargo una disminución en el valor de z^2 o un aumento en el valor de β , por pequeños que sean, sacan al sistema de la frontera crítica y transforman radicalmente el tipo de exploración; se dice entonces que la caminata ha experimentado una transición de fase orden-desorden, pasando de una exploración o búsqueda “desordenada” fractal o superdifusiva libre de escala a una exploración “ordenada” con una escala finita, la cual adquiere carácter cuasi-Browniano o difusivo cuando $z^2 < \beta$.
- **Régimen superdifusivo, $z^2 > \beta$:** La caminata se encuentra en un régimen fractal libre de escala, $\lambda \rightarrow \infty$. Cuando $\beta \approx 1$, para nodos muy alejados de la red, $k \gg 1$, la probabilidad es prácticamente constante $P(k; \beta) \approx (1/2)(1 - 1/\beta)$ y se tiene una situación que corresponde a una caminata que explora un espacio de estados equiprobables (los nodos muy alejados del origen). Es este comportamiento “atípico” de la probabilidad el

responsable de la divergencia en el segundo momento. En forma más específica si $\beta = 1 + e$, con $e \ll 1$, entonces, la probabilidad de hacer un desplazamiento $\ell = z^k$ vale $P(k; e) = e / (2(1+e)^{(k+1)}) \approx e/2$ cuando $ke \ll 1$, el “campo cercano”. Para el “campo lejano” donde se presentan variaciones tales que $ke \gg 1$, $P(k; \varepsilon) \sim 1/(2k)$. En ambos casos, idénticos cuando se toma $e = 1/k$, la probabilidad, aunque decae muy rápidamente, es prácticamente constante muy lejos del origen y todos los nodos son “primeros vecinos” de todos los demás.

10.3 Conclusiones

En la caminata de Weierstrass cuando se está muy cerca del valor $\beta \approx 1$ se relaja la “prohibición” de dar saltos grandes, la probabilidad de ir a cualquier sitio es entonces $a = (1/2)(1 - 1/\beta) \approx 0$ y el sistema se torna homogéneo (el análogo clásico de una partícula libre en ausencia de un campo externo) en el sentido de que todos los nodos de la red tienen la misma probabilidad (si bien pequeña, $P(k; \beta) = a/\beta^k$) de ser alcanzados desde cualquier nodo particular. Estos casos extremos sugieren que el inverso del “parámetro de castigo”, $T = 1/\beta$, el parámetro que modula la intensidad de las transiciones $p(j \rightarrow \pm z^k)$, hace las veces de un parámetro térmico, jugando el papel de una “temperatura”. Así, cuando la temperatura disminuye a su valor mínimo, $T \rightarrow 1$, la caminata se “congela” y las partículas efectúan una búsqueda local, prácticamente sin posibilidad de grandes exploraciones; por otro lado a grandes temperaturas, $T \gg 1$, las “fluctuaciones térmicas”, esto es, las transiciones a nodos muy distantes, aumentan sin control y la caminata puede explorar cualquier sitio con igual probabilidad (un resultado que recuerda el postulado fundamental de la física estadística del equilibrio: todos los estados accesibles del sistema son igualmente probables). La transición entre la búsqueda difusiva, $z^2 < \beta$, y la fractal, $z^2 > \beta$, se presenta en las vecindades de la “curva crítica” $z^2 = \beta$. En este caso aparecen vuelos largos con mayor frecuencia y la búsqueda está a punto de dejar el régimen cuasidifusivo para entrar al régimen superdifusivo.

La caminata aleatoria de Weierstrass puede presentar características de tipo fractal. Por ejemplo, la propiedad de autosimilaridad, esto es, la invarianza de escala o la ausencia de escalas físicas subyacentes, aparece cuando el grado de dispersión o el tamaño de las fluctuaciones diverge, es decir, cuando el segundo momento es infinito. La invarianza de escala de la caminata de Weierstrass se presenta cuando $z^2 > \beta$. Un análisis asintótico detallado de la función característica de la distribución de Weierstrass demuestra que en el régimen de saltos largos $|\ell| \gg 1$, la distribución del tamaño de pasos sigue una ley de potencias (Paul, *et al.*, 1999)

$$P(l) \sim \frac{1}{|l|^{1+\alpha}} \quad (10.13)$$

con $1 < \alpha < 2$. Las distribuciones de probabilidad que siguen una ley de potencias (Cauchy, Pareto, Levy) son distribuciones extremadamente útiles para analizar situaciones como las que se presentan durante los cambios bruscos o turbulentos como es el caso de los quiebres financieros.

La caminata de Weierstras, y otros procesos aleatorios con propiedades parecidas, puede ayudar a entender mejor situaciones de alta variabilidad en los índices de precios del mercado financiero. En este capítulo se ha utilizado la analogía existente entre las transiciones de fase de la caminata y las fluctuaciones desordenadas que pudieran presentarse durante la evolución de los precios. En la vida real las fluctuaciones no siempre tienen la misma intensidad.

Generalmente la evolución temporal de los precios se modela mediante $S(t)$ o $\ln[S(t)]$, como es descrito por el movimiento Browniano geométrico, donde el término de volatilidad σ crece a un ritmo constante, lo cual, evidentemente, no modela del todo una situación real. Como una salida a este problema se propone que en el movimiento Browniano geométrico, la volatilidad Browniana sea reemplazada por un proceso de Weierstrass que incluya una dependencia en el tiempo de los parámetros involucrados. Al ser dependientes del tiempo, los parámetros deberán contener información sobre la manera en que se correlacionan los agentes en el mercado de acciones. En este contexto, el régimen superdifusivo ($\alpha > \beta$) de la caminata de Weierstrass correspondería a una situación donde con alta probabilidad, o muy próximo a ello, se pueda presentar un régimen financiero turbulento altamente inestable con variaciones en los precios típicas de una dinámica de invariancia de escala; en este momento la mayoría de los agentes del mercado financiero se encuentran altamente correlacionados. En situaciones como la descrita, se puede pensar en un proceso $S(t)$ cuyas fluctuaciones presentan intensidades distintas en intervalos de tiempos distintos. Lo anterior, sin duda, representa una situación mucho más realista para entender la dinámica de la variación de los precios de una acción, la valuación de opciones y su evolución temporal.

10.4 Bibliografía

Black, Fischer y Scholes, Mayron. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*, Vol. 81, No. 3, 637-654.

Bouchaud, Jean-Philippe and Potters, Marc. (2009). *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, 2 edition, United Kindom.

Clásicos, Critica, Barcelona, 2001.

Courtault, Jean-Michel and Kabanov, Yuri, et al. (2000). Louis Bachelier on the centenary of Th'eorie de la Sp'eculation, *Mathematical Finance*, Vol. 10, No. 3, 341-353.

Einstein, Albert. (1905). Sobre el movimiento de partículas pequeñas suspendidas en líquidos en reposo exigido por la teoría cinético-molecular del calor, Einstein 1905: Un año milagroso,

Drakontos

- Gerber, Hans U., and Shiu, Elias S. W.(2004). Optimal dividends: analysis with Brownian motion, *North American Actuarial Journal*, Vol. 8, No.1, 1-20.
- Laloux, L., Potters, M. and et al. (1999). *Are financial crashes predictable?*, *Europhys. Lett*, Vol. 45, 1-5.
- Mantegna, R. and Stanley, H. (2007). *Introduction to Econophysics: Correlations and Complexity*. Cambridge University Press, United Kingdom.
- Mahnke, R., Kaupuvzs, J., and Lubashevsky, I. (2009). *Physics of Stochastic Processes. How Randomness Acts in Time*. WILEY-VCH, Germany.
- Mikosch, Thomas. (1998). *Elementary stochastic calculus, with finance in view*. World scientific, USA.
- Osborne, M.F.M. (1962). Periodic Structure in the Brownian Motion of Stock Prices, *Operations Research*, Vol. 10, No. 3, 345-379.
- Paul, Wolfgang y Baschnagel, Jorg. (1999). *Stochastic Processes. From Physics to Finance*. Springer, Germany
- Sornette, D. (2004). *Why Stock Markets Crash: Critical Events in Complex Financial Systems*, Princeton University Press, USA.
- Sornette, et al. (1999). Predicting Financial Crashes Using Discrete Scale Invariance. *J. of Risk* 1, 5.
- Sottinen, Tommi. (2001). Fractional Brownian motion, random walks and binary market models, *Finance Stochast.*, 5, 343-355.
- Wiersema, Ubbo F. (2008). *Brownian Motion Calculus*. John Wiley & Sons, England.

Capítulo 11

Un modelo GARCH con asimetría condicional autorregresiva para modelar series de tiempo: Una aplicación para los rendimientos del Índice de Precios y Cotizaciones de la BMV

Rocío Durán-Vázquez

Departamento de Finanzas y Contaduría, Universidad de las Américas Puebla

Arturo Lorenzo-Valdés

Departamento de Finanzas y Contaduría, Universidad de las Américas Puebla

Antonio Ruiz-Porras

Departamento de Métodos Cuantitativos. Universidad de Guadalajara, CUCEA

Resumen

Desarrollamos un modelo GARCH con asimetría condicional autorregresiva para describir series de tiempo. Esto significa que, en añadidura a la media y varianza condicionales, asumimos que el sesgo describe el comportamiento de las series de tiempo. Analíticamente usamos la metodología propuesta por Fernández y Steel (1998) para definir el comportamiento de las perturbaciones del modelo. Usamos el enfoque desarrollado por Brooks, et. al. (2005) para construirlo. Más aun, mostramos su utilidad a través de modelar la serie de rendimientos diarios del Índice de Precios y Cotizaciones (IPC) de México durante el periodo comprendido entre el 3 de enero de 2008 y el 29 de septiembre de 2009.

11.1 Introducción

La modelación de series de rendimientos bursátiles es necesaria para entender el funcionamiento y evaluación de los mercados financieros. Particularmente, la adecuada modelación de las mismas es necesaria para analizar decisiones de inversión, de valuación de activos y de administración de riesgos. Por esta razón, la modelación de las series bursátiles ha tenido una gran importancia en los ámbitos teórico y empírico. Esta modelación se ha centrado en el comportamiento de las medias y las varianzas de las series de rendimientos. Esto en consistencia con los trabajos de Markowitz (1952) y Tobin (1958) sobre la media y la varianza de los portafolios financieros.

Tradicionalmente, los análisis econométricos usados para modelar las series financieras han usado un enfoque discreto. Los más conocidos modelos discretos son aquellos conocidos como ARMA y de la familia GARCH.¹ Particularmente, los modelos que asumen procesos ARMA se usan para explicar el comportamiento de la media condicional de las series. Los modelos de la familia GARCH explican el comportamiento de las series centrándose en su varianza condicional (y su volatilidad). Implícitamente, se asume que los primeros dos momentos de la distribución de probabilidad de las series, y sus variaciones en el tiempo, son suficientes para describir el comportamiento de las series financieras.

En este contexto, es necesario señalar que las características de las series financieras hacen difícil su modelación econométrica usando dos momentos solamente. Incluso hay quienes concluyen que ninguno de los modelos econométricos existentes es capaz de describir de manera integral dichas características (Malmsten y Teräsvirta, 2010). Usualmente las series financieras manifiestan curtosis excesivas, clusters de volatilidad, volatilidades no constantes y distribuciones no normales. En añadidura a las anteriores, la volatilidad de los rendimientos financieros frecuentemente suele experimentar de “efectos apalancamiento” cuando hay correlaciones negativas entre el rendimiento y la volatilidad.²

En esta investigación, desarrollamos un modelo GARCH con asimetría condicional autorregresiva. Esto significa que, en añadidura a la media y varianza, asumimos que el sesgo (i.e. el tercer momento), es necesario para describir las series de rendimientos. Así, el modelo describe el comportamiento de las series suponiendo que sesgo de las perturbaciones puede variar en el tiempo. En este contexto, no nos limitamos al planteamiento teórico del modelo. También ilustramos su utilidad modelando la serie de rendimientos diarios del Índice de Precios y Cotizaciones (IPC) para el periodo comprendido entre el 3 de enero de 2008 y el 29 de septiembre de 2009.

¹ Los nombres ARMA y GARCH derivan de las principales características que definen a los modelos econométricos. Concretamente, ARMA es acrónimo de “*Autoregressive Moving Average*”; GARCH es acrónimo de “*Generalized Autoregressive Conditional Heteroscedasticity*”. Así los modelos ARMA se caracterizan por tener una estructura donde se conjugan procesos autorregresivos y de media móvil. Los modelos de la familia GARCH, por su parte, se caracterizan por asumir una estructura autorregresiva y la existencia de heteroscedasticidad condicional en el término de perturbación.

² Los efectos apalancamiento se pueden describir como aquellas reacciones asimétricas que ocurren en los rendimientos financieros como consecuencia de “shocks” informativos. Por esta razón, cuando existen efectos apalancamiento, suele decirse que las malas noticias tienen un impacto mayor que las buenas noticias.

Metodológicamente desarrollamos esta investigación en cinco etapas. En la primera etapa revisamos la literatura. En la segunda etapa construimos la distribución de probabilidad asimétrica usada para definir el comportamiento de las perturbaciones en el modelo con base en la metodología de Fernández y Steel (1998). En la tercera etapa construimos el modelo GARCH con asimetría condicional autorregresiva con base en el enfoque de Brooks, et. al. (2005). En la cuarta etapa construimos y analizamos el comportamiento estadístico de la serie de rendimientos. Finalmente, en la quinta etapa, estimamos los parámetros del modelo propuesto y evaluamos su bondad de ajuste.

El estudio está organizado en seis secciones. La Sección 11.2 incluye la revisión de la literatura. Particularmente revisamos algunos modelos que consideran múltiples momentos y mencionamos algunos estudios que usan los mismos para estudiar el mercado bursátil mexicano. En la Sección 11.3 construimos la distribución de probabilidad asimétrica necesaria para definir el comportamiento de las perturbaciones. En la sección 11.4 desarrollamos el modelo GARCH con asimetría condicional autorregresiva. En la Sección 11.5 mostramos el ejercicio empírico de modelación del Índice de Precios y Cotizaciones. En la última sección, sintetizamos los resultados y mencionamos algunas líneas de investigación futura.

11.2 Revisión de la literatura

Los modelos GARCH son los más usados para modelar series económicas y financieras.³ Estos modelos, para efectos analíticos, se clasifican con base en sus especificaciones funcionales. Estas especificaciones se refieren a las formas en que la información rezagada de una variable y los momentos de la misma describen a la media, a la varianza condicional y a la distribución probabilística de las series. En la literatura se considera que existen más de un centenar de especificaciones distintas (Bollerslev, 2010). Sin embargo, en la gran mayoría de los casos, se asume que los primeros dos momentos de la distribución, y sus variaciones en el tiempo, son suficientes para explicar el comportamiento de las series.⁴

Metodológicamente los modelos GARCH pueden describir adecuadamente las series cuando sus perturbaciones estandarizadas siguen un comportamiento normal o t de Student. Sin embargo, es reconocido que las series de rendimientos no satisfacen esta condición (i.e. el sesgo es diferente de cero). Particularmente si el sesgo es negativo, la función de probabilidad es asimétrica a la izquierda (i.e. la probabilidad de tener rendimientos relativamente altos es mayor que la de tener rendimientos bajos). Si el sesgo es positivo, la función es asimétrica a la derecha. Esta consideración sugiere que los modelos simétricos tienden a ponderar inadecuadamente los rendimientos y riesgos.⁵

³ Revisiones amplias de los usos de estos modelos en la modelación econométrica de series son Poon y Granger (2003) y Bauwens, Laurent y Rombouts (2006).

⁴ Adviértase que si la función de probabilidad es normal, los dos primeros momentos son suficientes para definir la forma y comportamiento probabilístico de la serie analizada. En este caso, lo mismo que en otras funciones de probabilidad simétricas, el tercer momento de la distribución de probabilidad, y el sesgo, son iguales a cero.

⁵ Adviértase que asumir simetría en la función de los rendimientos equivale a subestimar (sobrestimar) las probabilidades de pérdida cuando la distribución de los rendimientos es asimétrica a la derecha (izquierda).

La asimetría condicional autorregresiva en los modelos también se justifica si los inversionistas son racionales. En este contexto resulta interesante señalar que Chunchachinda, et. al. (1997) muestran que los inversionistas cambian sus elecciones de portafolio cuando suponen sesgos en los mercados. Así sus hallazgos confirman indirectamente la utilidad de los modelos con asimetría condicional. Más aun, Chen, Hong y Stein (2001) también hallan evidencia de que los modelos con asimetría condicional pudieran ser convenientes para los inversionistas. De hecho, ellos son capaces de predecir “cracs” bursátiles usando estimaciones de los sesgos de los rendimientos financieros.

Las consideraciones anteriores justifican el desarrollo de modelos GARCH con sesgo y curtosis condicionales autorregresivos. Particularmente, en lo que se refiere a los modelos con sesgos condicionales, Hansen (1994) es quien primero propone un modelo con una distribución t de Student asimétrica. Harvey y Siddique (1999) y (2000) extienden ese trabajo y proponen una metodología para estimar el coeficiente de asimetría condicional asociado a la distribución t asimétrica.⁶ En lo que se refiere a los modelos con curtosis condicional destacan los de Jondeau y Rockinger (2003), León, Rubio y Serna (2005), Brooks, et. al. (2005), y White, Kim y Manganeli (2008).

En México, es reciente el uso de los modelos GARCH para describir el comportamiento de series bursátiles de manera agregada o desagregada. Particularmente Guzmán-Plata (1998), describe treinta y tres acciones mediante modelos ARCH-M. Asimismo Ortiz-Ramírez (2010) describe el IPC mediante un modelo GARCH-M. Cermeño-Bazán y Solís-Montes (2012) comparan los rendimientos accionarios en distintos sectores mediante modelos GARCH. Yamazaki-Tanabe y Díaz-Hernández (2012) describen la volatilidad del mercado usando modelos GARCH multivariados y datos de alta frecuencia. Estos estudios son relevantes aunque suponen que las distribuciones de probabilidad son simétricas.

En la literatura existen algunos estudios que han analizado los efectos asimétricos de la información en los rendimientos bursátiles mexicanos. Estos estudios no asumen sesgos condicionales; sino impactos diferenciados de las noticias sobre la volatilidad de los rendimientos (v.g. “efectos apalancamiento”). Entre estos estudios se encuentran aquellos de López-Herrera (2004), Cermeño-Bazán y Solís-Montes (2012) y Lorenzo-Valdés y Ruíz-Porras (2012). Estos estudios usan, entre otros, modelos EGARCH para describir los rendimientos. Su importancia radica en que los mismos proveen evidencia que justifica el uso de modelos GARCH con sesgo condicional autorregresivos con fines de modelación.

Finalizamos esta sección enfatizando que el desarrollo y uso de modelos de la familia ARCH resulta un área de gran interés en la econometría financiera. Las dinámicas de los mercados financieros han hecho manifiesta la necesidad de describir y predecir sus movimientos de la mejor manera posible. Esta necesidad ha fomentado el continuo desarrollo de modelos teóricos y su uso empírico en los mercados financieros. En el contexto del mercado bursátil mexicano, esta necesidad está asociada a fenómenos como la crisis financiera global y la

⁶ El coeficiente de asimetría es una medida del sesgo de la distribución de la probabilidad de una serie de tiempo. Se le define como el tercer momento central estandarizado.

integración de México a la economía global.⁷ De hecho, creemos que estas razones justifican la pertinencia de la investigación desarrollada.

11.3 Metodología

En esta sección definimos la función de distribución asimétrica que explica las perturbaciones (ε) en nuestro modelo. Esta función de probabilidad asimétrica la construimos con base en la metodología de Fernández y Steel (1998). Usamos esta metodología porque la misma nos permite transformar, de una manera muy sencilla, una distribución simétrica unimodal cualquiera en una asimétrica.⁸ La metodología solamente nos requiere usar un parámetro escalar, λ , para hacer dicha transformación. Particularmente, en nuestro caso, usamos dicha metodología para transformar una distribución normal en una distribución “normal sesgada”.⁹

Planteamos la transformación propuesta considerando la función de densidad de una normal estandarizada:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (11.1)$$

Siguiendo a Fernández y Steel (1998), introducimos la asimetría en esta distribución mediante factores de escala inversos en los valores positivos y negativos de x . Estos factores los definimos con el parámetro escalar $\lambda > 0$. Particularmente, si este escalar es fijo, la distribución de transformación queda definida como:

$$f(x|\lambda) = \frac{2}{(\lambda + 1/\lambda)} \left\{ f\left(\frac{x}{\lambda}\right) I_{(0,\infty)}(x) + f(\lambda x) I_{(-\infty,0)}(x) \right\}, \quad (11.2)$$

donde $I_A = \begin{cases} 1 & \text{si sucede } A \\ 0 & \text{no sucede } A \end{cases}$

Sustituyendo en la ecuación (11.2) a la distribución normal, la función de densidad de la normal sesgada queda como:

$$f(x|\lambda) = \frac{2}{(\lambda + 1/\lambda)\sqrt{2\pi}} \left\{ e^{-x^2/(2\lambda^2)} I_{(0,\infty)}(x) + e^{-(x\lambda)^2/2} I_{(-\infty,0)}(x) \right\}. \quad (11.3)$$

Estadísticamente, la función normal sesgada se comporta como una normal para valores positivos y el comportamiento de otra normal para los valores negativos. Si $\lambda = 1$, tenemos una distribución normal simétrica. Si $\lambda < 1$, tenemos una distribución asimétrica sesgada a la izquierda y si $\lambda > 1$, tenemos una distribución sesgada a la derecha (véase la Figura 11.1):

⁷ Existen varios trabajos que muestran la dependencia de los movimientos bursátiles en México con respecto a los mercados financieros de Estados Unidos. Un trabajo reciente donde se muestra dicha dependencia es el de Durán-Vázquez, Lorenzo-Valdés y Ruiz-Porras (2012).

⁸ La metodología de Fernández y Steel (1998) se ha usado en otros estudios (i.e. Lambert y Laurent 2001 y 2002).

⁹ Fernández y Steel (1998) ejemplifican el uso de su metodología construyendo una distribución t de Student “sesgada”.

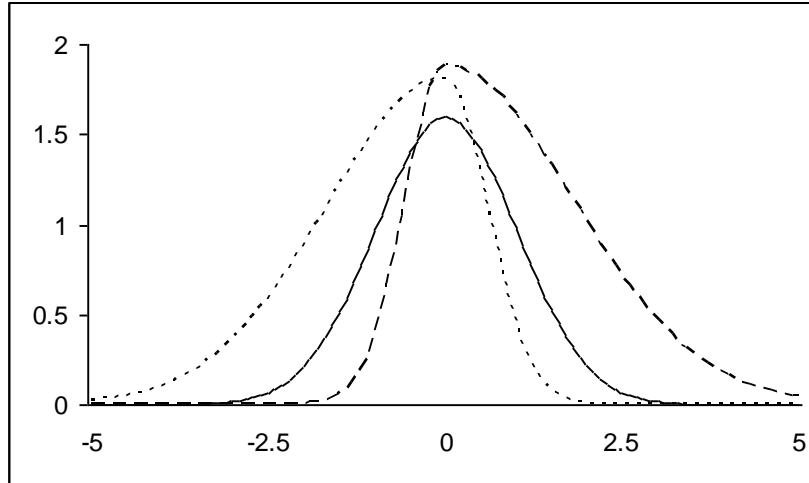


Figura 11.1. Funciones de densidad con $\lambda = 1$ (línea continua), $\lambda < 1$ (línea discontinua) y $\lambda > 1$ (línea punteada). Elaboración propia.

Las propiedades de la función en la ecuación (11.3) son:

- 1) $E[x] = \sqrt{\frac{2}{\pi}} \left(\lambda - \frac{1}{\lambda} \right) \equiv m$
- 2) $Var[x] = m^2 \left(\frac{\pi}{2} - 1 \right) + 1$
- 3) $E[(x - m)^3] = m \left[\left(2 - \frac{\pi}{2} \right) m^2 + 1 \right]$

Estadísticamente la variable x se distribuye como una normal sesgada con valor esperado igual a m . Esta variable no puede ser usada directamente para explicar las perturbaciones porque tiene un valor esperado diferente a cero. Por esta razón definimos la variable $\varepsilon = x - m$. Esta nueva variable tiene un valor esperado igual a cero y su segundo y tercer momentos centrales son idénticos a aquellos de x . Así, la variable ε resulta equivalente a la perturbación estandarizada utilizada en los modelos GARCH tradicionales. En este contexto, no sobra hacer notar que la varianza de ε no es unitaria. De hecho, esta situación explicará parcialmente la transformación de ε que se realizara en la siguiente sección.

11.4 Modelo GARCH con asimetría condicional autorregresiva

En esta sección construimos el modelo GARCH con asimetría condicional autorregresiva con base en el enfoque de Brooks, et. al. (2005). Este modelo lo construimos asumiendo la media, la varianza y la asimetría condicionales autorregresivos y un comportamiento de las perturbaciones dado por ε . Particularmente, dado que nuestro interés consiste en modelar series de

rendimientos, definimos la tasa de rendimiento continuo en el periodo t , r_t , como el cambio en los logaritmos del precio de un activo, P_t :¹⁰

$$r_t = \ln P_t - \ln P_{t-1}. \quad (11.4)$$

Estadísticamente supondremos que ε_t son variables aleatorias independientes e idénticamente distribuidas (iid) como la normal sesgada descrita en la sección anterior.¹¹ En este contexto, consideraremos ahora una transformación de ε_t que denotaremos como δ_t . Esta transformación la usaremos para definir el proceso de las perturbaciones del modelo GARCH, u_t . Por conveniencia, este último proceso lo construiremos para que su varianza condicional sea σ_t^2 y para que su distribución de probabilidad dependa de un parámetro de asimetría condicional λ_t . Este proceso lo definiremos como:

$$u_t = \delta_t \varepsilon_t. \quad (11.5)$$

La transformación δ_t la definimos como una función de la varianza condicional y de la asimetría a fin de que satisfaga los requerimientos de la varianza de u_t :

$$\delta_t = \left(\frac{\sigma_t^2}{m_t^2 [\pi/2 - 1] + 1} \right)^{1/2}. \quad (11.6)$$

Si definimos el conjunto de información como Ω_t , la transformación (11.6) permite que la variable u_t , tenga definidos los siguientes momentos:

$$\begin{aligned} E[u_t] &= E[u_t | \Omega_{t-1}] = 0 \\ \text{Var}[u_t | \Omega_{t-1}] &= E[u_t^2 | \Omega_{t-1}] = \sigma_t^2 \\ E[u_t^3 | \Omega_{t-1}] &= \delta_t^3 m_t \left[\left(2 - \frac{\pi}{2} \right) m^2 + 1 \right] \end{aligned} \quad (11.7)$$

Estadísticamente, los momentos anteriores existen cuando σ_t^2 y λ_t son procesos autorregresivos. Esto conlleva a que σ_t^2 y λ_t sean medibles en el tiempo t . Más aun, conlleva a que m_t y δ_t también sean medibles en t .

Económicamente, el modelo GARCH con asimetría condicional autorregresiva queda definido por las siguientes expresiones:

$$\begin{aligned} r_t &= v_t + u_t \\ u_t &= \delta_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \\ \lambda_t &= \gamma_0 + \gamma_1 [u_{t-1}^2 b_t] + \gamma_2 \lambda_{t-1} \end{aligned} \quad (11.8)$$

¹⁰ Adviértase que esta formulación supone que los precios siguen un proceso continuo de interés compuesto.

¹¹ Brooks et. al. (2005), a diferencia nuestra, asumen una distribución t de Student.

La primera expresión define la dinámica de los rendimientos. En dicha expresión $v_t = E[r_t | \Omega_{t-1}]$ es la media condicional de los rendimientos y se describe mediante un proceso ARMA. La segunda expresión define al proceso de perturbaciones u_t . La tercera expresión describe la volatilidad condicional $\sigma_t^2 = Var[r_t | \Omega_{t-1}] = E[u_t^2 | \Omega_{t-1}]$ como un GARCH(1,1). La última expresión describe el parámetro de asimetría condicional de manera similar a un proceso GARCH(1,1).¹² Por esta razón la expresión de asimetría debe cumplir las mismas restricciones que este último proceso. En términos prácticos esto significa que los parámetros γ_0, γ_1 y γ_2 deben ser positivos y que $\gamma_1 + \gamma_2 < 1$.

El coeficiente de asimetría condicional (CA_t) lo podemos construir a partir de la asimetría condicional (λ_t). Este coeficiente es el siguiente:

$$CA_t \equiv \frac{E[(r_t - v_t)^3 | \Omega_{t-1}]}{\sigma_t^3} = \frac{\delta_t^3}{\sigma_t^3} m \left[\left(2 - \frac{\pi}{2} \right) m^2 + 1 \right]. \quad (11.9)$$

Finalizamos esta sección indicando que el conjunto de especificaciones (11.8) define nuestro modelo GARCH con asimetría condicional autorregresiva. Este modelo lo hemos construido asumiendo que la media, la varianza y el sesgo son necesarios para describir el comportamiento de las series. En virtud de que la relevancia de incorporar la asimetría la hemos sustentado en consideraciones empíricas, la validación del modelo solo puede hacerse empíricamente. Particularmente, aquí validamos el modelo describiendo el comportamiento del Índice de Precios y Cotizaciones (IPC). Los resultados de dicho ejercicio de modelación econométrica los mostramos en la siguiente sección.

11.5 Ejercicio de modelación econométrica del Índice de Precios y Cotizaciones

En esta sección usamos el modelo GARCH con asimetría condicional autorregresiva, descrito en (11.8), para describir el rendimiento del Índice de Precios y Cotizaciones (IPC)¹³. Particularmente los datos empleados en este estudio comprenden los precios de cierre diarios del 3 de enero de 2008 al 29 de septiembre de 2009. Así la muestra de datos comprende 455 observaciones diarias.

¹² Matemáticamente es necesario hacer notar que en la expresión donde se define la asimetría condicional, aparece el término b_t . Este término se calcula como $b_t = \frac{\lambda_t}{\delta_t^3 \varepsilon_t^3}$. El objeto de este término que garantizar que la expresión de la asimetría condicional sea insesgada. Adviértase que $E[u_t^3 b_t | \Omega_{t-1}] = \lambda_t$. Analíticamente este termino lo deducimos de manera similar a como garantizamos que la expresión de la varianza condicional fuera insesgada. Adviértase que, en este último caso, $E[u_t^2 | \Omega_{t-1}] = \sigma_t^2$

¹³ El Índice de Precios y Cotizaciones (IPC) es el indicador representativo del comportamiento accionario de la Bolsa Mexicana de Valores (BMV). Este índice expresa el rendimiento del mercado accionario en función de las variaciones de precios de una muestra balanceada, ponderada y representativa de las acciones cotizadas. Este índice mide el cambio diario del valor de capitalización de las 35 emisoras de mayor bursatilización de la BMV.

En virtud de que nuestro interés consiste en analizar la serie de rendimientos, construimos la serie de rendimientos agregados del mercado bursátil mexicano con base en la ecuación (11.4). Particularmente, la Tabla 11.1 muestra las estadísticas descriptivas de la serie estimada de rendimientos.

	r_{IPC}
Media	-0.000105
Máximo	0.104407
Mínimo	-0.072661
desv. estand.	0.019986
Asimetría	0.371623
Curtosis	7.024864
Observ.	454

Tabla 11.1. Estadísticas descriptivas. Elaboración propia

Analíticamente es interesante hacer notar que la media estimada es negativa y el coeficiente de asimetría puntual estimado es positivo. Como hemos mencionado anteriormente, los sesgos positivos indican que las probabilidades de obtener rendimientos bajos son mayores que aquellas de obtener altos rendimientos. El estimado del sesgo, si bien atípico, puede explicarse si se consideran los efectos de la crisis financiera global durante el periodo analizado. Es bien conocido que estos efectos redujeron en buena medida los rendimientos bursátiles a nivel internacional.¹⁴ Por estas razones, no sobra indicar que consideramos que el ejercicio de modelación de los rendimientos bursátiles es interesante por sí mismo.

Económicamente usamos un proceso autorregresivo de orden uno para definir la media condicional de los rendimientos. Esto hace que el modelo estimado sea:

$$\begin{aligned}
 r_t &= \phi_0 + \phi_1 r_{t-1} + u_t \\
 u_t &= \delta_t \varepsilon_t \\
 \sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \\
 \lambda_t &= \gamma_0 + \gamma_1 [u_{t-1}^2 b_t] + \gamma_2 \lambda_{t-1}
 \end{aligned}
 \tag{11.10}$$

Los resultados de la estimación realizada se presentan en las Tablas 11.2 y 11.3. Cabe aclarar que para revisar la especificación del modelo, en este caso, a diferencia de los GARCH en donde se

¹⁴ El Centro de Estudios de las Finanzas Públicas (2009) de la Cámara de Diputados de México muestra que en el año 2008, el IPC tuvo una variación porcentual de -25.8%. La mayoría de las emisoras registraron pérdidas debido a la volatilidad e inestabilidad del mercado cambiario y de divisas, así como la depreciación de la moneda nacional. En el año 2009 se reflejó una ligera recuperación en concordancia con la situación económica del país.

diagnostican los residuales estandarizados, aquí utilizamos los residuales transformados por delta, es decir $\hat{\varepsilon}_t = \hat{u}_t / \hat{\delta}_t$.

	Coefficiente	Error estándar
ϕ_0	-0.000755	0.00011
ϕ_1	0.08233	0.002145
α_0	2.98E-06	8.62E-08
α_1	0.094671	0.001766
α_2	0.898395	0.000871
γ_0	0.040364	0.003808
γ_1	0.012946	0.000385
γ_2	0.953923	0.001032

Tabla 11.2. Se presentan los parámetros, estimados por máxima verosimilitud, con sus respectivos errores estándar. Elaboración propia.

	Residuales transformados		Res. transformados al cuadrado		Residuales transformados al cubo	
	Q	Prob.	Q	Prob.	Q	Prob.
1	0.0659	0.7970	0.0022	0.9630	0.5136	0.4740
10	4.2782	0.9340	11.0480	0.3540	16.9380	0.0760
20	15.7060	0.7350	21.9560	0.3430	27.9380	0.1110
30	25.7150	0.6900	27.8480	0.5790	48.2370	0.0190

Tabla 11.3. Estadísticos de Ljung-Box para los órdenes 1, 10, 20 y 30 de los residuales transformados y de los residuales transformados al cuadrado y al cubo. Elaboración propia.

En los resultados se aprecia que todos los parámetros son significativamente diferentes de cero y que el ajuste es bueno, observando que, en general, no hay dependencia lineal, cuadrática o al cubo. En este último caso, para la autocorrelación de orden 30 en los residuales al cubo, se rechaza la hipótesis nula de dependencia cúbica con un nivel de significancia del 5%, pero no a un nivel de 1%. Lo anterior nos indica que el modelo presenta un buen ajuste y se puede utilizar para realizar inferencias.

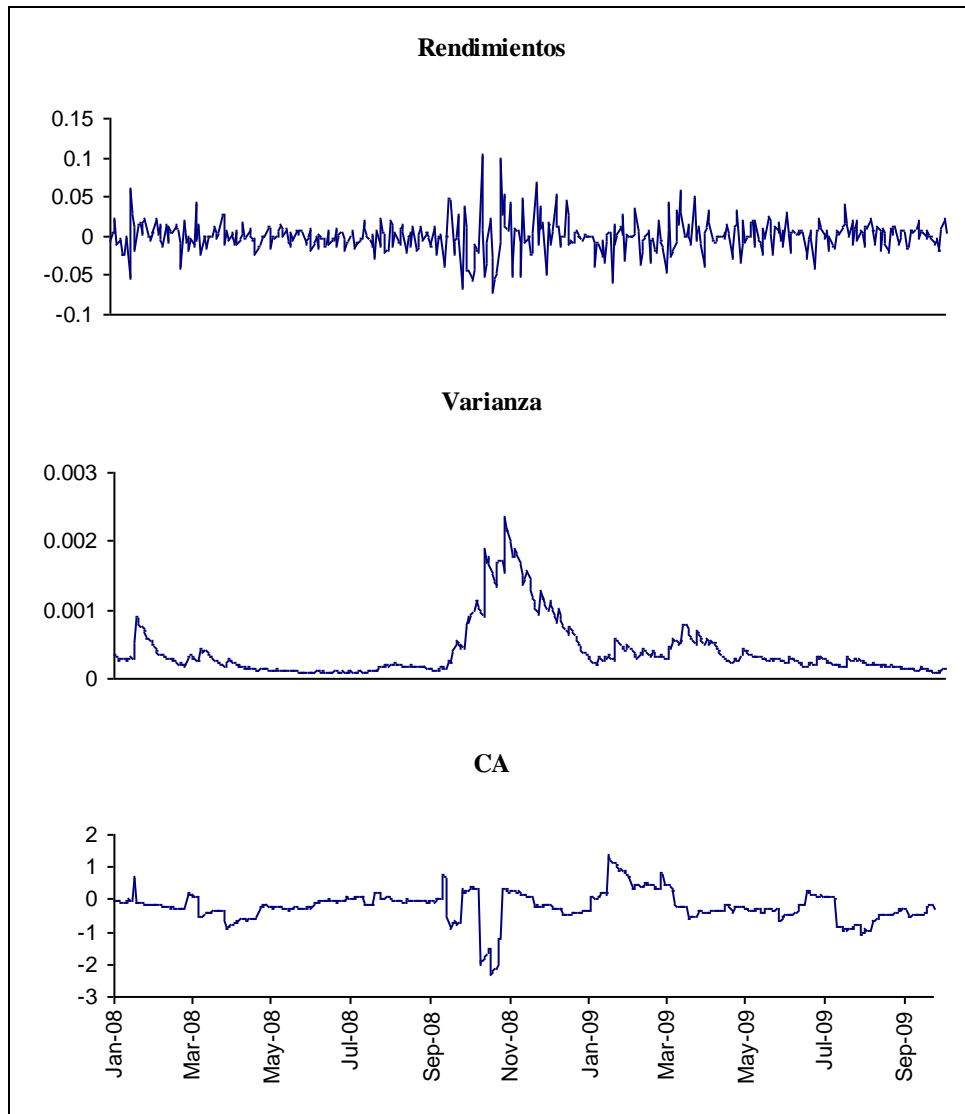


Figura 11.2. Gráficas de los rendimientos del IPC, varianza condicional del IPC y coeficiente de asimetría condicional (CA) del IPC. Los dos últimos estimados mediante el modelo. Elaboración propia.

En la Figura 11.2 presentamos las gráficas de los rendimientos y de las series de la volatilidad y el coeficiente de asimetría obtenidos mediante el modelo desarrollado. Visualmente podemos apreciar que la asimetría es más pronunciada hacia la parte negativa cuando aumenta la volatilidad. Por tanto es posible inferir que al aumentar la asimetría negativa, la volatilidad de los rendimientos y, eventualmente, su media tienden a incrementarse. Evidentemente esta inferencia, de mantenerse en el tiempo, pudiera resultar de importancia para tomar decisiones financieras diversas; como son las referidas a construir portafolios de inversión, valorar activos y administrar riesgos.

Académicamente las gráficas también muestran los efectos de la crisis financiera global en el mercado mexicano. De acuerdo el Fondo Monetario Internacional (2009), se distinguen

cuatro periodos diferenciados en la crisis global. Lo interesante de los mismos es que estos tienen correspondencia con el desempeño bursátil observado en las gráficas. Particularmente el primer periodo, que comprende de agosto 2007 a mayo 2008 se identifica un periodo de inestabilidad en el IPC. En el segundo periodo, de junio a octubre de 2008, ocurren caídas en la Bolsa Mexicana de Valores como respuesta al comportamiento observado en los mercados financieros internacionales. En el tercer periodo, que comprende de noviembre de 2008 a marzo de 2009, se registran altibajos e inestabilidad en el mercado. En el cuarto periodo, que comprende de marzo a julio de 2009, el comportamiento del mercado mexicano refleja una cierta estabilidad

Finalizamos esta sección indicando que los resultados del ejercicio de modelación validan la conveniencia de usar el modelo GARCH con asimetría condicional autorregresiva. Como hemos mostrado, el modelo parece reflejar adecuadamente el comportamiento de los rendimientos del Índice de Precios y Cotizaciones. El ajuste ocurre a pesar de lo atípico del periodo analizado. Particularmente el modelo sugiere que aumentos en la asimetría negativa tienden a incrementar la volatilidad de los rendimientos y, eventualmente, su media. Más aún, las estimaciones de la media, varianza y del coeficiente de asimetría parecen validar la existencia de las cuatro etapas de la crisis financiera global.

11.6 Conclusiones y líneas de investigación futura

En este estudio hemos desarrollado un modelo GARCH con asimetría condicional autorregresiva para modelar series financieras. Esto significa que, en añadidura a la media y varianza, hemos asumido que el sesgo es necesario para describir el comportamiento de las series. Analíticamente hemos usado la metodología de Fernández y Steel (1998) para definir el comportamiento de las perturbaciones. Asimismo hemos usado el enfoque de Brooks, et. al. (2005) para construir el modelo. Más aun, hemos mostrado su utilidad modelando la serie de rendimientos diarios del Índice de Precios y Cotizaciones (IPC) considerando el periodo entre el 3 de enero de 2008 y el 29 de septiembre de 2009.

Los resultados de esta investigación validan la conveniencia de usar el modelo GARCH con asimetría condicional autorregresiva. Empíricamente el modelo parece reflejar adecuadamente el comportamiento de los rendimientos del Índice de Precios y Cotizaciones. El ajuste ocurre a pesar de lo atípico del periodo analizado. Particularmente el modelo sugiere que aumentos en la asimetría negativa tienden a incrementar la volatilidad de los rendimientos y, eventualmente, su media. Más aún, las estimaciones de la media, varianza y del coeficiente de asimetría parecen validar la existencia de las cuatro etapas de la crisis financiera global.

Estos resultados tienen implicaciones econométricas y financieras. Desde una perspectiva econométrica, los resultados validan la conveniencia de usar modelos que consideren momentos de orden superior a la varianza para describir los rendimientos bursátiles en economías emergentes.¹⁵ Desde una perspectiva financiera, los resultados sugieren que el coeficiente de

¹⁵ En las economías en desarrollo los estudios que usan modelos GARCH que asumen momentos de orden superior a la varianza son muy escasos. Entre estos se encuentra el de Ahmad (2011). Su estudio modela la inflación en Egipto.

asimetría condicional puede utilizarse con provecho en la valuación de activos (opciones, notas estructuradas), en la construcción y evaluación del desempeño de portafolios de inversión, así como en la administración de riesgos.¹⁶ Por tanto, los resultados sugieren que el modelo pudiera ser una herramienta útil para tomar decisiones.

Finalizamos este estudio indicando algunas líneas de investigación futura. Una primera se refiere al desarrollo de otros modelos GARCH con asimetría condicional. Esto en virtud de que es factible usar los procedimientos usados aquí en modelos similares (EGARCH, IGARCH, GARCH-M, etc.). Una segunda línea, y natural extensión de este estudio, se refiere al desarrollo de modelos con curtosis condicional autorregresiva. Una tercera línea se refiere al uso de modelos GARCH multivariados y con saltos para describir y predecir series.¹⁷ Muy probablemente, el desarrollo de investigación alrededor de estas líneas será de gran utilidad para entender la dinámica de las series económicas y financieras.

11.7 Bibliografía

- Ahmad, D.A., (2011), “Modelling the density of inflation using autoregressive conditional heteroscedasticity, skewness, and kurtosis models”, *Ensayos Revista de Economía*, 30(2), 1-28
- Bollerslev, T. (2010), “Glossary to ARCH (GARCH)”, en Bollerslev, T., Russell, J.R. y M.W. Watson, (eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle* (Oxford University Press, Oxford), 137-163
- Brooks, C., S. Burke, S. Heravi y G. Persaud, (2005), “Autoregressive conditional kurtosis”, *Journal of Financial Econometrics*, 3(3), 399–421
- Campbell, R.H. y A. Siddique, (2000), “Conditional skewness in asset pricing tests”, *Journal of Finance*, 55(3), 1263–1296
- Centro de Estudios de las Finanzas Públicas, (2009), *El Mercado de Valores en México*, LXI Legislatura, Cámara de diputados. CEFP-124-2009.

¹⁶ Los modelos de valuación que se definen con base en una estructura media-varianza asumen comportamientos simétricos. Esto es, tienden a sobre y a subvaluar las probabilidades de pérdidas y ganancias. Modelos tan conocidos y usados como el CAPM y el de Black-Scholes asumen comportamientos simétricos.

¹⁷ Un estudio teórico reciente vinculado a esta línea de investigación es el de Sánchez Torres, Ortiz Arango y Venegas Martínez (2012).

- Cermeño-Bazán, R. y M.P. Solís-Montes, (2012), “Impacto de sorpresas macroeconómicas de México y Estados Unidos sobre el mercado accionario mexicano”, *Economía Mexicana*, 21(1), 35-67
- Chen, J., H. Hong y J. Stein, (2001), “Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices”, *Journal of Financial Economics*, 61(3), 345-381
- Chunchachinda, P., K. Dandapani, S. Hamid, y A. Prakash, (1997), “Portfolio selection and skewness: Evidence from international stock markets”, *Journal Banking and Finance*, 21(2), 143–167
- Durán-Vázquez, R., A. Lorenzo-Valdés y A. Ruiz-Porras, (2012), “Valuation of Latin-American stock prices with alternative versions of the Ohlson model: An investigation of cointegration relationships with time-series and panel-data”, en R. Espinosa Ramirez (coord.), in *Research Issues in International Economic Relations*, (Universidad de Guadalajara, México), 185-207 [por aparecer]
- Fernández, C. y M.F.J. Steel, (1998), “On Bayesian modeling of fat tails and skewness”, *Journal of the American Statistical Association*, 93(441), 359–371
- FMI, (2009), “Comunicado del Comité Monetario y Financiero Internacional de la Junta de Gobernadores del Fondo Monetario Internacional”, *Comunicado de Prensa No. 09/139* 25 de abril 2009. <http://www.imf.org/external/spanish/np/sec/pr/2009/pr09139s>
- Guzmán-Plata, M.P. (1998), “Los modelos CAPM y ARCH-M: Obtención de los coeficientes beta para una muestra de 33 acciones que cotizan en la Bolsa Mexicana de Valores”, *Economía, Teoría y Práctica*, 9, 63-76
- Hansen, B.E, (1994), “Autoregressive conditional density estimation”, *International Economic Review*, 35(3), 705–730
- Harvey, C. y A. Siddique, (1999), “Autoregressive conditional skewness”, *Journal of Financial and Quantitative Analysis*, 34(4), 465–487
- Harvey, C. y A. Siddique, (2000), “Conditional skewness in asset pricing tests”, *Journal of Finance*, 55(3), 1263-1295
- Jondeau, E. y M. Rockinger, (2003), “Conditional volatility, skewness and kurtosis: Existence, persistence and comovements”, *Journal of Economic Dynamics and Control*, 27(10), 1699-1737

- Lambert, P. y S. Laurent (2001), “Modeling financial time series using GARCH-type models with a skewed student distribution for the innovations”, Lovaina, Université Catholique de Louvain-Institut de Statistique, *Discussion Paper 0125*
- Lambert, P. y S. Laurent (2002), “Modeling skewness dynamics in series of financial data using skewed location-scale distributions”, Lovaina, Université Catholique de Louvain-Institut de Statistique, *Discussion Paper 0119*
- León A., G. Rubio y G. Serna, (2005), “Autoregressive conditional volatility, skewness and kurtosis”, *The Quarterly Review of Economics and Finance*, 45(4-5), 599–618
- López-Herrera, F. (2004), “Modelado de la volatilidad y pronóstico del índice de precios y cotizaciones de la Bolsa Mexicana de Valores”, *Contaduría y Administración*, 213, 43-72
- Lorenzo-Valdés, A. y A. Ruiz-Porras (2012), “Modelación de los rendimientos bursátiles mexicanos mediante los modelos TGARCH y EGARCH: Un estudio econométrico para 30 acciones y el Índice de Precios y Cotizaciones”, en S. Coronado Ramírez y L. Gatica Arreola (coords.), *Modelos no Lineales en Series Económicas y/o Financieras*, (Universidad de Guadalajara, Guadalajara), 46-81
- Malmsten, H. y T. Teräsvirta, (2010), “Stylized facts of financial time series and three popular models of volatility”, *Journal of Pure and Applied Mathematics*, 3(3), 413-447
- Markowitz, H. (1952), “Portfolio selection”, *Journal of Finance*, 7(1), 77-91
- Ortíz-Ramírez, A., (2010), “Valuación de opciones europeas sobre el IPC con un modelo GARCH para la volatilidad”, en Ortíz-Arango, F. (coord.), *Avances Recientes en Valuación de Activos y Administración de Riesgos, Vol. 1.* (Universidad Panamericana, México D.F.), 217-233
- Sánchez-Torres, F.J., Ortíz-Arango, F y Venegas-Martínez, F. (2012), “Volatilidad estocástica y procesos de difusión GARCH”, en Ortíz-Arango, F. y F. López-Herrera, (coords.), *Avances Recientes en Valuación de Activos y Administración de Riesgos, Vol. 3.* (Universidad Panamericana, México D.F.), 143-154
- Tobin, J. (1958), “Liquidity preference as behavior towards risk”, *Review of Economic Studies*, 25(2), 65-86

White, H., TH. Kim y S. Manganelli, (2008), “Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR”, Francfort, European Central Bank, *Working Paper* 957

Yamazaki-Tanabe, E. y A. Díaz-Hernández (2012), “Modelado de la volatilidad del mercado accionario mexicano con datos de alta frecuencia”, en Ortíz-Arango, F. y F. López-Herrera, (coords.), *Avances Recientes en Valuación de Activos y Administración de Riesgos, Vol. 3.* (Universidad Panamericana, México D.F.), 23-45

Capítulo 12

Sobre la no unicidad de soluciones del problema de decisión de consumo y portafolio en presencia de saltos

Francisco Venegas-Martínez

Escuela Superior de Economía, Instituto Politécnico Nacional

Abigail Rodríguez Nava

Departamento de Producción Económica, UAM-X

Francisco López Herrera

División de Investigación, Facultad de Contaduría y Administración, UNAM

Resumen

Este capítulo desarrolla un modelo estocástico que describe el proceso de toma de decisiones de un consumidor-inversionista racional que desea integrar un portafolio en un ambiente de riesgo sujeto a dos restricciones: una de tipo presupuestal y otra sobre para asegura una riqueza final con un cierto nivel de confianza. Se muestra la no unicidad de la solución del problema del consumidor en presencia de saltos. Para ello se exhibe que el precio de estado no es determinado de manera única por una medida martingala. Por último, se obtiene la de mínima entropía cruzada.

12.1 Introducción

Una de las características importantes de los mercados financieros es que éstos ocasionalmente presentan saltos bruscos, ya sea hacia arriba o hacia abajo, en los precios de los activos. Esta investigación modela explícitamente la existencia de saltos en los precios de los activos e incluye una restricción sobre el nivel de la riqueza final en el proceso de toma de decisiones de consumo y portafolio de un individuo racional. Existe una extensa literatura sobre el modelado de saltos en los precios de los activos; véanse, por ejemplo: Cox y Ross (1976), Ball y Torous (1985), Page y Sanders (1986), Cao (2001), Chandrasekhar Reddy Gukhal (2004) y Venegas-Martínez (2001), (2006a) y (2008b). As importante destacar que este trabajo extiende, en varias direcciones, las investigaciones de Gavira-Durón y Venegas-Martínez (2011) y Venegas-Martínez y López-Herrera (2013), entre otros.

En vista de los saltos en los precios de los activos observados a finales de 2008 y a principios de 2009 en los mercados financieros de casi todo el mundo y de las consecuencias negativas que tuvo la burbuja especulativa generada por dichos mercados, es imprescindible modelar el comportamiento de los precios de los activos utilizando procesos estocásticos con saltos, así como incorporar en el proceso de toma de decisiones una restricción adicional sobre la riqueza final de tal forma que exceda un umbral determinado con un cierto nivel de confianza.

Aunque el problema de asignar proporciones de la riqueza a la tenencia de los diferentes activos disponibles en los mercados financieros se ha estudiado ampliamente, poca atención se ha puesto en el hecho de que cuando hay saltos repentinos y extremos en los precios de los activos los inversionistas desean al final de su horizonte de planeación que su riqueza exceda un determinado umbral (nivel de subsistencia) con cierto nivel de confianza. Por supuesto, si el nivel de confianza fuera de $(1-\alpha)\times 100\%$ siempre existe una probabilidad de magnitud α de que el umbral no se exceda.

En el presente trabajo se desarrolla un modelo estocástico que describe el proceso de toma de decisiones de un consumidor-inversionista racional en el proceso de integrar un portafolio en un ambiente de riesgo sujeto a dos restricciones: una de tipo presupuestal y otra que asegura que la riqueza exceda un umbral con un cierto nivel de confianza. Se muestra la no unicidad de la solución del problema del consumidor en presencia de saltos. Para ello se exhibe que el precio de estado no es determinado de manera única por una martingala. Por último, ya que la martingala asociada al precio de estado no es única se obtiene la de mínima entropía cruzada.

Esta investigación está organizada de la siguiente forma: la sección 12.2 proporciona las características de la economía; la sección 12.3 describe los activos disponibles en la economía y la dinámica de los rendimientos de los mismos; la sección 12.4 plantea la restricción presupuestal del consumidor en términos de la dinámica estocástica de la riqueza real marginal; a través de la sección 12.5 se plantea el problema de optimización del consumidor racional; en el transcurso de la sección 12.6 se establecen las condiciones de primer orden para una solución interior del problema del consumidor representativo; en la sección 12.7 se replantea el problema de optimización del consumidor en términos de una restricción estática que involucra el valor esperado y descontado de la riqueza del agente; en la sección 12.8 se muestra la no unicidad de los precios de estado; en la sección 12.9 se determina el precio de estado con la martingala de mínima entropía; y, por último, en la sección 12.10 se presentan las conclusiones.

12.2 Características de la economía bajo estudio

Considere una economía que produce y consume un solo bien genérico (el numerario) y que está poblada por un número finito de individuos idénticos en gustos y dotaciones que toman decisiones en un horizonte infinito. La incertidumbre en la economía en cuestión está representada por dos espacios de probabilidad: el primero, asociado con difusiones (pequeños movimientos que se observan todos los días en los precios de los activos), equipado con una filtración, $(\Omega, F_w, (F_{tW})_{t \in [0, T]}, P_w)$ sobre el cual está definido un movimiento browniano estándar, $dW_t \sim \mathcal{N}(0, dt)$, de tal forma que

$$E[dW_t] = 0$$

y

$$\text{Var}[dW_t] = dt,$$

y en el segundo espacio de probabilidad, asociado con los saltos, $(\Omega, F_N, (F_{tN})_{t \in [0, T]}, P_N)$, en el que está definido un proceso de Poisson dN_t con parámetro de intensidad ϕ , de tal forma que

$$P_N\{\text{un salto de longitud uno durante } dt\} = P_N\{dN_t = 1\} = \phi dt$$

y

$$P_N\{\text{más de un salto durante } dt\} = P_N\{dN_t > 1\} = o(dt),$$

así

$$P_N\{\text{ningún salto durante } dt\} = 1 - \phi dt + o(dt),$$

donde, como siempre, $o(dt)/dt \rightarrow 0$ cuando $dt \rightarrow 0$. Observe que el tiempo medio esperado entre dos saltos es $1/\phi$. Se puede mostrar, fácilmente, que $E_N[dN_t] = \text{Var}_N[dN_t] = \phi dt$.

En lo que sigue, todos los procesos estocástico que se introduzca con posterioridad se suponen adaptados a la filtración $F_t \equiv F_{tW} \times F_{tN}$ en el espacio de probabilidad generado por $P \equiv P_W \otimes P_N$. Asimismo, se supone que todas las igualdades que involucran procesos estocásticos se cumplen P -casi seguramente (con probabilidad uno). Por supuesto, se supone que todos los procesos están bien definidos, sin establecer explícitamente las condiciones de regularidad que lo garanticen.

12.3 Activos disponibles y sus rendimientos

Las posibilidades de inversión, en los mercados financieros de esta economía, están representadas por un bono de vencimiento instantáneo (el cual en automático es cupón cero), libre de riesgo de incumplimiento (emitido por el gobierno), que paga tasa de interés (anualizada) constante y continuamente capitalizable r . La tasa r es real en el sentido de que los bonos pagan intereses en términos de bienes. Se supone que el precio del bono, B_t , y el precio del activo riesgoso (una acción), S_t , tienen las siguientes dinámicas:

$$dB_t = B_t r dt, \tag{12.1)1}$$

y

$$dS_t = S_t [(\mu - \phi\nu)dt + \sigma dW_t + \nu dN_t], \tag{12.2)}$$

donde μ es el rendimiento medio anualizado de la acción, σ es la desviación estándar de los rendimientos anualizados alrededor de μ y, por último, ν es el tamaño medio esperado del salto. Los parámetros μ , σ y ν son llamados, respectivamente, de tendencia, volatilidad y tamaño del salto, respectivamente. Observe que en (2) se puede reescribir como

$$dS_t = S_t [\mu dt + \sigma dW_t + \nu X_t],$$

de tal forma que el proceso $X_t = dN_t - \phi dt$ satisface

$$E_N[X_t] = 0$$

y

$$\text{Var}_N[X_t] = \phi dt,$$

lo que asegura que X_t sea una martingala. Se supone también que la acción paga dividendos a tasa δ de tal forma que

$$dD_t = \delta S_t dt. \quad (12.3)$$

Se supone además que los procesos dW_t y dN_t son independientes entre sí, lo cual implica que $\text{Cov}(dW_t, dN_t) = 0$.

12.4. Restricción presupuestal del agente

Se supone que el agente tiene una dotación inicial de riqueza conocida, en términos de bienes, a_0 . Asimismo se supone que la riqueza real del individuo está dada por el portafolio

$$\Pi_t \equiv A_t = B_t + S_t, \quad 0 \leq t \leq T, \quad (12.4)$$

donde el valor inicial del portafolio

$$\Pi_0 = a_0 = B_0 + S_0$$

es conocido. Como puede observarse el portafolio Π_t , $0 \leq t \leq T$, es autofinanciable. Los cambios marginales en el valor del portafolio se deben exclusivamente a cambios marginales en los precios de los activos, de tal manera que

$$\begin{aligned} d\Pi_t &= dB_t + dS_t + dD_t \\ &= a_t \left(\frac{B_t}{a_t} \right) \frac{dB_t}{B_t} + a_t \left(\frac{S_t}{a_t} \right) \frac{dS_t + dD_t}{S_t} \\ &= a_t (1 - \theta_t) \frac{dB_t}{B_t} + a_t \theta_t \frac{dS_t + dD_t}{S_t} \end{aligned} \quad (12.5)$$

donde $\theta_t = S_t / a_t$ y $1 - \theta_t = B_t / a_t$ son, respectivamente, la proporción de la riqueza real que se asigna a la tenencia del activo riesgoso y la proporción, complementaria, que se destina a la compra

de bonos. Para calcular el cambio marginal de la riqueza, da_t , se requiere restar el consumo por unidad del tiempo del valor del portafolio, es decir,

$$da_t = d\Pi_t - dc_t, \quad 0 \leq t \leq T. \quad (12.6)$$

Después de sustituir las ecuaciones (12,1)-(12.3) en la expresión (12.6) y de escribir $dc_t = c_t dt$, se tiene que (12.5) se transforma en

$$da_t = a_t \left(r + \theta_t \lambda \sigma - \frac{c_t}{a_t} \right) dt + a_t \theta_t \sigma dW_t + a_t \theta_t v dN_t, \quad (12.7)$$

donde

$$\lambda = \frac{\mu + \delta - \phi v - r}{\sigma} = -\frac{\phi v}{\sigma} + \bar{\lambda} \quad (12.8)$$

con

$$\bar{\lambda} = \frac{\mu + \delta - r}{\sigma} \quad (12.9)$$

La última cantidad representa el premio al riesgo de mercado-

12.5. Problema de decisión del consumidor representativo

Considere un consumidor-inversionista con preferencias y dotaciones exógenas que toma decisiones en un horizonte infinito. Se supone que el individuo maximiza su satisfacción por un bien genérico de consumo de carácter perecedero. Asimismo, se supone que el individuo es adverso al riesgo y tomador de precios, es decir, no tiene influencia alguna sobre los precios de los activos.

El agente está dotado con una riqueza inicial, a_0 , y elige, en cada instante, de su horizonte de planeación $[0, T]$, la proporción de su riqueza que asignará a la tenencia de un bono libre de riesgo de incumplimiento y la proporción que destinará a un activo riesgoso. La utilidad esperada del tipo von Neumann-Morgenstern, u_0 , al tiempo $t=0$, del individuo, el cual se supone adverso al riesgo y competitivo (tomador de precios), está dada por:

$$\text{Maximizar}_{c_t, \theta_t} \quad E^{\mathbb{P}} \left[\int_0^{\infty} \ln(c_t) e^{-\rho t} dt \mid \mathcal{F}_0 \right]$$

donde ρ es la tasa subjetiva de descuento (o tasa de descuento intertemporal). Esta tasa es un indicador de que tan ansioso está el individuo por el consumo presente. En otras palabras, entre mayor sea ρ , más ansioso estará el agente por consumir en el presente que en el futuro. Como siempre, se supone que la función de utilidad $u(\cdot)$ es dos veces continuamente diferenciable, estrictamente

creciente, estrictamente cóncava y con $\lim_{x \rightarrow 0} u'(x) = \infty$ y $\lim_{x \rightarrow \infty} u'(x) = 0$ (las llamadas condiciones de Inada).

12.6. Condiciones de primer orden

A continuación se plantea el problema de un consumidor que maximiza utilidad por el bien de consumo sujeto a la restricción (7). De esta manera, el problema se plantea como:

$$\begin{aligned} & \text{Maximizar}_{c_t, \theta_t} \quad E^P \left[\int_0^\infty \ln(c_t) e^{-\rho t} dt \mid \mathcal{F}_0 \right] \\ & \text{sujeto a:} \quad da_t = a_t \left(r + \theta_t \lambda \sigma - \frac{c_t}{a_t} \right) dt + a_t \theta_t \sigma dW_t + a_t \theta_t \nu dN_t \end{aligned}$$

En este caso la ecuación de Hamilton-Jacobi-Bellman (H-JB) está dada por (véase, por ejemplo, Venegas-Martínez (2008)):

$$\begin{aligned} 0 = & \ln(c_t) - \rho V(a_t) + V'(a_t) a_t \left(r + \theta_t \lambda \sigma - \frac{c_t}{a_t} \right) + \frac{1}{2} V''(a_t) a_t^2 \theta_t^2 \sigma^2 \\ & + [V[a_t(1 + \theta_t \nu)] - V(a_t)] \phi \end{aligned}$$

Se supone como candidato de solución de la ecuación anterior a $V(a_t) = \beta_0 - \beta_1 \ln(a_t)$, de tal forma que

$$0 = \ln(c_t) - \rho \beta_0 - \rho \beta_1 \ln(a_t) + \beta_1 \left(r + \theta_t \lambda \sigma - \frac{c_t}{a_t} \right) - \frac{1}{2} \beta_1 \theta_t^2 \sigma^2 + \beta_1 \ln(1 + \theta_t \nu) \phi$$

donde

$$\beta_1 = \frac{1}{\rho} .$$

La derivada con respecto θ_t conduce a

$$\lambda \sigma - \theta_t \sigma^2 + \frac{\nu \phi}{1 + \theta_t \nu} = 0 \quad (12.10)$$

Equivalentemente,

$$(1 + \theta_t \nu) \lambda \sigma + \nu \phi = (1 + \theta_t \nu) \theta_t \sigma^2$$

así

$$\theta_t^2 \nu \sigma^2 + \theta_t (\sigma^2 - \nu \lambda \sigma) - (\lambda \sigma + \nu \phi) = 0.$$

Observe que si $\nu = 0$, entonces $\theta_t = \frac{\lambda}{\sigma}$. Si $\nu > 0$, entonces

$$\theta_t = \frac{\nu \lambda \sigma - \sigma^2 + \sqrt{(\nu \lambda \sigma - \sigma^2)^2 + 4\nu \sigma^2 (\lambda \sigma + \nu \phi)}}{2\nu \sigma^2},$$

es decir,

$$\theta_t = \frac{\nu \lambda \sigma - \sigma^2 + \sqrt{(\nu \lambda \sigma + \sigma^2)^2 + 4\phi \nu^2 \sigma^2}}{2\nu \sigma^2}$$

ó

$$\theta_t = \frac{\lambda - \sigma' + \sqrt{(\lambda + \sigma')^2 + 4\phi}}{2\sigma}$$

donde $\sigma' = \sigma / \nu$. Si $\phi = 0$, entonces $\theta_t = \lambda / \sigma$. Sea

$$\Lambda = \frac{\lambda - \sigma' + \sqrt{(\lambda + \sigma')^2 + 4\phi}}{2}. \quad (12.11)$$

De tal forma que

$$\theta_t = \frac{\Lambda}{\sigma}.$$

Observe que Λ satisface la siguiente identidad

$$\Lambda^2 \nu + \Lambda \sigma - \Lambda \nu \lambda - (\lambda \sigma + \nu \phi) = 0.$$

Además, por otro lado

$$da_t = a_t (r + \theta_t \lambda \sigma - \rho) dt + a_t \theta_t \sigma dW_t + a_t \theta_t \nu dN_t$$

Note que, en virtud del lema de Ito,

$$d \ln(a_t) = \left(r + \theta_t \lambda \sigma - \frac{1}{2} \theta_t^2 \sigma^2 - \rho \right) dt + \theta_t \sigma dW_t + \ln(1 + \nu \theta_t) dN_t$$

y

$$E[d \ln(a_t) | \mathcal{F}_t] = \left(r + \theta_t \lambda \sigma - \frac{1}{2} \theta_t^2 \sigma^2 + \phi \ln(1 + \nu \theta_t) \right) dt$$

La derivada igual a cero de la tasa de crecimiento de la riqueza real esperada, expresada en la ecuación anterior, es

$$\lambda\sigma - \theta_t \sigma^2 + \frac{v\phi}{1+v\theta_t} = 0,$$

lo cual coincide plenamente con (12.10). Observe ahora que el proceso de la riqueza se puede reescribir como:

$$da_t = a_t(r + \Lambda\lambda - \rho)dt + a_t\Lambda dW_t + a_t\bar{v} dN_t$$

con

$$\bar{v} = \frac{\Lambda v}{\sigma}.$$

De esta manera,

$$a_t = a_0 e^{(r-\rho+\Lambda\lambda)t} \exp\left\{-\frac{1}{2}\Lambda^2 t + \Lambda W_t + \ln(1+\bar{v})N_t\right\}$$

ó

$$a_t = a_0 \exp\left\{[-\rho + \Lambda(\lambda - \Lambda) + \phi\bar{v}]t\right\} \exp\left\{rt + \frac{1}{2}\Lambda^2 t + \Lambda W_t + \ln(1+\bar{v})N_t - \phi\bar{v}t\right\}.$$

En consecuencia,

$$c_t = \rho a_0 \exp\left\{[-\rho + \Lambda(\lambda - \Lambda) + \phi\bar{v}]t\right\} \exp\left\{rt + \frac{1}{2}\Lambda^2 t + \Lambda W_t + \ln(1+\bar{v})N_t - \phi\bar{v}t\right\}$$

La expresión anterior se puede escribir como

$$c_t = \frac{e^{-\rho t}}{\eta \xi_t} \tag{12.12}$$

con

$$\eta = \frac{1}{\rho a_0}$$

y

$$\begin{aligned} \xi_t &= \exp\left\{(\Lambda(\Lambda - \lambda) - \phi\bar{v})t\right\} \exp\left\{-rt - \frac{1}{2}\Lambda^2 t - \Lambda W_t - \ln(1+\bar{v})N_t + \phi\bar{v}t\right\} \\ &= \exp\left\{(\Lambda(\Lambda - \lambda) - \phi\bar{v})t\right\} e^{-rt} p_t \end{aligned}$$

donde

$$\xi_0 = 1$$

y

$$p_t = \exp \left\{ -\frac{1}{2} \Lambda^2 t - \Lambda W_t - \ln \left(1 + \frac{\Lambda v}{\sigma} \right) N_t + \frac{\Lambda v \phi}{\sigma} t \right\} \quad (12.13)$$

es una martingala. Por lo tanto,

$$\xi_t c_t = \frac{1}{\eta} e^{-\rho t}$$

Note que la ecuación anterior implica que

$$\xi_t a_t = a_0 e^{-\rho t}$$

En consecuencia, se debe cumplir que

$$\lim_{t \rightarrow \infty} \xi_t a_t = \lim_{t \rightarrow \infty} a_0 e^{-\rho t} = 0$$

Observe también que

$$\xi_t = \exp \left\{ \left(\Lambda (\Lambda - \lambda) - \phi \bar{v} \right) t \right\} e^{-r t} p_t$$

de tal manera que

$$d\xi_t = -\xi_t \left[\left(r - \Lambda (\Lambda - \lambda) + \phi \bar{v} \right) dt + \Lambda dW_t + \bar{v} dN_t \right]$$

y

$$d \ln(\xi_t) = \left(-r + \bar{v} \phi + \Lambda \lambda - \frac{1}{2} \Lambda^2 \right) dt - \Lambda dW_t - \ln(1 + \bar{v}) dN_t.$$

Los resultados anteriores proporcionan dos procesos ξ_t y p_t , en la siguiente sección se mostrará que estos no únicos.

12.7. Problema de decisión del consumidor con restricción estática

Considere ahora el problema

$$\begin{aligned} \text{Maximizar}_{c_s} \quad & E^P \left[\int_0^\infty \ln(c_t) e^{-\rho t} dt \mid \mathcal{F}_0 \right] \\ \text{sujeto a:} \quad & E^P \left[\int_0^\infty \xi_t c_t dt \mid \mathcal{F}_0 \right] = \xi_0 a_0. \end{aligned}$$

La restricción en este problema es equivalente a (12.7) (Venegas-Martínez y López-Herrera, 2013). Las condiciones de primer orden para una solución interior del problema anterior se obtienen a través del Lagrangiano

$$\mathcal{L} \equiv \ln(c_t) e^{-\rho t} - \eta \xi_t c_t$$

En este caso, la condición necesaria de máximo está dada por

$$0 = \frac{\partial \mathcal{L}}{\partial c_t} = \frac{1}{c_t} e^{-\rho t} - \eta \xi_t,$$

De esta manera,

$$c_t = \frac{e^{-\rho t}}{\eta \xi_t}$$

lo cual coincide con el resultado (12.12). Observe también que

$$E^P \left[\int_0^\infty \xi_t c_t dt \mid \mathcal{F}_0 \right] = a_0 \int_0^\infty \rho e^{-\rho t} dt = a_0,$$

y

$$\Lambda = \frac{\nu\lambda - \sigma + \sqrt{(\nu\lambda + \sigma)^2 + 4\phi}}{2\nu},$$

lo cual coincide con (12.11). Por otro lado, dado que

$$a_t \xi_t = a_0 e^{-\rho t},$$

se sigue

$$d \ln(a_t) = -d \ln(\xi_t) - \rho dt$$

pero

$$d \ln(\xi_t) = \left(-r + \frac{\Lambda \nu \phi}{\sigma} - \frac{1}{2} \Lambda^2 \right) dt - \Lambda dW_t - \ln \left(1 + \frac{\Lambda \nu}{\sigma} \right) dN_t.$$

En consecuencia,

$$d\ln(a_t) = \left(r + \theta_t \lambda \sigma - \frac{1}{2} \theta_t^2 \sigma^2 - A \right) dt + \theta_t \sigma dW_t + \ln(1 + v\theta_t) dN_t,$$

donde A es una constante por determinar tal que $c_t = Aa_t$. Si se igualan las partes deterministas se sigue que

$$r + \theta_t \lambda \sigma - \frac{1}{2} \theta_t^2 \sigma^2 - A = r - \frac{\Lambda v \phi}{\sigma} + \frac{1}{2} \Lambda^2 - \left[\rho + \Lambda^2 - \frac{1}{\sigma} \Lambda (\sigma \lambda + v \phi) \right]$$

Así,

$$\theta_t \lambda \sigma - \frac{1}{2} \theta_t^2 \sigma^2 = -\frac{\Lambda v \phi}{\sigma} + \frac{1}{2} \Lambda^2 - \Lambda^2 + \frac{1}{\sigma} \Lambda (\sigma \lambda + v \phi) + (A - \rho).$$

Po lo tanto,

$$\theta_t \lambda \sigma - \frac{1}{2} \theta_t^2 \sigma^2 = \Lambda \lambda - \frac{1}{2} \Lambda^2 + (A - \rho)$$

ó

$$\lambda(\theta_t \sigma - \Lambda) = \frac{1}{2} (\theta_t^2 \sigma^2 - \Lambda^2) = \frac{1}{2} (\theta_t \sigma + \Lambda)(\theta_t \sigma - \Lambda) + (A - \rho).$$

Observe que $\theta_t = \Lambda / \sigma$ y $A = \rho$ son solución de la ecuación anterior.

12.8 No unicidad de precios de estado

En esta sección se mostrará la no unicidad de los precios de estado. Para ello, considere un proceso de la forma

$$dP_t = V \phi P_t dt - \Lambda P_t dW_t - V P_t dN_t$$

con

$$V = V(\lambda, v, \sigma, \phi, \Lambda)$$

por determinar. Si recordamos que

$$da_t = (a_t r + a_t \theta_t \sigma \lambda - c_t) dt + a_t \theta_t \sigma dW_t + a_t \theta_t v dN_t$$

y se define $y_t = a_t e^{-rt} P_t$, el valor presente de la riqueza, al aplicar el lema de Itô a y_t , se obtiene que (ver Apéndice)

$$\begin{aligned}
 d(a_t e^{-rt} P_t) &= \left[-ra_t e^{-rt} P_t + e^{-rt} P_t (a_t r + a_t \theta_t \sigma \lambda - c_t) + V \phi e^{-rt} a_t P_t - a_t e^{-rt} P_t \theta_t \sigma \Lambda \right] dt \\
 &\quad + (a_t e^{-rt} P_t \theta_t \sigma - a_t e^{-rt} P_t \Lambda) dW_t + a_t e^{-rt} P_t [(1-V)(1+\theta_t \nu) - 1] dN_t \\
 &= e^{-rt} a_t P_t \theta_t \sigma (\lambda - \Lambda) dt - e^{-rt} P_t c_t dt + V \phi e^{-rt} a_t P_t dt + a_t e^{-rt} P_t (\theta_t \sigma - \Lambda) dW_t \\
 &\quad + a_t e^{-rt} P_t [(1-V)(1+\theta_t \nu) - 1] dN_t
 \end{aligned}$$

Sea $\zeta_t = e^{-rt} P_t$, se mostrará que este precio de estado (una martingala) es diferente de ξ_t , dado en (12.13). En efecto,

$$d(\zeta_t a_t) = -\zeta_t c_t dt + \zeta_t a_t (\theta_t \sigma - \Lambda) dW_t + \zeta_t a_t \theta_t \sigma (\lambda - \Lambda) dt + \zeta_t a_t V \phi dt + \zeta_t a_t [(1-V)(1+\theta_t \nu) - 1] dN_t,$$

ó

$$d(\zeta_t a_t) = -\zeta_t c_t dt + \zeta_t a_t (\theta_t \sigma - \Lambda) dW_t + \zeta_t a_t \theta_t \sigma (\lambda - \Lambda) dt + \zeta_t a_t \{ [\theta_t \nu (1-V) - V] dN_t + V \phi dt \},$$

En consecuencia,

$$d(\xi_t a_t) = -\xi_t c_t dt + \xi_t a_t (\theta_t \sigma - \Lambda) dW_t + \xi_t a_t \theta_t \sigma (\lambda - \Lambda) dt + \xi_t a_t \theta_t \nu (1-V) dN_t - \xi_t a_t V (dN_t - \phi dt),$$

Si $\theta_t = \theta$ es constante y se supone que $\zeta_t a_t = a_0 e^{-\rho t}$ se obtiene que $\lim_{t \rightarrow \infty} \zeta_t a_t = 0$, así

$$-a_0 = -E \left[\int_0^{\infty} \zeta_t c_t dt \mid \mathcal{F}_0 \right] + E \left[\int_0^{\infty} \zeta_t a_t (\theta \sigma - \Lambda) dW_t \mid \mathcal{F}_0 \right] + a_0 \theta \sigma (\lambda - \Lambda) + a_0 \nu \phi \theta (1-V) = 0$$

y

$$E \left[\int_0^{\infty} \zeta_t c_t dt \mid \mathcal{F}_0 \right] = E \left[\int_0^{\infty} \zeta_t a_t (\theta \sigma - \Lambda) dW_t \mid \mathcal{F}_0 \right] + a_0 \{ 1 + \theta [\sigma \lambda - \sigma \Lambda + \nu \phi (1-V)] \}$$

por lo cual se requiere que

$$\sigma \lambda - \sigma \Lambda + \nu \phi (1-V) = 0$$

ó

$$\nu \phi V + \sigma \Lambda - (\sigma \lambda + \nu \phi) = 0.$$

Si se supone que Λ satisface

$$\Lambda^2 \nu - \Lambda \nu \lambda + \Lambda \sigma - (\lambda \sigma + \nu \phi) = 0$$

entonces

$$\nu \phi V = \Lambda^2 \nu - \Lambda \nu \lambda$$

Por lo tanto,

$$\phi V = \Lambda^2 - \Lambda\lambda = \Lambda(\Lambda - \lambda)$$

De esta manera, $V = V(\lambda, \nu, \sigma, \phi, \Lambda)$ está dada por

$$V = \frac{\Lambda(\Lambda - \lambda)}{\phi}.$$

Consecuentemente, $\theta = \Lambda / \sigma$ y

$$\theta^2 \nu \sigma^2 + \theta(\sigma^2 - \nu\lambda\sigma) - (\lambda\sigma + \nu\phi) = 0$$

y

$$dP_t = \Lambda(\Lambda - \lambda)P_t dt - \Lambda P_t dW_t - \frac{\Lambda(\Lambda - \lambda)}{\phi} P_t dN_t$$

Por lo tanto,

$$\mathbb{E} \left[\int_0^\infty \xi_t c_t dt \mid \mathcal{F}_0 \right] = a_0$$

También

$$P_t = \exp \left\{ -\frac{1}{2} \Lambda^2 t - \Lambda W_t - \ln \left(1 + \frac{\Lambda(\Lambda - \lambda)}{\phi} \right) N_t + \Lambda(\Lambda - \lambda)t \right\} \quad (12.14)$$

Por lo anterior, se tiene que

$$\zeta_t = e^{-rt} P_t = \exp \left\{ -rt - \frac{1}{2} \Lambda^2 t - \Lambda W_t - \ln \left(1 + \frac{\Lambda(\Lambda - \lambda)}{\phi} \right) N_t + \Lambda(\Lambda - \lambda)t \right\}.$$

Considere ahora la ecuación diferencial estocástica

$$dP_t = \Lambda(\Lambda - \lambda)P_t dt - \Lambda P_t dW_t - \frac{\Lambda(\Lambda - \lambda)}{\phi} P_t dN_t$$

y como

$$da_t = (a_t r + a_t \theta_t \sigma \lambda - c_t) dt + a_t \theta_t \sigma dW_t + a_t \theta_t \nu dN_t$$

si se define $y_t = a_t e^{-rt} P_t$, el valor presente de la riqueza al aplicar el lema de Itô a y_t , se obtiene que

$$\begin{aligned} d(a_t e^{-rt} P_t) &= \left[-ra_t e^{-rt} P_t + e^{-rt} P_t (a_t r + a_t \theta_t \sigma \lambda - c_t) + \Lambda(\Lambda - \lambda) e^{-rt} a_t P_t - a_t e^{-rt} P_t \theta_t \sigma \Lambda \right] dt \\ &\quad + (a_t e^{-rt} P_t \theta_t \sigma - a_t e^{-rt} \theta_t P_t \Lambda) dW_t + a_t e^{-rt} P_t \left[\left(1 - \frac{\Lambda(\Lambda - \lambda)}{\phi} \right) (1 + \theta_t \nu) - 1 \right] dN_t \\ &= e^{-rt} a_t P_t \theta_t \sigma (\lambda - \Lambda) dt - e^{-rt} P_t c_t dt + \Lambda(\Lambda - \lambda) e^{-rt} a_t P_t dt + a_t e^{-rt} P_t (\theta_t \sigma - \Lambda) dW_t \\ &\quad + a_t e^{-rt} P_t \left[\left(1 - \frac{\Lambda(\Lambda - \lambda)}{\phi} \right) (1 + \theta_t \nu) - 1 \right] dN_t \\ &= e^{-rt} a_t P_t (\theta_t \sigma - \Lambda) (\lambda - \Lambda) dt - e^{-rt} P_t c_t dt + a_t e^{-rt} P_t (\theta_t \sigma - \Lambda) dW_t \\ &\quad + a_t e^{-rt} P_t \left[\theta_t \nu \left(1 - \frac{\Lambda(\Lambda - \lambda)}{\phi} \right) - \frac{\Lambda(\Lambda - \lambda)}{\phi} \right] dN_t \end{aligned}$$

Sea $\zeta_t = e^{-rt} P_t$, entonces

$$\begin{aligned} d(\zeta_t a_t) &= -\zeta_t c_t dt + a_t (\theta_t \sigma - \Lambda) dW_t + \zeta_t a_t \theta_t \sigma (\lambda - \Lambda) (\lambda - \Lambda) dt \\ &\quad + \zeta_t a_t \left[\theta_t \nu \left(1 - \frac{\Lambda(\Lambda - \lambda)}{\phi} \right) - \frac{\Lambda(\Lambda - \lambda)}{\phi} \right] dN_t, \end{aligned}$$

Si $\theta_t = \theta$ es constante

$$\theta \nu \left(1 - \frac{\Lambda(\Lambda - \lambda)}{\phi} \right) - \frac{\Lambda(\Lambda - \lambda)}{\phi} = 0$$

y si $\theta = \frac{\Lambda}{\sigma}$, entonces

$$\frac{\Lambda}{\sigma} \nu \left(\frac{\phi - \Lambda(\Lambda - \lambda)}{\phi} \right) - \frac{\Lambda(\Lambda - \lambda)}{\phi} = 0$$

ó

$$\frac{\Lambda}{\sigma} \nu (\phi - \Lambda(\Lambda - \lambda)) - \Lambda(\Lambda - \lambda) = 0$$

equivalentemente

$$\nu (\phi - \Lambda(\Lambda - \lambda)) - \sigma(\Lambda - \lambda) = 0$$

lo cual implica

$$\nu \phi - \Lambda^2 \nu + \Lambda \nu \lambda - \sigma \Lambda + \sigma \lambda = 0$$

así

$$\Lambda^2 v - \Lambda v \lambda + \Lambda \sigma - (\lambda \sigma + v \phi) = 0,$$

entonces

$$\theta^2 v \sigma^2 + \theta(\sigma^2 - v \lambda \sigma) - (\lambda \sigma + v \phi) = 0$$

Si supone que $\lim_{t \rightarrow \infty} \zeta_t a_t = 0$, se sigue que

$$\mathbb{E} \left[\int_0^\infty \zeta_t c_t dt \mid \mathcal{F}_0 \right] = a_0$$

Por lo tanto,

$$da_t = (a_t r + a_t \theta_t \sigma \lambda - c_t) dt + a_t \theta_t \sigma dW_t + a_t \theta_t v dN_t.$$

De esta manera,

$$c_t = \frac{\rho a_0 e^{-\rho t}}{\zeta_t}$$

lo cual implica

$$c_t = \rho a_0 e^{-\rho t} \exp \left\{ rt + \frac{1}{2} \Lambda^2 t + \Lambda W_t + \ln \left(1 + \frac{\Lambda(\Lambda - \lambda)}{\phi} \right) N_t - \Lambda(\Lambda - \lambda)t \right\}$$

y

$$a_t = a_0 e^{-\rho t} \exp \left\{ rt + \frac{1}{2} \Lambda^2 t + \Lambda W_t + \ln \left(1 + \frac{\Lambda(\Lambda - \lambda)}{\phi} \right) N_t - \Lambda(\Lambda - \lambda)t \right\}.$$

Observe que

$$\zeta_t a_t = a_0 e^{-\rho t}$$

de tal forma que ζ_t proporciona una solución diferente a la de ξ_t obtenida en la sección 7.

12.9 Martingala de mínima entropía

La presente sección obtiene el precio de estado como la martingala que minimiza la entropía cruzada. Sea

$$\frac{dP}{dQ} = \pi_t = \exp \left\{ -\frac{1}{2} \Lambda^2 t + \Lambda W_t + \ln(1 + \bar{v}) N_t - \phi \bar{v} t \right\},$$

la entropía cruzada es igual

$$\begin{aligned} H &= \int \frac{dP}{dQ} \ln \left(\frac{dP}{dQ} \right) dQ = \int \ln \left(\frac{dP}{dQ} \right) dP = E^P \left[\ln \left(\frac{dP}{dQ} \right) \right] = E^P [\ln(\pi_t)] \\ &= E^P \left[-\frac{1}{2} \Lambda^2 t + \Lambda W_t + \ln(1 + \bar{v}) N_t - \phi \bar{v} t \right] \\ &= \left[-\frac{1}{2} \Lambda^2 + \phi \ln(1 + \bar{v}) - \phi \bar{v} \right] t \end{aligned}$$

Si $\bar{v} = \Lambda \nu / \sigma$, entonces $\bar{v} = \theta \nu$, así

$$\frac{H}{t} = -\frac{\theta^2 \sigma^2}{2} + \phi \ln(1 + \nu \theta) - \theta \phi \nu$$

y

$$\frac{\partial}{\partial \theta} \left(\frac{H}{t} \right) = -\theta \sigma^2 + \frac{\nu \phi}{1 + \theta \nu} - \phi \nu = 0$$

Si $\bar{\lambda} = 0$, entonces

$$\lambda \sigma = -\phi \nu$$

y

$$-\theta \sigma^2 + \frac{\nu \phi}{1 + \theta \nu} + \lambda \sigma = 0$$

como en (12.10), en cuyo caso $\theta = \frac{\lambda}{\sigma} - \frac{1}{\nu}$ y $\Lambda = \lambda - \frac{\sigma}{\nu}$, de tal forma que

$$\bar{v} = \theta \nu = \frac{\lambda \nu}{\sigma} - 1,$$

Por lo tanto,

$$\frac{dP}{dQ} = \pi_t = \exp \left\{ -\frac{1}{2} \Lambda^2 t + \Lambda W_t + \ln \left(\frac{\lambda \nu}{\sigma} \right) N_t - \phi \left(\frac{\lambda \nu}{\sigma} - 1 \right) t \right\},$$

así

$$d\pi_t = -\phi \left(\frac{\lambda \nu}{\sigma} - 1 \right) \pi_t dt - \Lambda \pi_t dW_t + \left(\frac{\lambda \nu}{\sigma} - 1 \right) \pi_t dN_t,$$

Si se denota

$$\Lambda = \lambda - \frac{\sigma}{\nu}$$

entonces

$$\pi_t = \exp \left\{ -\frac{1}{2} \Lambda^2 t + \Lambda W_t + \ln(1 + \nu\Lambda) N_t - \phi(\nu\Lambda) t \right\}$$

con $\Lambda = \lambda - \frac{\sigma}{\nu}$, por tanto

$$d\pi_t = -\phi\nu\Lambda \pi_t dt - \Lambda \pi_t dW_t + \nu\Lambda \pi_t dN_t$$

y

$$da_t = (a_t r + a_t \theta_i \sigma \lambda - c_t) dt + a_t \theta_i \sigma dW_t + a_t \theta_i \nu dN_t.$$

Si se define $y_t = a_t e^{-rt} \pi_t$, el valor presente de la riqueza al aplicar el lema de Itô a y_t , se obtiene que (ver Apéndice)

$$\begin{aligned} d(a_t e^{-rt} \pi_t) &= \left[-ra_t e^{-rt} \pi_t + e^{-rt} \pi_t (a_t r + a_t \theta_i \sigma \lambda - c_t) - \nu\Lambda \phi e^{-rt} a_t \pi_t - a_t e^{-rt} \pi_t \theta_i \sigma \Lambda \right] dt \\ &\quad + (a_t e^{-rt} \pi_t \theta_i \sigma - a_t e^{-rt} \pi_t \Lambda) dW_t + a_t e^{-rt} \pi_t [(1 + \nu\Lambda)(1 + \theta_i \nu) - 1] dN_t \\ &= e^{-rt} a_t P_t \theta_i \sigma (\lambda - \Lambda) dt - e^{-rt} \pi_t c_t dt - \nu\Lambda \phi e^{-rt} a_t \pi_t dt + a_t e^{-rt} \pi_t (\theta_i \sigma - \Lambda) dW_t \\ &\quad + a_t e^{-rt} P_t [(1 + \nu\Lambda)(1 + \theta_i \nu) - 1] dN_t \end{aligned}$$

Sea $\chi_t = e^{-rt} \pi_t$, entonces

$$d(\chi_t a_t) = -\chi_t c_t dt + \chi_t a_t (\theta_i \sigma - \Lambda) dW_t + \chi_t a_t \theta_i \sigma (\lambda - \Lambda) dt - \chi_t a_t \nu\Lambda \phi dt + \chi_t a_t [(1 + \nu\Lambda)(1 + \theta_i \nu) - 1] dN_t,$$

ó

$$d(\chi_t a_t) = -\chi_t c_t dt + \chi_t a_t (\theta_i \sigma - \Lambda) dW_t - \chi_t a_t \theta_i \sigma (\lambda - \Lambda) dt + \chi_t a_t \{ [\theta_i \nu (1 + \nu\Lambda) + \nu\Lambda] dN_t - \nu\Lambda \phi dt \},$$

lo cual implica

$$d(\chi_t a_t) = -\chi_t c_t dt + \chi_t a_t (\theta_i \sigma - \Lambda) dW_t + \chi_t a_t \theta_i \sigma (\lambda - \Lambda) dt + \chi_t a_t \theta_i \nu (1 + \nu\Lambda) dN_t + \chi_t a_t \nu\Lambda (dN_t - \phi dt),$$

Si $\theta_i = \theta$ es constante y se supone que $\chi_t a_t = a_0 e^{-\rho t}$ por lo que $\lim_{t \rightarrow \infty} \chi_t a_t = 0$

$$-a_0 = -E \left[\int_0^{\infty} \chi_t c_t dt \mid \mathcal{F}_0 \right] + E \left[\int_0^{\infty} \chi_t a_t (\theta \sigma - \Lambda) dW_t \mid \mathcal{F}_0 \right] + a_0 \theta \sigma (\lambda - \Lambda) + a_0 \nu \phi \theta (1 + \nu\Lambda) = 0$$

así

$$E \left[\int_0^{\infty} \chi_t c_t dt \mid \mathcal{F}_0 \right] = E \left[\int_0^{\infty} \chi_t a_t (\theta \sigma - \Lambda) dW_t \mid \mathcal{F}_0 \right] + a_0 \{ 1 + \theta [\sigma \lambda - \sigma \Lambda + \nu \phi (1 + \nu\Lambda)] \}$$

Con lo cual se requiere que

$$\sigma\lambda - \sigma\Lambda + \nu\phi(1 + \nu\Lambda) = 0$$

Equivalentemente

$$\sigma\lambda - \sigma\Lambda + \nu\phi + \nu^2\phi\Lambda = 0$$

ó

$$-\phi\nu^2\Lambda + \sigma\Lambda - (\sigma\lambda + \nu\phi) = 0$$

Se tiene que cumplir que $\sigma = \phi\nu^2$ ó $\nu = \sqrt{\sigma/\phi}$ en cuyo caso también se cumple que

$$\Lambda^2\nu - \Lambda\nu\lambda + \sigma\Lambda - (\lambda\sigma + \nu\phi) = 0$$

con $\Lambda = \lambda - \frac{\sigma}{\nu} = \bar{\lambda} - \frac{\phi\nu}{\sigma} - \frac{\sigma}{\nu} = \bar{\lambda} - \sqrt{\frac{\phi}{\sigma}} - \sqrt{\phi\sigma}$, así

$$\lambda = \frac{\mu + \delta - \phi\nu - r}{\sigma} = \bar{\lambda} - \sqrt{\frac{\phi}{\sigma}}$$

ó

$$\nu\Lambda = \sqrt{\frac{\sigma}{\phi}}\Lambda = \sqrt{\frac{\sigma}{\phi}}\bar{\lambda} - 1 - \sigma$$

Por lo tanto

$$\bar{\nu} = \frac{\Lambda\nu}{\sigma} = \frac{1}{\sqrt{\sigma\phi}}\Lambda = \frac{1}{\sqrt{\sigma\phi}} - \frac{1}{\sigma} - 1$$

Consecuentemente,

$$\theta = \frac{\bar{\lambda}}{\sigma} - \frac{1}{\sigma}\sqrt{\frac{\phi}{\sigma}} - \sqrt{\frac{\phi}{\sigma}},$$

así

$$d\pi_t = -\phi\left(\sqrt{\frac{\sigma}{\phi}}\bar{\lambda} - \sigma - 1\right)\pi_t dt - \Lambda\pi_t dW_t + \left(\sqrt{\frac{\sigma}{\phi}}\bar{\lambda} - \sigma - 1\right)\pi_t dN_t$$

De esta manera,

$$\pi_t = \exp \left\{ -\frac{1}{2} \Lambda^2 t + \Lambda W_t + \ln \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma \right) N_t - \phi \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma - 1 \right) t \right\}$$

y

$$da_t = (a_t r + a_t \theta_t \sigma \lambda - c_t) dt + a_t \theta_t \sigma dW_t + a_t \theta_t \sqrt{\frac{\sigma}{\phi}} dN_t$$

y se define $y_t = a_t e^{-rt} \pi_t$, el valor presente de la riqueza al aplicar el lema de Itô a y_t , se obtiene que

$$\begin{aligned} d(a_t e^{-rt} \pi_t) &= \left[-ra_t e^{-rt} \pi_t + e^{-rt} \pi_t (a_t r + a_t \theta_t \sigma \lambda - c_t) - \phi \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma - 1 \right) \phi e^{-rt} a_t \pi_t - a_t e^{-rt} \pi_t \theta_t \sigma \Lambda \right] dt \\ &\quad + (a_t e^{-rt} \pi_t \theta_t \sigma - a_t e^{-rt} \pi_t \Lambda) dW_t + a_t e^{-rt} \pi_t \left[\left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma \right) \left(1 + \theta_t \sqrt{\frac{\sigma}{\phi}} \right) - 1 \right] dN_t \\ &= e^{-rt} a_t P_t \theta_t \sigma (\lambda - \Lambda) dt - e^{-rt} \pi_t c_t dt - \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma - 1 \right) \phi e^{-rt} a_t \pi_t dt + a_t e^{-rt} \pi_t (\theta_t \sigma - \Lambda) dW_t \\ &\quad + a_t e^{-rt} P_t \left[\left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma \right) \left(1 + \theta_t \sqrt{\frac{\sigma}{\phi}} \right) - 1 \right] dN_t \end{aligned}$$

Sea $\tilde{\chi}_t = e^{-rt} \pi_t$, entonces

$$\begin{aligned} d(\tilde{\chi}_t a_t) &= -\tilde{\chi}_t c_t dt + \tilde{\chi}_t a_t (\theta_t \sigma - \Lambda) dW_t + \tilde{\chi}_t a_t \theta_t \sigma (\lambda - \Lambda) dt - \tilde{\chi}_t a_t \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma - 1 \right) \phi dt \\ &\quad + \tilde{\chi}_t a_t \left[\theta_t \sqrt{\frac{\sigma}{\phi}} \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma \right) + \sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma - 1 \right] dN_t, \end{aligned}$$

Equivalentemente,

$$\begin{aligned} d(\tilde{\chi}_t a_t) &= -\tilde{\chi}_t c_t dt + \tilde{\chi}_t a_t (\theta_t \sigma - \Lambda) dW_t + \tilde{\chi}_t a_t \theta_t \sigma (\lambda - \Lambda) dt + \tilde{\chi}_t a_t \theta_t \sqrt{\frac{\sigma}{\phi}} \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma \right) dN_t \\ &\quad + \tilde{\chi}_t a_t \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma - 1 \right) (dN_t - \phi dt), \end{aligned}$$

Si $\theta_t = \theta$ es constante y se supone que $\tilde{\chi}_t a_t = a_0 e^{-\rho t}$ por lo que $\lim_{t \rightarrow \infty} \tilde{\chi}_t a_t = 0$

$$-a_0 = -E \left[\int_0^{\infty} \tilde{\chi}_t c_t dt \mid \mathcal{F}_0 \right] + a_0 \theta \sigma (\lambda - \Lambda) + a_0 \nu \phi \theta (1 + \nu \Lambda) = 0$$

$$\mathbb{E} \left[\int_0^{\infty} \tilde{\chi}_t c_t dt \mid \mathcal{F}_0 \right] = \mathbb{E} \left[\int_0^{\infty} \tilde{\chi}_t a_t (\theta \sigma - \Lambda) dW_t \mid \mathcal{F}_0 \right] + a_0 \left\{ 1 + \theta \left[\sigma \lambda - \sigma \Lambda + \phi \sqrt{\frac{\sigma}{\phi}} \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma \right) \right] \right\}$$

Si $\bar{\lambda} = 0$, entonces

$$\begin{aligned} \sigma \lambda - \sigma \Lambda + \phi \sqrt{\frac{\sigma}{\phi}} \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma \right) &= \sigma(\lambda - \Lambda) + \sigma \bar{\lambda} - \sigma \sqrt{\sigma \phi} \\ &= \sigma \sqrt{\sigma \phi} + \sigma \bar{\lambda} - \sigma \sqrt{\sigma \phi} \\ &= \sigma \bar{\lambda} \\ &= 0 \end{aligned}$$

En virtud de que $\mathbb{E} \left[\int_0^{\infty} \tilde{\chi}_t c_t dt \mid \mathcal{F}_0 \right] = a_0$, se sigue

$$da_t = (a_t r + a_t \theta_t \sigma \lambda - c_t) dt + a_t \theta_t \sigma dW_t + a_t \theta_t \sqrt{\frac{\sigma}{\phi}} dN_t$$

así

$$da_t = a_t \left(r + \phi \left(\frac{\lambda v}{\sigma} - 1 \right) - \rho \right) dt + a_t \sqrt{\sigma \phi} \left(\frac{1}{\sigma} + 1 \right) dW_t + a_t \left(\frac{\lambda v}{\sigma} - 1 \right) dN_t$$

La solución de la ecuación anterior es

$$a_t = a_0 e^{(r - \rho + \Lambda \lambda)t} \exp \left\{ -\frac{1}{2} \Lambda^2 t + \Lambda W_t - \ln \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma \right) N_t \right\}$$

lo cual implica que

$$\begin{aligned} a_t &= a_0 \exp \left\{ \left[-\rho + \Lambda(\lambda - \Lambda) + \phi \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma - 1 \right) \right] t \right\} \\ &\quad \times \exp \left\{ rt + \frac{1}{2} \Lambda^2 t + \Lambda W_t + \ln \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma \right) N_t - \phi \left(\sqrt{\frac{\sigma}{\phi}} \bar{\lambda} - \sigma - 1 \right) t \right\}, \end{aligned}$$

con lo que se concluye que

$$c_t = \frac{\rho a_0 e^{-\rho t}}{\bar{\chi}_t}.$$

Esta expresión proporciona la regla óptima de consumo con la martingala de mínima entropía cruzada.

12.10 Conclusiones

En esta investigación se desarrolló un modelo estocástico que describe el proceso de toma de decisiones de un consumidor-inversionista racional en el proceso de integrar un portafolio en un ambiente de riesgo sujeto a dos restricciones: una de tipo presupuestal y otra de sobre la riqueza con un cierto nivel de confianza.

Se mostró la no unicidad de la solución del problema del consumidor racional en presencia de saltos de Poisson. Para ello se demostró que el precio de estado no una martingala única. Finalmente, se obtuvo el precio de estado de mínima entropía cruzada, la cual es una medida de información relativa a una densidad *a priori*.

12.11 Bibliografía

- Ball, C. A., and W. N. Torous (1985). On Jumps in Common Stock Prices and Their Impact on Call Option Pricing. *Journal of Finance*, Vol. 40, No. 1, pp.155-173.
- Cao, M. (2001). Systematic Jump Risks in a Small Open Economy: Simultaneous Equilibrium Valuation of Options on the Market Portfolio and the Exchange Rate. *Journal of International Money and Finance*, Vol. 20, No. 2, pp. 191-218.
- Chandrasekhar Reddy Gukhal, C. R. (2004). The Compound Option Approach to American Options on Jump-Diffusions. *Journal of Economic Dynamics and Control*, Vol. 28, No. 10, pp. 2055-2074.
- Cox, J. C., and S. Ross (1976). The Valuation of Options for Alternative Stochastic Processes. *Journal of Financial Economics*, Vol. 3, No. 1-2, pp. 145-166.
- Gavira-Durón, N. y F. Venegas-Martínez (2011). Decisiones óptimas de consumo y portafolio: un enfoque de precios de estado de Arrow-Debreu. *Revista Contaduría y Administración*, No. 234, pp. 151-172.
- Page, F. H., and A. B. Sanders (1986). A General Derivation of the Jump Process Option Pricing Formula. *Journal of Financial and Quantitative Analysis*, Vol. 21, No. 4, pp. 437-446.
- Venegas-Martínez, F. (2001). Temporary Stabilization: A Stochastic Analysis. *Journal of Economic Dynamics and Control*, Vol. 25, No. 9, pp. 1429-1449.
- Venegas-Martínez, F. (2006a). Stochastic Temporary Stabilization: Undiversifiable Devaluation and Income Risks. *Economic Modelling*, Vol. 23, No. 1, pp. 157-173.
- Venegas-Martínez, F. (2006b). Fiscal Policy in a Stochastic Temporary Stabilization Model: Undiversifiable Devaluation Risk. *Journal of World Economic Review*, Vol. 1, No. 1, pp. 13-38.

Venegas-Martínez, F. (2008). Riesgos Financieros y Económicos. 1ra. edición. Thomson. México.

Venegas-Martínez, F. y F. López-Herrera (2013). Decisiones de consumo y portafolio con un nivel de confianza sobre la riqueza final: horizonte de planeación finito *EconoQuantum*, Revista de Economía y Negocios, por aparecer.

Apéndice

En este apéndice se presenta el lema de Ito para dos procesos de difusión con saltos.

$$dX_t = X_t(\mu_x dt + \sigma_x dW_t + \nu_x dN_t)$$

$$dY_t = Y_t(\mu_y dt + \sigma_y dW_t + \nu_y dN_t)$$

$$z = f(X_t, Y_t, t)$$

$$\begin{aligned} dz = & \left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial X_t} \mu_x X_t + \frac{\partial f}{\partial Y_t} \mu_y Y_t + \frac{1}{2} \frac{\partial^2 f}{\partial X_t^2} \sigma_x^2 X_t^2 + \frac{1}{2} \frac{\partial^2 f}{\partial Y_t^2} \sigma_y^2 Y_t^2 + \frac{\partial^2 f}{\partial X_t \partial Y_t} \sigma_x \sigma_y X_t Y_t \right) dt \\ & + \left(\frac{\partial f}{\partial X_t} \sigma_x X_t + \frac{\partial f}{\partial Y_t} \sigma_y Y_t \right) dW_t \\ & + \left[f(X_t(1 + \nu_x), Y_t(1 + \nu_y), t) - f(X_t, Y_t, t) \right] dN_t. \end{aligned}$$

Capítulo 13

Una Nota sobre el uso de Martingalas en el modelo estándar de mercado para valorar derivados de tasas de interés

Carlos Alexander Grajales Correa

Departamento de Estadística y Matemáticas, Facultad de Ciencias Económicas, Universidad de Antioquia

Francisco Venegas-Martínez

Escuela Superior de Economía, Intituto Politécnico Nacional

Resumen

Este capítulo estudia un derivado conocido como *cap* de tasa de interés, el cual se negocia internacionalmente en el mercado OTC. Se realiza una descripción financiera de dicho instrumento y se desarrollan los detalles matemáticos para calcular su precio a partir de martingalas en el modelo estándar de mercado. Por último, se ilustra el modelo propuesto con un ejemplo de valuación de un contrato *cap* y se examinan algunas medidas de sensibilidad en su precio, haciendo uso del software DerivaGem (un software libre), todo con el fin de generar estrategias útiles para un coberturista o un emisor del producto.

13.1 Introducción

En el modelo desarrollado por Black-Scholes (1973) y Merton (1973) para valuación de opciones europeas se establece un conjunto de supuestos, entre los cuales se plantea que la tasa de interés libre de riesgo (de incumplimiento) permanece constante durante el período de vida del instrumento financiero. Además tal valuación parte del principio conocido como valuación neutral al riesgo. Sin embargo, relajar el supuesto de tasa constante está más acorde con la realidad contingente observada, sobre la cual se entiende que la dinámica de tasas obedece a una naturaleza estocástica. Esto conduce a nuevas estrategias de cobertura de portafolios y a nuevas formulaciones para valorar derivados sobre tasas de interés.

Así, en este nuevo ambiente, se hace necesario subrayar la importancia de un resultado heredado de la teoría de la medida, llamado medida equivalente de martingala, con la cual se valua un derivado financiero, f , a través de una extensión del principio de valuación tradicional en riesgo neutral y del establecimiento de una unidad de medida, dada por el

precio de un activo financiero, g , llamada numerario. Por otro lado, un proceso estocástico, X , es una martingala si tiene tendencia igual a cero, esto es,

$$dX = \sigma dw,$$

con σ constante y w un proceso de Wiener o movimiento browniano. De esta manera, el principio mencionado establece que la razón f/g será una martingala en algún “mundo” definido por la prima de riesgo del activo subyacente del derivado f (Shreve, 2005; Svoboda 2004).

En este sentido, este trabajo estudia un derivado financiero conocido como *cap* de tasas de interés, negociado internacionalmente, en el mercado OTC, y el cual forma parte del mercado de derivados más grande del mundo, el de tasas de interés (para más detalles véase Venegas-Martínez (2008)). Un contrato *cap* es un contrato financiero que se cotiza internacionalmente sobre mostrador y ofrece cobertura contra la eventualidad que una tasa flotante, en una nota de tasa flotante, resulte mayor a cierto nivel. De esta manera, para un derivado f se desarrolla un argumento que hace uso de una martingala especial, f/g , consistente con el precio de un bono cero cupón, g , para valorar de un contrato *cap* con el modelo estándar de mercado, lo cual generaliza el modelo de Black (1976). Al final, en coherencia con lo anterior, se desarrolla un ejemplo de valuación de un *cap* y se identifican gráficamente algunas medidas de sensibilidad en su precio, esto mediante el software DerivaGem, todo con el fin de generar estrategias para un coberturista o un emisor del producto.

13.2 Martingalas y numerarios

Suponga, en primer lugar, que un derivado, f , depende de un solo activo subyacente, θ , que no paga dividendos, y que sus dinámicas están dadas por

$$\begin{aligned}d\theta &= m\theta dt + s\theta dw \\df &= \mu f dt + \sigma f dw\end{aligned}$$

donde m, s, μ, σ son funciones de θ y t ; m y μ representan, respectivamente, las tendencias de θ y f , mientras que s y σ son las volatilidades de las mismas. Con base en lo anterior, puede mostrarse que cualquier derivado f satisface la relación

$$\mu - r = \lambda \sigma.$$

La constante λ , así definida, se conoce como la prima de riesgo de mercado de θ , y es útil para definir el exceso de rentabilidad esperada del derivado respecto a la tasa libre de riesgo, por unidad de volatilidad, σ . Un mundo tradicional de riesgo neutral se define como aquel donde λ es cero, y así $\mu = r$. De manera alternativa, puede escogerse un mundo en el cual λ es diferente de cero, con lo cual se redefiniría la tasa de crecimiento de f pero no su volatilidad. La elección del valor de λ es también conocida como cambio de medida. Algún valor de λ conduce al mundo real, mientras que si λ se hace igual a la volatilidad del numerario, el mundo de destino es llamado mundo forward de riesgo neutral respecto al numerario, y es así como se puede viajar a través de diferentes mundos. De este modo,

el resultado de medida equivalente de martingala establece que, si λ es la volatilidad del numerario, entonces el cociente f/g es una martingala para todo f , esto es,

$$f_0 = g_0 \mathbb{E}_g \left[\frac{f_T}{g_T} \right] \quad (1)$$

donde los subíndices 0 y T indican, respectivamente, el valor del activo correspondiente en el tiempo presente $t = 0$, y en el tiempo posterior $t = T$, y \mathbb{E}_g es el operador esperanza en el mundo definido por la volatilidad del numerario.

Típicamente, se toma el numerario g como una cuenta monetaria en la cual una unidad monetaria, $g_0 = 1$, se capitaliza a

$$g_T = 1 \times \exp(r_t T)$$

en un tiempo T , donde r_t , en primera instancia, es determinista. Sin embargo, si se supone una tasa r_t estocástica, el numerario adopta la forma

$$dg = r g dt,$$

y así,

$$\begin{aligned} g_T &= \exp \left(\int_0^T r dt \right) \\ &= \exp(\bar{r}T). \end{aligned}$$

Finalmente, de acuerdo con (1), como la volatilidad del numerario es cero, significa que para un derivado f , f/g es martingala en un mundo tradicional de riesgo neutral, esto es,

$$f_0 = \hat{\mathbb{E}}[\exp(-\bar{r}T)],$$

donde se ha cambiado por conveniencia de identificación, la notación \mathbb{E}_g por $\hat{\mathbb{E}}$. Es en este ambiente donde usualmente se valúa el precio futuro, F_{fut} , del activo subyacente en un contrato de futuro; también se valúan opciones sobre acciones, índices o divisas. Este tipo de contratos se negocia frecuentemente en bolsa (un mercado listado reconocido por la autoridades financieras para negociar dichos contratos).

De manera alternativa, puede elegirse el numerario g como el precio al tiempo t de un bono cupón cero, $P(t, T)$, que paga una unidad monetaria en el vencimiento en T . Consecuentemente con (1),

$$f_0 = P(0, T) E_T[f_T],$$

donde se ha cambiado por conveniencia de identificación, la notación E_g por E_T . En este ambiente, se valúa por ejemplo, el precio *forward*, F_{fw} , del activo subyacente a un contrato *forward* o el precio de una opción sobre un bono. Sin embargo, un caso de especial interés se presenta cuando el derivado f es una tasa *forward* de interés. Si en el tiempo t , la tasa es $R(t, T, T^*)$, aplicable y compuesta en un periodo posterior $[T, T^*]$, y si $P(t, T^*)/P(t, T)$ es el precio *forward* de un bono cupón cero en el mismo período, entonces puede mostrarse

que la tasa forward es una martingala en un mundo *forward* de riesgo neutral con respecto del precio del bono cero cupón $g_t = P(t, T^*)$, esto es,

$$R(0, T, T^*) = E_{T^*} [f_T/g_T],$$

donde

$$f_t = [P(t, T) - P(t, T^*)] / (T^* - T),$$

y se ha cambiado, por conveniencia, la notación E_g por E_{T^*} . Este caso particular resulta ser conveniente para valuar contratos *cap* y *floor*, los cuales, respectivamente, son útiles para cubrirse, en un tiempo $t < T$, contra una subida o una bajada significativa de la tasa realizada $R(T, T, T^*)$ con respecto de un nivel de tasa fijo R_K . Estos contratos (además de opciones sobre *swaps* que no se exponen aquí) son ampliamente negociados en los mercados internacionales sobre mostrador.

13.3 Caso de valuación de un contrato cap

En esta sección se desarrolla una ilustración para valuar un contrato *cap*. Se puede ver que los resultados obtenidos son consistentes con el modelo de Black (1976). En este caso, la tasa fija R_K se conoce como tasa *cap*, y la tasa realizada, que define el referente de comparación con R_K y que tiene naturaleza estocástica, usualmente se asocia a la tasa *LIBOR* observada con una periodicidad determinada; esta periodicidad es llamada tenor.

Así, para el diseño de un contrato *cap*, los elementos básicos a definir son R_K , el tenor, la tasa de comparación R_k (*LIBOR*) y la vida del contrato T . Si el contrato inicia en la fecha t_1 y paga en las fechas

$$t_2, t_3, \dots, t_n, t_{n+1} = t_1 + T,$$

entonces para una posición larga el contrato pagará en t_{k+1} la cantidad, conocida como *caplet*,

$$c_{k+1, k+1} = L\tau (R_k - R_K)^+, \quad k = 1, \dots, n,$$

donde

$$\tau = t_{k+1} - t_k,$$

es el tenor, L es el capital principal sobre el cual se gana interés, y

$$R_k = R(t_k, t_k, t_{k+1}).$$

Así, el pago $c_{t, k+1}$, visto en un tiempo $t < t_k$, puede traducirse como una opción call europea sobre la *LIBOR* observada en el vencimiento t_k , y donde el precio *strike* es R_K . En suma, el *cap* es un portafolio de n opciones call europeas con las características descritas. Ahora bien para valuar el *caplet* $c_{t, k+1}$, en el tiempo t , se hace uso del numerario

$$g_t = P(t, t_{k+1}),$$

y se establece así que la razón c/g es una martingala en un mundo *forward* de riesgo neutral con respecto a g . Por lo tanto, de acuerdo con (1),

$$c_{0, k+1} = P(0, t_{k+1}) E_{k+1} [c_{k+1, k+1}].$$

El valor de $c_{0,k+1}$ se obtiene mediante la siguiente secuencia de igualdades (Hull y White 1990):

$$\begin{aligned} c_{0,k+1} &= L\tau P(0, t_{k+1}) E_{k+1} \left[(R_k - R_K)^+ \right] \\ &= L\tau P(0, t_{k+1}) [E_{k+1}(R_k) N(d_1) - R_K N(d_2)] \\ &= L\tau P(0, t_{k+1}) [E_{k+1}(R(t_k, t_k, t_{k+1})) N(d_1) - R_K N(d_2)] \\ &= L\tau P(0, t_{k+1}) [(R(t_0, t_k, t_{k+1})) N(d_1) - R_K N(d_2)] \\ &= L\tau P(0, t_{k+1}) [F_k N(d_1) - R_K N(d_2)] \end{aligned}$$

donde F_k es la tasa forward dada por $R(t_0, t_k, t_{k+1})$, que en la práctica se calcula a partir de la curva cero *LIBOR/SWAP*, denotada por $r(t)$, y

$$d_1 = [\ln(F_k/R_K) + \sigma_k^2 t_k / 2] / (\sigma_k \sqrt{t_k});$$

$$d_2 = d_1 - \sigma_k \sqrt{t_k};$$

y $\sigma_k \sqrt{t_k}$ es la volatilidad de la variable aleatoria $\ln(R_k)$ cuando ésta se supone lognormal. Finalmente, si $r(t)$ está expresada en composición anual compuesta continua, haciendo la conversión de tasas se tendrá que,

$$F_k = [\exp(r(t_{k+1})t_{k+1} - r(t_k)t_k) - 1] / (t_{k+1} - t_k).$$

En la literatura, a la metodología de valuación presentada se le conoce como modelo estándar de mercado, en este caso, para un contrato *cap* (Hull, 2012; Brigo y Mercurio, 2006).

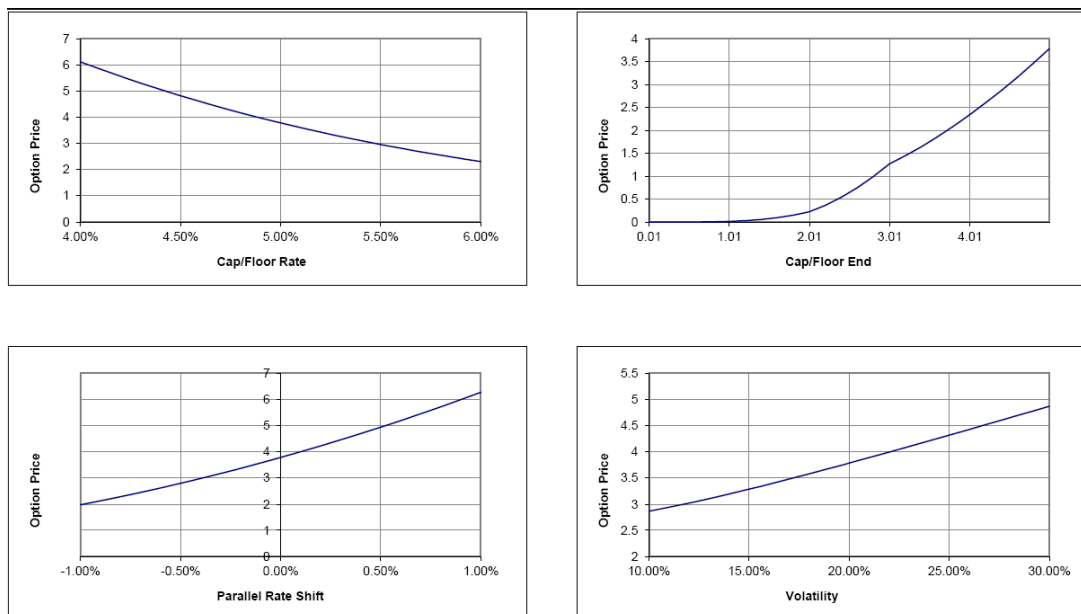
En relación a lo anterior, el software *DerivaGem* Hull(2012) es una calculadora en Excel, de distribución libre, sobre derivados financieros que puede descargarse del sitio web: www.rotman.utoronto.ca/~hull/software/. El método de valuación presentado se identifica en el software bajo el nombre *Pricing Model Black European*.

Para ilustrar numéricamente la valuación, y de acuerdo a la notación anterior, considere que se tienen los siguientes datos: $L = 100$, $t_1 = 0$, $T = 5$ años, $R_K = 5\%$, $\sigma_k = 20\%$, y un tenor trimestral. La curva cero $(t, r(t))$ que se supone en este ejemplo es:

$$(0, 3.0\%), (1, 3.6\%), (2, 4.0\%), (3, 4.6\%), (4, 4.9\%), (5, 5.20\%).$$

El valor de la prima que se obtiene es aproximadamente 3.7850, con lo cual se cubre cada uno de los *caplets* del *cap* frente a la eventualidad, que la tasa LIBOR, en cada fecha de pago, esté por encima del nivel 5%.

Otros resultados del software se tienen en la Gráfica 1 donde se muestra de izquierda a derecha y de arriba a abajo, la valuación del contrato, así como la sensibilidad en su precio con respecto de la variación de la tasa *cap*, R_K , del tiempo, de la curva cero cupón cuando ésta se perturba por un desplazamiento paralelo, con respecto de la volatilidad σ_k . Cada resultado refleja una relación no lineal entre las variables señaladas y deja ver que, para el comprador, el contrato ofrece una estrategia de cobertura contra la eventualidad de tener una tasa flotante mayor a R_K en una nota de deuda flotante.

Gráfica 13.1. Valuación y sensibilidad de un *cap*.

13.4 Conclusiones

Dada la necesidad de valorar un derivado f que depende de la naturaleza estocástica de la tasa de la interés, se ha expuesto un argumento que muestra cómo partiendo del principio de valuación forward en riesgo neutral, con respecto a un numerario g , se consigue que la razón f/g sea una martingala en un mundo donde g es una cuenta monetaria, o bien el precio de un bono cero cupón. En este último caso, y en particular, se muestra que la martingala referida proporciona una formulación analítica y comprensible en la valuación un contrato *cap* de tasas.

En la aplicación presentada, en el transcurso de la sección 13.3, se dejan ver los altos riesgos delta y vega, los cuales deberían cubrirse. Una posible metodología para realizar tal cobertura está en la agenda de trabajo.

Por último se ha mostrado que el software DerivaGem es útil para valorar un contrato *cap* y para explorar algunas medidas de sensibilidad en su precio, los cuales pueden proporcionar información valiosa a un coberturista o a un emisor del producto.

13.5 Bibliografía

- Black, F. and M. Scholes (1973). "The Pricing of Options and Corporate Liabilities". *The Journal of Political Economy*, Vol. 81, No. 3, pp. 637-654.
- Black, F. (1976). "The Pricing of Commodity Contracts". *Journal of Financial Economics*, Vol. 3, No. 1-2, pp. 167-79.

- Brigo, D. and F. Mercurio (2006). Interest Rate Models, Theory and Practice, with Smile, Inflation and Credit. Springer.
- Hull, J. and A. White (1990). "Pricing Interest Rate Derivative Securities", *The Review of Financial Studies*, Vol. 3, No. 4, pp. 573-92.
- Hull, J. (2012). Options, Futures and Other Derivatives. Prentice Hall, eighth edition.
- Merton, R. (1973). "Theory of Rational Option Pricing". *Bell Journal of Economics and Management Science*, Vol 4, No. 1, pp. 141-83.
- Shreve, S. (2005). Stochastic Calculus for Finance I and II. Springer.
- Svoboda, S. (2004). Interest Rate Modelling. Palgrave, Macmillan.
- Venegas-Martínez, F. (2008). Riesgos Financieros y Económicos, 2da. edición. Cengage. México.

Capítulo 14

Utilidad Diferencial Recursiva Estocástica (UDRE) vs. Hamilton-Jacobi-Bellman (HJB)

Francisco Venegas-Martínez

Escuela Superior de Economía, Instituto Politécnico Nacional

Isela Elizabeth Téllez-León

Escuela Superior de Economía, Instituto Politécnico Nacional

Francisco Ortiz Arango

Escuela de Ciencias Económicas y Empresariales, Universidad Panamericana

Resumen

En este capítulo se utiliza el concepto de utilidad diferencial recursiva estocástica (UDRE), desarrollado por Duffie y Epstein (1991), para permitir separar en las decisiones de consumo y portafolio los efectos de la elasticidad de sustitución intertemporal de consumo y la aversión relativa al riesgo. Particularmente, se muestra que en las decisiones de consumo e inversión en un portafolio que contiene un activo riesgoso, UDRE sí permite que en la proporción de la riqueza real que se destina al activo riesgo la aversión al riesgo tenga un efecto negativo. Asimismo, la regla óptima de consumo es afectada por la aversión al riesgo.

14.1 Introducción

Las teorías de valuación de activos y de decisiones de consumo y portafolio basada, tradicionalmente, en la utilidad esperada, aditiva e intertemporal de Von Neumann-Morgenstern, en tiempo continuo, cuentan con una alternativa muy atractiva para la investigación, la utilidad recursiva, la cual se encuentra formalmente definida en el artículo “Asset Pricing with Stochastic Differential Utility” de Darel Duffie y Larry G. Epstein en 1991. Este trabajo fue ampliado en 1992 con la participación de Costis Skiadas, y lleva por título “Stochastic Differential Utility”.

Los modelos convencionales de valuación de activos y de decisiones de consumo y portafolio, en economía financiera, basado en el consumo (CCAPM) de Lucas (1978) y Breeden (1979) suponen que las preferencias de los agentes tienen una representación del tipo Von Neumann-Morgenstern. Estos modelos han sido criticados por dos razones. En primer lugar, porque se tiene evidencia de que el modelo no funcionan bien empíricamente; véanse, por ejemplo, Hansen y Singleton (1983); Mehra y Prescott (1985);

Mankiw and Shapiro (1986), y Breeden, Gibbons, y Litzenberger (1989). En segundo lugar, la especificación de la utilidad confunde la aversión al riesgo y la sustitución intertemporal dentro del índice de satisfacción instantánea. Con base en estos dos inconvenientes, Epstein y Weil (1990) propusieron el concepto de utilidad intertemporal recursiva en tiempo discreto. Esas funciones de utilidad permiten la separación entre la sustitución intertemporal por el consumo y la aversión al riesgo; esta separación ha sido aprovechada por Epstein (1988), Campbell (1990) y Kandel y Stambaugh (1990) en varias extensiones teóricas, aprovechando que esto también conlleva a relaciones entre los rendimientos de los activos y la tasa de consumo, lo que concuerda con los datos observados; véanse, al respecto, Epstein y Zin (1990, 1991a, 1991b), Giovannini y Weil (1989), y Bufman y Leiderman (1990).

Por otra parte, la separación entre aversión al riesgo y sustitución intertemporal se logran a un costo razonable, en términos de la relajación necesaria de los axiomas de la teoría intertemporal de la utilidad esperada. Duffie y Epstein (1992) formulan y analizan la utilidad recursiva en tiempo continuo, la cual llamaron utilidad diferencial estocástica. En su artículo: “Asset Pricing with Stochastic Differential Utility” obtienen mediante la aplicación de un marco de agente-representativo una serie de nuevos modelos de valuación de activos y de decisiones de consumo e inversión en portafolios. También aprovechan la potencia del análisis en tiempo continuo para desarrollar extensiones alternativas del CCAPM de Breeden (1979).

Si se supone que la función de utilidad del agente representativo es homotética, se tiene que el exceso de rendimientos esperados de los activos son obtenidos por un modelo lineal de dos factores (los factores son el crecimiento del consumo agregado y la rentabilidad de un portafolio de mercado). Es decir, se tiene una combinación lineal del CCAPM y el CAPM estático de Sharpe (1964) y Lintner (1965). Al respecto, un estudio en tiempo discreto que supone que el consumo y la rentabilidad de los activos se distribuyen en forma conjunta lognormal se encuentra en Epstein y Zin (1991a).

Por otro lado, el modelo de dos factores es susceptible a la crítica de Roll (1977) sobre la cartera agregada del CAPM, porque presuntamente incluye activos no comercializables, lo que implica que los rendimientos del portafolio son imperfectamente observables. Por lo tanto, en “Asset Pricing with Stochastic Differential Utility”, se describe una extensión alternativa al CCAPM que es inmune a dicha crítica. Para ello se requiere que el consumo en cada instante tenga al menos dos componentes, y con ello se obtiene (bajo determinados supuestos) un modelo multifactorial del exceso de rendimientos esperados que se distingue del CCAPM por la inclusión de un precio relativo de los factores.

Por último, en el modelo propuesto en “Asset Pricing with Stochastic Differential Utility” se destacan las ventajas teóricas de la flexibilidad ofrecida por la utilidad diferencial estocástica al separar la aversión al riesgo y la sustitución intertemporal. Esto se realiza mediante una generalización del modelo de Cox, Ingersoll, y Ross (1985a). En este sentido, se puede caracterizar la respuesta de la estructura de plazos de la tasa de interés a variaciones en la aversión al riesgo. Se destaca que tal caracterización no es posible bajo la especificación de la utilidad aditiva Von Neumann-Morgenstern adoptada por Cox, Ingersoll, y Ross, en donde un cambio en la concavidad de la función de utilidad instantánea afecta tanto a la aversión al riesgo como al grado de sustitución intertemporal, por lo que la interpretación de los resultados correspondientes de estática comparativa son ambiguos.

En “Asset Pricing with Stochastic Differential Utility” se incluye un resultado adicional de valuación de activos como una extensión de la fórmula usual de valor descontado-esperado para valuar flujos de dividendos. Esta fórmula se utiliza en Duffie y Epstein (1992) para el término de estructura, pero es de interés más amplio en la literatura especializada.

La literatura que estudia las estructuras alternativas de preferencias generalizadas y las implicaciones de valuación de activos pueden encontrarse en Singleton (1990), Sundaresan (1989), Constantinides (1990), y Huang (1987) Hindy (1990), Detemple y Zapatero (1989) y Heaton (1991), entre otros. Estos autores estudian un concepto de utilidad que relaja el supuesto de aditividad intertemporal, por ejemplo, la formación de hábito al consumo o la durabilidad de los bienes de consumo. Una consecuencia es que el consumo pasado afecta directamente a las preferencias actuales, mientras que para la utilidad recursiva, el consumo pasado influye en las decisiones actuales sólo a través de su efecto sobre las actuales. Aunque estas rutas alternativas para generalizar utilidad intertemporal tienen un considerable interés en la investigación no proporcionan una separación entre la sustitución intertemporal y la aversión al riesgo (Epstein, 1992), ni

permiten obtener modelos de determinación de precios de nuevos activos como se describe en Duffie y Epstein (1992). Se trata pues de una cuestión empírica que sigue sin resolverse, de cualquier forma se busca que la utilidad recursiva o una de las alternativas anteriores, o quizás una combinación adecuada, explique mejor el comportamiento observado en el consumo y el rendimiento de los activos. Por último, es importante mencionar otros trabajos que aplican el concepto de utilidad diferencial recursiva estocástica en crecimiento económico son los de Chatterjee, Giuliano y Turnovsky (2003).

En este capítulo se utiliza el concepto de utilidad diferencial recursiva estocástica (UDRE), desarrollado por Duffie y Epstein (1991), para examinar en las decisiones de un consumidor-inversionista racional. Particularmente, se muestra que en las decisiones de inversión en un portafolio que contiene un activo riesgoso, UDRE sí afecta negativamente la proporción de la riqueza real que se destina al activo riesgo la aversión al riesgo tenga un efecto negativo. De la misma manera, la regla óptima de consumo es afectada por la aversión al riesgo; la aversión al riesgo aparece en la propensión marginal al consumo del individuo.

Este capítulo está organizado como sigue: en la sección 14.2 se introduce el coeficiente de aversión relativa al riesgo y el coeficiente de Arrow-Pratt; en el transcurso de la sección 14.3 se discute sobre la elasticidad de sustitución intertemporal; a través de la sección 14.4 se plantea el problema control del consumidor racional; en la sección 14.5 se introduce el concepto de utilidad diferencial recursiva estocástica; en la sección 14.6 se resuelve el problema del consumidor con el enfoque tradicional de Hamilton-Jacobi-Bellman; a través de la sección 14.7 se realiza un análisis comparativo entre las soluciones obtenidas por ambos métodos; y, por último, en la sección 14.8, se presentan las conclusiones.

14.2 Coeficiente de aversión relativa al riesgo

El coeficiente de aversión relativa al riesgo, R , mide la curvatura de la función de utilidad. La concavidad será mayor cuanto mayor sea la segunda derivada en valor absoluto de la función de utilidad. Por tanto, la aversión al riesgo será mayor cuanto mayor sea la segunda derivada. No obstante, la aversión al riesgo no se puede medir exclusivamente por la segunda derivada, se debe tener también en cuenta la primera derivada que mide la utilidad marginal.

Considere un consumidor racional que desea maximizar su satisfacción por el consumo de un bien genérico, c_t , con función de utilidad

$$u(c_t) = \frac{c_t^{1-\frac{1}{\varepsilon}}}{1-\frac{1}{\varepsilon}}, \quad \varepsilon > 0$$

donde ε es un parámetro de preferencia. Observe que la tasa de sustitución intertemporal está dada por

$$R = -\frac{\frac{du'(c_t)}{dc_t}}{\frac{u'(c_t)}{c_t}} = -\frac{du'(c_t)}{dc_t} \frac{c_t}{u'(c_t)} = -\frac{u''(c_t)c_t}{u'(c_t)} = \frac{1}{\varepsilon}$$

ya que

$$u'(c_t) = c_t^{-\frac{1}{\varepsilon}} \quad y \quad u''(c_t) = -\frac{1}{\varepsilon} c_t^{-(1+\frac{1}{\varepsilon})}$$

Note que la expansión en serie de Taylor de $u(c_t)$ alrededor de $E[c_t]$ satisface

$$u(c_t) = u(E[c_t]) + u'(E[c_t])(c_t - E[c_t]) + \frac{1}{2}u''(E[c_t])(c_t - E[c_t])^2 + \dots$$

Si se toman valores esperados en la expresión anterior se sigue que

$$\begin{aligned} E[u(c_t)] &= u(E[c_t]) + u'(E[c_t])E[(c_t - E[c_t])] + \frac{1}{2}u''(E[c_t])E[(c_t - E[c_t])^2] + \dots \\ &= u(E[c_t]) + \frac{1}{2}u''(E[c_t])\text{Var}[c_t] + \dots \end{aligned}$$

ya que

$$E[(c_t - E[c_t])] = 0$$

si $u''(E[c_t]) < 0$, entonces

$$E[u(c_t)] < u(E[c_t])$$

Si $u'(c_t) > 0$, es decir, si $u(c_t)$ es estrictamente creciente, entonces existe $\varphi = \varphi(t) > 0$ (porque $E[c_t]$ depende de t) tal que

$$E[u(c_t)] = u(E[c_t] - \varphi(t)) < u(E[c_t])$$

Además,

$$u(E[c_t] - \varphi(t)) - u(E[c_t]) \approx -u'(E[c_t])\varphi(t).$$

Por lo tanto

$$E[u(c_t)] \approx u'(E[c_t])\varphi(t) + u(E[c_t] - \varphi(t)) + \frac{1}{2}u''(E[c_t])\text{Var}[c_t]$$

ó

$$u'(E[c_t])\varphi(t) \approx -\frac{1}{2}u''(E[c_t])\text{Var}[c_t]$$

donde

$$\varphi(t) = \frac{1}{2}\bar{R}(t)\Psi(T)$$

y

$$\Psi(t) = \frac{\text{Var}[c_t]}{E[c_t]}$$

es el coeficiente de variación y

$$\bar{R}(t) = -\frac{u''(E[c_t])E[c_t]}{u'(E[c_t])}$$

es el coeficiente de aversión relativa al riesgo del consumo esperado. Lo anterior conlleva a definir

$$R = -\frac{u''(\hat{c}_t)\hat{c}_t}{u'(\hat{c}_t)}$$

donde $\hat{c}_t \equiv E[c_t]$. Cuando esta razón se mantiene constante es llamada el coeficiente de aversión relativa al riesgo. Es importante mencionar, al respecto, que Constantinides, Donaldson, and Mehra, (2002) encuentran evidencia empírica de que R es mayor a 1 .

14.2.1 Coeficiente de Arrow-Pratt

El análisis de las decisiones económicas en situaciones de incertidumbre requiere de medidas cuantitativas de la actitud de un individuo racional frente al riesgo. Las más utilizadas son conocidas como “coeficientes de Arrow-Pratt”, que fueron formuladas independientemente y por primera vez por Kenneth Arrow y John Pratt y consisten en los siguientes cocientes:

$$\kappa = \frac{R}{\hat{c}_t} = -\frac{u''(\hat{c}_t)}{u'(\hat{c}_t)}$$

14.3 Elasticidad de sustitución intertemporal

La elasticidad de sustitución intertemporal es uno de los parámetros de preferencias clave en los modelos macroeconómicos intertemporales. Diversos estudios han puesto de manifiesto una posible subestimación de ésta para el caso de distintos países. Es práctica habitual estimar el citado parámetro empleando únicamente datos de consumo de bienes no duraderos y servicios, omitiendo los flujos de servicios que el consumo duradero genera. Este modo de proceder sólo es admisible si la utilidad intratemporal es separable entre los diferentes componentes del consumo. Contrastar tal separabilidad para el caso español es uno de los objetivos de este trabajo, además de analizar la incidencia que la consideración del consumo duradero tiene sobre los valores estimados de la elasticidad de sustitución intertemporal. Epstein and Zin (1991a) encuentran evidencia empírica de que ε es menor que 1. La elasticidad de sustitución intertemporal, ζ , se define como

$$\frac{\frac{d(c_{t_2}/c_{t_1})}{c_{t_2}/c_{t_1}}{d(p_{t_2}/p_{t_1})/p_{t_2}/p_{t_1}} = -\frac{\frac{d(c_{t_2}/c_{t_1})}{c_{t_2}/c_{t_1}}{d(u'(c_{t_2})/u'(c_{t_1}))}}{u'(c_{t_2})/u'(c_{t_1})} = -\frac{d \ln(c_{t_2}/c_{t_1})}{d \ln(u'(c_{t_2})/u'(c_{t_1}))} = \varepsilon$$

ya que

$$d \ln(u'(c_{t_2}) / u'(c_{t_1})) = d \ln(c_{t_2}^{-\frac{1}{\varepsilon}} / c_{t_1}^{-\frac{1}{\varepsilon}}) = -\frac{1}{\varepsilon} \ln(c_{t_2} / c_{t_1}).$$

Esto hace que exista confusión entre la tasa de sustitución intertemporal, ε , y la aversión al riesgo, $1/\varepsilon$. A continuación se presenta una alternativa de función de utilidad total descontada que enmienda esta limitación

14.4 Problema del consumidor racional

Considere un consumidor racional que se desea resolver el problema

$$\text{Maxizar } E_0 \left[\int_0^{\infty} \frac{c_t^{1-1/\varepsilon}}{1-R} e^{-\rho t} dt \right]_{c_t}$$

$$\text{sujeto a: } da_t = a_t(1-\theta_t)dR_b + a_t\theta_t dR_s - c_t dt$$

donde a_t es la riqueza real del individuo, c_t es el consumo, θ_t es la proporción de la riqueza que asigna a la tenencia de un activo riesgoso y $1-\theta_t$ es la proporción de la riqueza destinada a la tenencia bonos libres de riesgo de incumplimiento. Para más detalles véase Venegas-Martínez (2008). Los rendimientos de los activos son: para el bono

$$dR_b = \frac{db_t}{b_t} = r dt$$

y para el activo riesgoso

$$dR_s = \frac{dS_t}{S_t} = \mu dt + \sigma dW_t$$

donde

$$dW_t \sim N(0, dt).$$

Si se sustituyen los rendimientos de los activos en la restricción presupuestal se obtiene:

$$da_t = a_t \left(r + \theta_t(\mu - r) - \frac{c_t}{a_t} \right) dt + a_t \theta_t \sigma dW_t$$

ó

$$da_t = a_t \left(g(\theta_t) - \frac{c_t}{a_t} \right) dt + a_t \theta_t \sigma dW_t$$

donde

$$g(\theta_t) = r + \theta_t(\mu - r)$$

14.5 Utilidad diferencial recursiva estocástica

En esta sección se introduce el concepto de utilidad diferencial recursiva estocástica (UDRE) que dista del enfoque tradicional de Hamilton-Jacobo-Bellman en la definición de la función de valor y en el uso de aproximaciones a la función de densidad exponencial y a la de un binomio elevado a una fracción. Considere un consumidor racional que desea tomar decisiones de consumo y portafolio con base en la siguiente función de valor:

$$J(a_t) = \max_{c_s | s \in [t, \infty)} E_t \left[\int_t^\infty \frac{c_s^{1-\frac{1}{\epsilon}}}{1-R} e^{-\rho(s-t)} ds \right]$$

De esta manera,

$$J(a_t) = \max_{c_s | s \in [t, \infty)} \left\{ \int_t^\infty \frac{c_s^{1-\frac{1}{\epsilon}}}{1-R} e^{-\rho(s-t)} ds - e^{-\rho dt} E_{t+dt} \left[\int_{t+dt}^\infty \frac{c_s^{1-\frac{1}{\epsilon}}}{1-R} e^{-\rho(s-(t+dt))} ds \right] \right\}$$

Consecuentemente,

$$\begin{aligned} J(a_t) &= \max_{c_s | s \in [t, \infty)} \left\{ \frac{c_t^{1-\frac{1}{\epsilon}}}{1-R} e^{-\rho s} dt + o(dt) - e^{-\rho dt} E_{t+dt} \left[\int_{t+dt}^\infty \frac{c_s^{1-\frac{1}{\epsilon}}}{1-R} e^{-\rho(s-(t+dt))} ds \right] \right\} \\ &= \max_{c_s | s \in [t, t+dt)} \left\{ \frac{c_t^{1-\frac{1}{\epsilon}}}{1-R} dt + o(dt) + e^{-\rho dt} E_t \max_{c_s | s \in [t+dt, \infty)} E_{t+dt} \left[\int_{t+dt}^\infty \frac{c_s^{1-\frac{1}{\epsilon}}}{1-R} e^{-\rho(s-(t+dt))} ds \right] \right\} \end{aligned}$$

Por lo tanto,

$$J(a_t) = \max_{c_s | s \in [t, t+dt)} \left\{ \frac{c_t^{1-\frac{1}{\epsilon}}}{1-R} dt + o(dt) + e^{-\rho dt} E_t J(a_t + da_t) \right\}$$

$$V(a_t) = A a_t^{(1-R)}$$

Sea

$$J(a_t) = V(a_t)^{\frac{1-\frac{1}{\epsilon}}{1-R}} = \left[A a_t^{(1-R)} \right]^{\frac{1-\frac{1}{\epsilon}}{1-R}} = A^{\frac{1-\frac{1}{\epsilon}}{1-R}} a_t^{1-\frac{1}{\epsilon}}$$

Se define ahora

$$V(a_t)^{\frac{1-\frac{1}{\epsilon}}{1-R}} = \max_{c_s | s \in [t, t+dt)} \left\{ \frac{c_t^{1-\frac{1}{\epsilon}}}{1-R} dt + o(dt) + e^{-\rho dt} E_t \left[\left[V(a_t + da_t) \right]^{\frac{1-\frac{1}{\epsilon}}{1-R}} \right] \right\}$$

Si se denota

$$U(t) = (1-R)V(a_t)^{\frac{1-\frac{1}{\epsilon}}{1-R}}$$

Se tiene que

$$U(t) = \max_{c_s | s \in [t, t+dt)} \left\{ c_t^{1-\frac{1}{\epsilon}} dt + o(dt) + e^{-\rho dt} E_t [U(t+dt)] \right\}$$

Ahora bien,

$$\begin{aligned} V(a_t + da_t) &= V(a_t) + dV(a_t) \\ &= V(a_t) + \left[V'(a_t) a_t (g(\theta_t) - B) + \frac{1}{2} V''(a_t) \sigma_w^2 \right] dt + V'(a_t) \sigma_w dW_t \end{aligned}$$

En consecuencia

$$\left[V(a_t + da_t) \right]^{\frac{1-\frac{1}{\varepsilon}}{1-R}} = \left[V(a_t) + \left[V'(a_t) a_t (g(\theta_t) - B) + \frac{1}{2} V''(a_t) \sigma_w^2 \right] dt + V'(a_t) \sigma_w dW_t \right]^{\frac{1-\frac{1}{\varepsilon}}{1-R}} \quad \text{Ahor}$$

a se sustituye

$$V(a_t) = A a_t^{(1-R)}$$

así

$$\begin{aligned} \left[V(a_t + da_t) \right]^{\frac{1-\frac{1}{\varepsilon}}{1-R}} &= \\ &\left\{ \left[A a_t^{1-R} + \left[(1-R) A a_t^{1-R} (g(\theta_t) - B) - \frac{1}{2} (1-R) R A a_t^{1-R} \theta_t^2 \sigma^2 \right] \right] dt + A a_t^{1-R} \theta_t \sigma dW_t \right\}^{\frac{1-\frac{1}{\varepsilon}}{1-R}} \end{aligned}$$

ó

$$\left[V(a_t + da_t) \right]^{\frac{1-\frac{1}{\varepsilon}}{1-R}} = A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{1-\frac{1}{\varepsilon}} \left[\left[1 + (1-R)(g(\theta_t) - B) - \frac{1}{2} (1-R) R \theta_t^2 \sigma^2 \right] dt + \theta_t \sigma dW_t \right]^{\frac{1-\frac{1}{\varepsilon}}{1-R}}$$

Se aplica la aproximación $(1+x)^\alpha = 1 + \alpha x + o(x)$, se sigue que

$$\left[V(a_t + da_t) \right]^{\frac{1-\frac{1}{\varepsilon}}{1-R}} = A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{1-\frac{1}{\varepsilon}} \left[1 + \left(1 - \frac{1}{\varepsilon} \right) \left[(g(\theta_t) - B) - \frac{1}{2} R \theta_t^2 \sigma^2 \right] dt + \left(1 - \frac{1}{\varepsilon} \right) \theta_t \sigma dW_t \right] + o(dt)$$

Así

$$E_t \left[\left[V(a_t + da_t) \right]^{\frac{1-\frac{1}{\varepsilon}}{1-R}} \right] = A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{1-\frac{1}{\varepsilon}} \left[1 + \left(1 - \frac{1}{\varepsilon} \right) \left[(g(\theta_t) - B) - \frac{1}{2} R \theta_t^2 \sigma^2 \right] dt \right] + o(dt)$$

ó

$$\mathbf{E}_t \left[\left[V(a_t + da_t) \right]^{\frac{1-\frac{1}{\varepsilon}}{1-R}} \right] = \left[\mathbf{E}_t \left[V(a_t + da_t) \right] \right]^{\frac{1-\frac{1}{\varepsilon}}{1-R}}$$

Por lo tanto,

$$V(a_t)^{\frac{1-\frac{1}{\varepsilon}}{1-R}} = \max_{c_t | [t, t+dt]} \left\{ \frac{c_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}}}{1-R} dt + o(dt) + e^{-\rho dt} \mathbf{E}_t \left[V(a_t + da_t)^{\frac{1-\frac{1}{\varepsilon}}{1-R}} \right] \right\}$$

lo cual implica que

$$A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}} = \max_{c_t | [t, t+dt]} \left\{ \frac{c_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}}}{1-R} dt + o(dt) + e^{-\rho dt} A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}} \left[1 + \left(1 - \frac{1}{\varepsilon} \right) \left[(g(\theta_t) - B) - \frac{1}{2} R \theta_t^2 \sigma^2 \right] dt \right] \right\}$$

Si se utiliza ahora la aproximación

$$e^{-\rho dt} \approx 1 - \rho dt + o(dt)$$

Se tiene que

$$A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}} = \max_{c_t | [t, t+dt]} \left\{ \frac{c_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}}}{1-R} dt + o(dt) + A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}} \left[1 + \left(1 - \frac{1}{\varepsilon} \right) \left[(g(\theta_t) - B) - \frac{1}{2} R \theta_t^2 \sigma^2 \right] dt \right] - \rho A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}} dt \right\}$$

ó

$$0 = \max_{c_t | [t, t+dt]} \left\{ \frac{c_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}}}{1-R} dt + o(dt) + A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}} \left(1 - \frac{1}{\varepsilon} \right) \left[(g(\theta_t) - B) - \frac{1}{2} R \theta_t^2 \sigma^2 \right] dt - \rho A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}} dt \right\}$$

Si se supone que $c_t = Ba_t$, entonces

$$0 = \max_{c_t} \left\{ \frac{c_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}}}{1-R} + A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{\frac{1-\frac{1}{\varepsilon}}{\varepsilon}} \left\{ -\rho + \left(1 - \frac{1}{\varepsilon} \right) \left[(g(\theta_t) - B) - \frac{1}{2} R \theta_t^2 \sigma^2 \right] \right\} \right\}$$

Si c_t es óptimo, se obtiene que

$$0 = \frac{c_t^{1-\frac{1}{\varepsilon}}}{1-R} + A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{1-\frac{1}{\varepsilon}} \left\{ -\rho + \left(1 - \frac{1}{\varepsilon}\right) \left[\left((r + \theta_t(\mu - r)) - B \right) - \frac{1}{2} R \theta_t^2 \sigma^2 \right] \right\}$$

equivalentemente

$$0 = \frac{B^{1-\frac{1}{\varepsilon}}}{1-R} a_t^{1-\frac{1}{\varepsilon}} + A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} a_t^{1-\frac{1}{\varepsilon}} \left[-\rho + \left(1 - \frac{1}{\varepsilon}\right) \left[\left((r + \theta_t(\mu - r)) - B \right) - \frac{1}{2} R \theta_t^2 \sigma^2 \right] \right]$$

ó

$$0 = \frac{B^{1-\frac{1}{\varepsilon}}}{1-R} + A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} \left[-\rho + \left(1 - \frac{1}{\varepsilon}\right) \left[\left((r + \theta_t(\mu - r)) - B \right) - \frac{1}{2} R \theta_t^2 \sigma^2 \right] \right] \equiv H$$

Si se calcula ahora

$$\frac{\partial H}{\partial \theta_t} = 0,$$

entonces

$$\mu - r - R \theta_t^2 \sigma^2 = 0.$$

Por lo tanto

$$\theta = \left(\frac{\mu - r}{\sigma} \right) \frac{1}{\sigma R}.$$

Al sustituir en H el valor de θ , se sigue que

$$0 = \frac{B^{1-\frac{1}{\varepsilon}}}{1-R} + A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} \left[-\rho + \left(1 - \frac{1}{\varepsilon}\right) \left(r + \frac{\lambda^2}{2R} - B \right) \right] \equiv K$$

Si se deriva K con respecto de B , se tiene que

$$\frac{\partial K}{\partial B} = \frac{1-\frac{1}{\varepsilon}}{1-R} B^{-\frac{1}{\varepsilon}} - A^{\frac{1-\frac{1}{\varepsilon}}{1-R}} \left(1 - \frac{1}{\varepsilon}\right) = 0$$

así

$$B = (1-R)^{-\varepsilon} A^{\frac{1-\varepsilon}{1-R}}$$

Al sustituir en B en K , se tiene

$$0 = (1-R)^{-\varepsilon} A^{\frac{1-1/\varepsilon}{1-R}(1-\varepsilon)} + A^{\frac{1-1/\varepsilon}{1-R}} \left[-\rho + \left(1 - \frac{1}{\varepsilon}\right) \left(r + \frac{\lambda^2}{2R}\right) \right] - A^{\frac{1-1/\varepsilon}{1-R}(1-\varepsilon)} \left(1 - \frac{1}{\varepsilon}\right) (1-R)^{-\varepsilon}$$

Por lo tanto,

$$0 = (1-R)^{-\varepsilon} A^{\frac{1-1/\varepsilon}{1-R}(-\varepsilon)} - \rho + \left(1 - \frac{1}{\varepsilon}\right) \left(r + \frac{\lambda^2}{2R}\right) - A^{\frac{1-1/\varepsilon}{1-R}(-\varepsilon)} \left(1 - \frac{1}{\varepsilon}\right) (1-R)^{-\varepsilon}$$

ó

$$0 = (1-R)^{-\varepsilon} A^{\frac{1-1/\varepsilon}{1-R}(-\varepsilon)} \left[1 - \left(1 - \frac{1}{\varepsilon}\right) \right] - \rho + \left(1 - \frac{1}{\varepsilon}\right) \left(r + \frac{\lambda^2}{2R}\right)$$

así

$$\frac{(1-R)^{-\varepsilon}}{\varepsilon} A^{\frac{1-1/\varepsilon}{1-R}(-\varepsilon)} = -\rho + \left(1 - \frac{1}{\varepsilon}\right) \left(r + \frac{\lambda^2}{2R}\right)$$

ó

$$A^{\frac{1-\varepsilon}{1-R}} = \varepsilon \left[\rho - \left(1 - \frac{1}{\varepsilon}\right) \left(r + \frac{\lambda^2}{2R}\right) \right] (1-R)^{\varepsilon}$$

En consecuencia,

$$A = \left[\varepsilon \rho + (1-\varepsilon) \left(r + \frac{\lambda^2}{2R}\right) \right]^{\frac{1-R}{1-\varepsilon}} (1-R)^{\varepsilon \frac{1-R}{1-\varepsilon}}$$

ó

$$A = \left[\varepsilon \rho + (1-\varepsilon) \left(r + \frac{\lambda^2}{2R}\right) \right]^{\frac{1-R}{1-\varepsilon}} (1-R)^{\frac{1-R}{1-1/\varepsilon}}.$$

Si se sustituye A en la ecuación de B

$$B = (1-R)^{-\varepsilon} A^{\frac{1-1/\varepsilon}{1-R}(-\varepsilon)} = (1-R)^{-\varepsilon} A^{\frac{1-\varepsilon}{1-R}}$$

se sigue, finalmente, que

$$B = (1-R)^{-\varepsilon} (1-R)^{\varepsilon} \left(\varepsilon \rho + (1-\varepsilon) \left(r + \frac{\lambda^2}{2R} \right) \right),$$

con lo que se concluye la solución del problema de optimización del consumidor racional

14.6 Solución del problema del consumidor con Hamilton-Jacobi-Bellman

A continuación, para destacar las diferencias en la solución, se resolverá el problema de la sección anterior con el enfoque clásico de Hamilton-Jacobi-Bellman (HJB). Considere el problema

$$\begin{aligned} \text{Maximizar} \quad & E \left[\int_0^{\infty} \frac{c_s^{1-\frac{1}{\varepsilon}}}{1-R} e^{-\rho(s)} ds \middle| F_t \right] \\ \text{sujeto a:} \quad & da_t = a_t (1-\theta_t) dR_b + a_t \theta_t dR_s - c_t dt \\ & dR_b = r dt \quad dR_s = \mu dt + \sigma dW_t \end{aligned}$$

y sean

$$\begin{aligned} da_t &= a_t \left(r + \theta_t (\mu - r) - \frac{c_t}{a_t} \right) dt + a_t \theta_t \sigma dW_t \\ \frac{da_t}{a_t} &= \mu_a dt + \sigma_a dW_t \\ \mu_a &= r + \theta_t (\mu - r) - \frac{c_t}{a_t} \quad \sigma_a = \theta_t \sigma \end{aligned}$$

Se define

$$J(a_t, t) = \max_{c_s | s \geq t} E_t \left[\int_t^{\infty} \frac{c_s^{1-\frac{1}{\varepsilon}}}{1-R} e^{-\rho s} ds \right]$$

Ahora se aplica el teorema del valor medio del cálculo integral y recursividad

$$J(a_t, t) = \max_{c_s | s \in [t, t+dt]} \left\{ \frac{c_s^{1-\frac{1}{\varepsilon}}}{1-R} e^{-\rho s} dt + o(dt) + J(a_t + da_t, t + dt) \right\}$$

De la definición de diferencial, se sigue que

$$J(a_t, t) = \max_{c_s | s \in [t, t+dt]} \left\{ \frac{c_s^{1-\frac{1}{\varepsilon}}}{1-R} e^{-\rho s} dt + o(dt) + J(a_t, t) + dJ(a_t, t) \right\}$$

Al aplicar el lema de Ito, se obtiene

$$0 = \max_{c_s | s \in [t, t+dt]} \left\{ \frac{c_s^{1-\frac{1}{\varepsilon}}}{1-R} e^{-\rho t} dt + o(dt) + \left[\frac{\partial J(a_t, t)}{\partial t} + \frac{\partial J(a_t, t)}{\partial a_t} a_t \mu_a + \frac{1}{2} \frac{\partial^2 J(a_t, t)}{\partial a_t^2} a_t^2 \sigma_a^2 \right] dt + \frac{\partial J(a_t, t)}{\partial a_t} a_t \sigma_a dW_t \right\}$$

Ahora se toma el valor esperado

$$0 = \max_{c_s | s \in [t, t+dt]} \left\{ \frac{c_s^{1-\frac{1}{\varepsilon}}}{1-R} e^{-\rho t} dt + o(dt) + \left[\frac{\partial J(a_t, t)}{\partial t} + \frac{\partial J(a_t, t)}{\partial a_t} a_t \mu_a + \frac{1}{2} \frac{\partial^2 J(a_t, t)}{\partial a_t^2} a_t^2 \sigma_a^2 \right] dt \right\}$$

Se divide entre la cantidad infinitesimal y se toma el límite cuando dicho infinitesimal tiende a cero, así

$$0 = \max_{c_s | s \in [t, t+dt]} \left\{ \frac{c_s^{1-\frac{1}{\varepsilon}}}{1-R} e^{-\rho t} + \frac{\partial J(a_t, t)}{\partial t} + \frac{\partial J(a_t, t)}{\partial a_t} a_t \mu_a + \frac{1}{2} \frac{\partial^2 J(a_t, t)}{\partial a_t^2} a_t^2 \sigma_a^2 \right\}$$

Si c_t es óptimo, entonces

$$0 = \frac{c_t^{1-\frac{1}{\varepsilon}}}{1-R} e^{-\rho t} + \frac{\partial J(a_t, t)}{\partial t} + \frac{\partial J(a_t, t)}{\partial a_t} a_t \mu_a + \frac{1}{2} \frac{\partial^2 J(a_t, t)}{\partial a_t^2} a_t^2 \sigma_a^2$$

Esta es la ecuación de H-J-B, ahora el problema es encontrar la función de valor de la forma:

$$J(a_t, t) = V(a_t) e^{-\rho t}$$

Por lo tanto,

$$\frac{\partial J(a_t, t)}{\partial a_t} = V'(a_t) e^{-\rho t}$$

$$\frac{\partial^2 J(a_t, t)}{\partial a_t^2} = V''(a_t) e^{-\rho t}$$

$$\frac{\partial J(a_t, t)}{\partial t} = -\rho V(a_t) e^{-\rho t}$$

Al sustituir en la ecuación de H-J-B y dividir entre el factor de descuento, se tiene la siguiente condición de primer orden:

$$0 = \frac{c_t^{1-\frac{1}{\varepsilon}}}{1-R} - \rho V(a_t) + V'(a_t) a_t \mu_a + \frac{1}{2} V''(a_t) \sigma_a^2 a_t^2$$

Ahora se considera el candidato de solución

$$V(a_t) = \beta u(a_t) = \beta \frac{a_t^{1-\frac{1}{\varepsilon}}}{1-R}$$

donde

$$V'(a_t) = \frac{\beta(1-\frac{1}{\varepsilon})}{1-R} a_t^{-\frac{1}{\varepsilon}} \quad \text{y} \quad V''(a_t) = -\frac{1}{\varepsilon} \frac{\beta(1-\frac{1}{\varepsilon})}{1-R} a_t^{-(1+\frac{1}{\varepsilon})}$$

Al sustituir en las condiciones de primer orden se tiene

$$0 = \frac{c_t^{1-\frac{1}{\varepsilon}}}{1-R} - \rho \beta \frac{a_t^{1-\frac{1}{\varepsilon}}}{1-R} + \beta \frac{\beta(1-\frac{1}{\varepsilon})}{1-R} a_t^{-\frac{1}{\varepsilon}} a_t \mu_a + \frac{1}{2} \left(-\frac{1}{\varepsilon}\right) \frac{\beta(1-\frac{1}{\varepsilon})}{1-R} a_t^{-(1+\frac{1}{\varepsilon})} \sigma_a^2 a_t^2$$

$$0 = \frac{c_t^{1-\frac{1}{\varepsilon}}}{1-R} - \rho \beta \frac{a_t^{1-\frac{1}{\varepsilon}}}{1-R} + \beta \frac{\beta(1-\frac{1}{\varepsilon})}{1-R} a_t^{1-\frac{1}{\varepsilon}} \mu_a + \frac{1}{2} \left(-\frac{1}{\varepsilon}\right) \frac{\beta(1-\frac{1}{\varepsilon})}{1-R} a_t^{(1-\frac{1}{\varepsilon})} \sigma_a^2$$

Al multiplicar esta última expresión por el término que contiene a a_t , se tiene

$$0 = \frac{c_t^{1-\frac{1}{\varepsilon}}}{a_t^{1-\frac{1}{\varepsilon}}} - \rho \beta + \beta \left(1-\frac{1}{\varepsilon}\right) \mu_a + \frac{1}{2} \left(-\frac{1}{\varepsilon}\right) \beta \left(1-\frac{1}{\varepsilon}\right) \sigma_a^2 = H$$

Ahora con H y sus derivadas se forma el siguiente sistema de ecuaciones

$$\begin{cases} 0 = \frac{c_t^{1-\frac{1}{\varepsilon}}}{a_t^{1-\frac{1}{\varepsilon}}} - \rho\beta + \beta\left(1-\frac{1}{\varepsilon}\right)\mu_a + \frac{1}{2}\left(-\frac{1}{\varepsilon}\right)\beta\left(1-\frac{1}{\varepsilon}\right)\sigma_a^2 \\ \frac{\partial H}{\partial c_t} = 0 = \frac{1-\frac{1}{\varepsilon}}{a_t^{1-\frac{1}{\varepsilon}}} c_t^{-\frac{1}{\varepsilon}} - \beta \frac{\left(1-\frac{1}{\varepsilon}\right)}{a_t} \Rightarrow c_t^{-\frac{1}{\varepsilon}} = \beta a_t^{-\frac{1}{\varepsilon}} \Rightarrow c_t = \beta^{-\varepsilon} a_t \\ \frac{\partial H}{\partial \theta_t} = 0 = \beta\left(1-\frac{1}{\varepsilon}\right)(\mu-r) - \theta_t \sigma^2 \beta \frac{\left(1-\frac{1}{\varepsilon}\right)}{\varepsilon} \Rightarrow \theta_t = \frac{\mu-r}{\sigma^2} \varepsilon \end{cases}$$

A continuación se sustituyen las derivadas en H

$$\begin{aligned} 0 &= \frac{\beta^{-\varepsilon\left(1-\frac{1}{\varepsilon}\right)} a_t^{1-\frac{1}{\varepsilon}}}{a_t^{1-\frac{1}{\varepsilon}}} - \rho\beta + \beta\left(1-\frac{1}{\varepsilon}\right) \left(r + \frac{\varepsilon(\mu-r)^2}{\sigma^2} - \frac{\beta^{-\varepsilon} a_t}{a_t} \right) - \frac{1}{2} \beta \frac{1-\frac{1}{\varepsilon}}{\varepsilon} \frac{\varepsilon^2 (\mu-r)^2}{\sigma^4} \sigma^2 \\ 0 &= \beta^{(1-\varepsilon)} - \rho\beta + \beta\left(1-\frac{1}{\varepsilon}\right) \left(r + \frac{\varepsilon(\mu-r)^2}{\sigma^2} - \beta^{-\varepsilon} \right) - \frac{1}{2} \beta\left(1-\frac{1}{\varepsilon}\right) \frac{\varepsilon(\mu-r)^2}{\sigma^2} \end{aligned}$$

así

$$0 = \beta^{(-\varepsilon)} - \rho + \left(1-\frac{1}{\varepsilon}\right) \left(r + \frac{\varepsilon(\mu-r)^2}{\sigma^2} - \beta^{-\varepsilon} \right) - \frac{1}{2} \left(1-\frac{1}{\varepsilon}\right) \frac{\varepsilon(\mu-r)^2}{\sigma^2}$$

De esta manera,

$$\begin{aligned} 0 &= \beta^{(-\varepsilon)} \left(1 - \left(1-\frac{1}{\varepsilon}\right)\right) - \rho + \left(1-\frac{1}{\varepsilon}\right) \left(r + \frac{1}{2} \frac{\varepsilon(\mu-r)^2}{\sigma^2} \right) \\ 0 &= \beta^{-\varepsilon} \left(\frac{1}{\varepsilon}\right) - \rho + \left(1-\frac{1}{\varepsilon}\right) \left(r + \frac{1}{2} \frac{\varepsilon(\mu-r)^2}{\sigma^2} \right) \end{aligned}$$

lo cual implica que

$$\beta^{-\varepsilon} \left(\frac{1}{\varepsilon}\right) = \rho - \left(1-\frac{1}{\varepsilon}\right) \left(r + \frac{1}{2} \frac{\varepsilon(\mu-r)^2}{\sigma^2} \right)$$

ó

$$\beta^{-\varepsilon} = \varepsilon\rho + (1-\varepsilon) \left(r + \frac{1}{2} \frac{\varepsilon(\mu-r)^2}{\sigma^2} \right)$$

La trayectoria del consumo está dada por:

$$\begin{aligned} c_t &= \left[\varepsilon\rho + (1-\varepsilon) \left(r + \frac{1}{2} \frac{\varepsilon(\mu-r)^2}{\sigma^2} \right) \right] a_t \\ c_t &= \left[\varepsilon\rho + (1-\varepsilon) \left(r + \frac{1}{2} \varepsilon\lambda^2 \right) \right] a_t \end{aligned}$$

con

$$\lambda = \frac{\mu - r}{\sigma}$$

Lo cual representa el premio al riesgo.

14.7 Análisis comparativo entre UDRE y HJB

En esta sección se lleva a cabo un análisis comparativo entre las soluciones de los problemas de optimización con utilidad diferencial recursiva estocástica y Hamilton-Jacobi-Bellman. En el primer caso las decisiones de consumo y portafolio son:

$$\theta = \frac{\lambda}{\sigma R},$$

$$c_t = (1 - R)^{-\varepsilon} (1 - R)^\varepsilon \left(\varepsilon \rho + (1 - \varepsilon) \left(r + \frac{\lambda^2}{2R} \right) \right) a_t.$$

Mientras que en el segundo

$$\theta = \frac{\lambda \varepsilon}{\sigma},$$

$$c_t = \left[\varepsilon \rho + (1 - \varepsilon) \left(r + \frac{1}{2} \varepsilon \lambda^2 \right) \right] a_t.$$

Como puede observarse el primer caso, el parámetro de aversión al riesgo está en el denominador de la proporción de la riqueza que el agente asigna a la tenencia del activo riesgoso. Entre más cóncava sea la función, mayor será R , es decir, el individuo tiene mayor aversión al riesgo, lo que hace que la proporción de la riqueza que se dedica a la tenencia del activo riesgoso disminuya; situación que no se observa en el caso tradicional de HJB. Asimismo, en el primer caso la decisión de consumo se ve afectada por la aversión al riesgo; mientras que en el segundo sólo participa la elasticidad de sustitución intertemporal del consumo.

14.8 Conclusiones

El concepto de utilidad diferencial recursiva estocástica ha permitido separar en las decisiones de consumo y portafolio la elasticidad de sustitución intertemporal de consumo y la aversión relativa al riesgo.

Particularmente, en las decisiones de consumos e inversión en un portafolio que contiene un activo riesgoso, UDRE sí permite que en la proporción de la riqueza real que se destina al activo riesgo aversión al riesgo tenga un efecto negativo.

Por último, en la agenda de investigación futura se incorporarán saltos de Poisson en el enfoque de UDRE.

14.9 Bibliografía

- Breeden, D. (1979). An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities. *Journal of Financial Economics*, Vol. 7, pp. 265-296.
- Breeden, D., M. Gibbons, and R. Litzenberger (1989). Empirical Tests of the Consumption Oriented CAPM. *Journal of Finance*, Vol. 44, pp. 231-262.
- Bufman, G., and L. Leiderman (1990). Consumption and Asset Returns Under Non-Expected Utility. *Economics Letters*, Vol. 34, pp. 231-235.
- Campbell, J. (1990). Intertemporal Asset Pricing without Consumption. Working paper, Department of Economics, Princeton University.
- Chatterjee, S., P. Giuliano, and S. J. Turnovsky (2003). Capital Income Taxes and Growth in a Stochastic Economy: A Numerical Analysis of the Role of Risk Aversion and Intertemporal Substitution, Working paper, University of Washington, Seattle.
- Constantinides, G. (1990). Habit Formation: A Resolution of the Equity Premium Puzzle. *Journal of Political Economy*, Vol. 98, pp. 519-543.
- Constantinides, G. M., Donaldson, J. B., Mehra, R. (2002). Junior Can't Borrow: a New Perspective on the Equity Premium Puzzle. *Quarterly Journal of Economics*, Vol. 117, pp. 269-296.
- Cox, J., J. Ingersoll, and S. Ross (1985a). A Theory of the Term Structure of Interest Rates. *Econometrica*, Vol. 53, pp. 385-408.
- Detemple, J., and F. Zapatero (1989). Optimal Consumption Portfolio Policies with Habit Formation. Working paper, School of Business, Columbia University.
- Duffie, D., and L. Epstein (1992). Stochastic Differential Utility. *Econometrica*, Vol. 60, pp. 353-394.
- Duffie, D., and P. L. Lions (1990). PDE Solutions of Stochastic Differential Utility. Research Paper 1095, Graduate School of Business, Stanford University, forthcoming in *Journal of Mathematical Economics*.
- Epstein, L. (1988). Risk Aversion and Asset Prices. *Journal of Monetary Economics*, Vol. 22, pp. 179-192.
- Epstein, L. (1992). Behavior Under Risk: Recent Developments in Theory and Applications. in J. J. Laffont (ed.), *Advances in Economic Theory*, Cambridge University Press, Cambridge.
- Epstein, L., and S. Zin (1990). Consumption, Labor Supply and Portfolio Choice with Time and State Non-Separable Utility. Working paper, Department of Economics, University of Toronto.
- Epstein, L., and S. Zin (1991a). Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis. *Journal of Political Economy*, Vol. 99, pp. 263-286.
- Epstein, L., and S. Zin (1991b). The Independence Axiom and Asset Return. Working Paper 109. National Bureau of Economic Research.

- Giovannini, A., and P. Weil (1989). Risk Aversion and Intertemporal Substitution in the Capital Asset Pricing Model. Working Paper 2824, National Bureau of Economic Research.
- Giuliano, P., and S. J. Turnovsky (2002). Intertemporal Substitution, Risk Aversion, and Economic Performance in a Stochastically Growing Open Economy, Working paper, University of Washington, Seattle.
- Hansen, L., and K. Singleton (1983). Stochastic Consumption, Risk Aversion and the Temporal Behavior of Asset Returns. *Journal of Political Economy*, Vol. 91, pp. 249-265.
- Heaton, J. (1991). An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications. Working Paper 3245.91, Sloan School of Management, MIT.
- Hindy, A., and C.F. Huang (1990). Optimal Consumption and Portfolio Rules with Local Substitution. Working Paper No. 89, revised, Alfred P. Sloan School of Management, MIT. 435
- Huang, C.-F. (1987). An Intertemporal General Equilibrium Asset Pricing Model. *Econometrica*, Vol. 55, pp. 117-142.
- Kandel, S., and R. Stambaugh (1990). Asset Returns and Intertemporal Preferences,” Working Paper 3633, National Bureau of Economic Research.
- Lintner, J. (1965). The Valuation of Risky Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budgets, *Review of Economics and Statistics*, Vol. 47, pp. 13-37.
- Lucas, R. (1978). Asset Prices in an Exchange Economy. *Econometrica*, Vol. 46, pp. 1429-1445.
- Mankiw, G., and M. Shapiro (1986). Risk and Return: Consumption Beta versus Market Beta. *Review of Economics and Statistics*, Vol. 68, pp. 452-459.
- Mehra, R., and E. Prescott (1985). The Equity Premium: A Puzzle. *Journal of Monetary Economics*, Vol. 15, pp. 145-161.
- Roll, R. (1977). A Critique of the Asset Pricing Theory’s Tests: Part I: On Past and Potential Testability of the Theory, *Journal of Financial Economics*, Vol.4, pp. 129-176.
- Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, Vol. 19, pp. 425-442.
- Singleton, K. (1990), Specification and Estimation of Intertemporal Asset Pricing Models,” in B. Friedman and F. Hahn (eds.), *Handbook of Monetary Economics*, North Holland, Amsterdam, pp. 583-623.
- Sundaresan, S. M. (1989). Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth. *Review of Financial Studies*, Vol. 2, pp. 73-89.
- Venegas-Martínez, F. (2008). Riesgos Económicos y Financieros: Valuación de Productos Derivados y Decisiones Económicas bajo Incertidumbre, 2da. edición, Cengage. México.
- Weil, P. (1990). Non-Expected Utility in Macroeconomic. *Quarterly Journal of Economics*, Vol. 105, pp. 29-42.

Capítulo 15

Seguimiento y pronóstico de precios de tres acciones que cotizan en la BMV empleando Redes Neuronales Diferenciales

Francisco Ortiz-Arango

ECEE, Universidad Panamericana

Agustín Ignacio Cabrera-Llanos

UPIBI, Instituto Politécnico Nacional

Fernando Cruz-Aranda

ECEE, Universidad Panamericana

Resumen

La utilización de las redes neuronales artificiales como herramienta de descripción, análisis y pronóstico del comportamiento de variables económicas y financieras se ha convertido en los últimos años en un método cada vez más usado y confiable debido a los excelentes resultados obtenidos McNelis (2005) y Huang et al. (2007), en el pronóstico del rendimiento de acciones Vashisth y Chandra (2010), así como el pronóstico del precio de acciones Padhiary y Mishra (2011). A diferencia de los trabajos citados, en los cuales se utilizan redes neuronales artificiales discretas, en este trabajo se utiliza una clase particular de red neuronal artificial, la llamada red neuronal diferencial o dinámica (RND) para llevar a cabo un análisis completo de las series de precios de cierre diarios de tres acciones que cotizan en la Bolsa Mexicana de Valores (BMV): TELMEX, GMODELOC (Grupo Modelo) y CEMEX dentro de un periodo comprendido entre el 3 de enero de 2000 y 7 de enero de 2011. Dentro de este periodo de análisis podemos identificar tres subperiodos, cada uno con objetivos distintos: El primero se utiliza como periodo de aprendizaje o entrenamiento de la red y dura 6 meses; el segundo periodo, llamado de trabajo se emplea para llevar a cabo el seguimiento de los precios de cierre diarios de las tres acciones mencionadas y finalmente el tercer periodo que dura tres semanas se enfoca en llevar a cabo el pronóstico de los precios

de cierre diarios de las acciones bajo estudio. En ambos casos los resultados obtenidos con la red neuronal diferencial fueron sumamente precisos.

15.1 Introducción

La utilización de las redes neuronales artificiales como herramienta de descripción, análisis y pronóstico del comportamiento de variables económicas y financieras se ha convertido en una técnica cada vez más frecuente en los últimos años. Esta situación se debe primordialmente a dos aspectos muy relevantes:

- La gran precisión de sus resultados tanto en el proceso de seguimiento como en el de pronóstico de valores futuros de variables financieras, esto le ha conferido un estatus de gran confiabilidad McNelis (2005) y Huang et al. (2007)
- La relativa facilidad de su uso e implementación gracias al gran incremento en la capacidad de los sistemas de cómputo en los últimos años, tanto en el hardware como en el software de naturaleza cuantitativa.

De este modo podemos encontrar en la literatura reciente trabajos como el de Vashisth y Chandra (2010) en el que llevan a cabo el pronóstico del rendimiento de acciones; también hay trabajos como el de Padhiary y Mishra (2011) en el cual presentan resultados del pronóstico del precio de acciones. En los trabajos citados anteriormente los autores utilizaron redes neuronales artificiales de tipo discreto, a diferencia de estos artículos en este trabajo se utiliza una clase particular de red neuronal artificial, la llamada red neuronal diferencial o dinámica (RND), utilizando esta técnica se pueden citar los trabajos de Cabrera y Ortiz (2012) en el cual realizan el pronóstico del rendimiento del Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores. También están los trabajos de Ortiz et al. (2012a), donde llevan a cabo el modelado del comportamiento del tipo de cambio Peso-Dólar; también Ortiz et al. (2012b), en el que realizan el pronóstico de los índices accionarios de las Bolsas de Valores de España y México.

En este trabajo mediante el uso de una RND llevamos a cabo un análisis completo de las series de precios de cierre diarios de tres títulos accionarios que cotizan en la Bolsa Mexicana de Valores (BMV): TELMEXL (Teléfonos de México, S.A.B. de C.V.), GMODELOC (Grupo Modelo, S.A.B. de C.V.) y CEMEXCPO (Cemex, S.A.B. de C.V.) en un periodo comprendido entre el 3 de enero de 2000 y 7 de enero de 2011. Dentro de este periodo de análisis se establecieron tres subperiodos, cada uno con objetivos distintos:

- El primero se utiliza como periodo de aprendizaje o entrenamiento de la red y dura 6 meses, comprendidos entre el 3 de enero al 30 de junio de 2000.
- El segundo periodo, llamado de trabajo se emplea para llevar a cabo el seguimiento de los precios de cierre diarios de las tres acciones mencionadas entre el 3 de julio de 2000 y el 17 de diciembre de 2010.
- Finalmente el tercer periodo que dura tres semanas se enfoca en llevar a cabo el pronóstico de los precios de cierre diarios de las acciones bajo estudio y comprende del 20 de diciembre de 2010 al 7 de enero de 2011.

Al revisar los resultados obtenidos con la red neuronal diferencial puede apreciarse que los resultados obtenidos fueron sumamente precisos, lo cual alimenta la idea de reforzar el estudio y uso de esta técnica para llevar a cabo el análisis de predicción de valores de variables financieras.

El presente capítulo está conformado del siguiente modo, en la siguiente sección se plantea de manera resumida la descripción matemática de la metodología de las redes neuronales diferenciales; en la sección 15.3 se presentan los resultados obtenidos y finalmente se enlistan las conclusiones del trabajo realizado.

15.2 Metodología

En esta sección se plantea la descripción matemática de la red neuronal diferencial utilizada en este trabajo parte fundamental del mismo. La metodología consta fundamentalmente de 3 etapas: La primera consiste básicamente en "entonar" la RND y llevar a cabo el proceso de entrenamiento de la red, este paso es fundamental pues de este depende la precisión de los cálculos de la red en las etapas de seguimiento y de pronóstico. En este caso utilizamos los datos de las tres series de precios de cierre comprendidos entre el 3 de enero al 30 de junio de 2000.

Cabe hacer mención de que una ventaja adicional de las RND sobre las redes neuronales discretas es que dado que las RND se basan en su operación en criterios de convergencia emanados de la teoría de estabilidad de Lyapunov, esto permite entre otras cosas una mayor precisión y rapidez en la convergencia de las soluciones, lo cual se traduce para fines prácticos en una reducción considerable en el número de datos requeridos para llevar a cabo el proceso de entrenamiento de la red neuronal hasta casi ocho veces menos datos, esto es muy útil y práctico en las aplicaciones pues en ocasiones son pocos los datos disponibles y usar una red de este tipo permite ampliar tanto el periodo de seguimiento como el de pronóstico en el caso requerido.

La segunda etapa se enfoca al proceso de seguimiento de los valores de las tres series, se pretende que los valores generados por la red sean lo más cercanos a los valores de mercado de las acciones consideradas, los cuales están registrados en las series empleadas, el periodo considerado está entre el 3 de julio de 2000 y el 17 de diciembre de 2010.

Para llevar a cabo el proceso de seguimiento así como el de pronóstico de los valores de las series financieras citadas se utilizó el siguiente sistema de ecuaciones diferenciales no lineales:

$$\frac{d}{dt}x_t = \dot{\hat{x}}_t = f(x_t, u_t, t) + \xi_{1,t},$$

con x_0 conocida

$$y_t = Cx_t + \xi_{2,t}$$

$$x_t \in \mathbb{R}^n, y_t \in \mathbb{R}^p \quad t \geq 0, u_t \in \mathbb{R}^m$$

$$C \in \mathbb{R}^{p \times n}, \xi_{1,t} \in \mathbb{R}^{n \times 1} \text{ y } \xi_{2,t} \in \mathbb{R}^{p \times 1}$$

$$Cx_t > \xi_{2,t}$$

donde:

x_t es el vector de estados

y_t es el vector de salida o resultados

C es la matriz de transformación del estado x_t al estado y_t

$\xi_{1,t}$ y $\xi_{2,t}$ son ruido blanco el cual no puede ser medido pero si acotado

sobre la dinámica de los estados

u_t es el vector de alimentación de información externa

Además este sistema está sujeto a las siguientes restricciones:

$$\dot{\hat{x}}_t = A\hat{x}_t + W_{1,t}\sigma(\hat{x}_t) + W_{2,t}\phi(\hat{x}_t)\gamma(u_t)$$

$$\dot{W}_{1,t} = -K_1 P \Delta_t \sigma(\hat{x}_t)^T$$

$$\dot{W}_{2,t} = -K_2 P \Delta_t \gamma(u_t)^T \phi(\hat{x}_t)^T$$

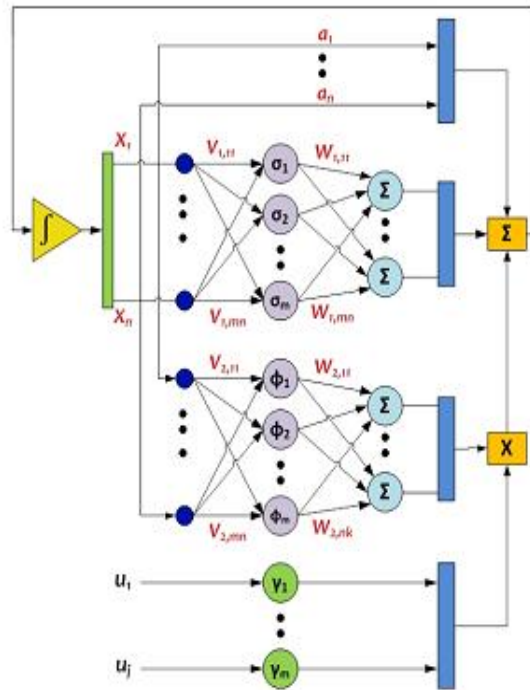
$$\sigma_i(\hat{x}_t) = \frac{a_i}{1 + e^{-b_i \hat{x}_t}} - c_i$$

$$\phi_{jj}(\hat{x}_t) = \frac{\bar{a}_j}{1 + e^{-\bar{b}_j \hat{x}_t}} - \bar{c}_j$$

$$A^T P + PA + PRP + Q = 0$$

Para poder resolver tanto sistema de ecuaciones diferenciales no lineales como el conjunto de restricciones se utilizó la plataforma de Matlab-Simulink. A continuación, en la figura 15.1 se muestra el esquema típico de una red neuronal diferencial, Cabrera y Ortiz (2012).

Figura 15.1. Esquema de una red neuronal diferencial.



En la tercera y última etapa se lleva a cabo el proceso de pronóstico de los valores de cierre diarios de cada una de las tres acciones, a lo largo del periodo del 20 de diciembre de 2010 al 7 de enero de 2011.

15.3 Resultados.

En esta sección se presentan los resultados obtenidos, en primera instancia se muestran las gráficas de los procesos de seguimiento, en las cuales podremos apreciar la gran precisión entre los valores calculados con la red con respecto a los valores reales de las series. Posteriormente mostraremos las gráficas de los errores porcentuales al comparar los datos reales con los valores obtenidos mediante la red para los pronósticos de los precios de cierre diarios de las acciones de TELMEXL, GMODELOC y CEMEXCPO. A través de las cuales podremos verificar su gran precisión como método predictor de valores de variables financieras.

En las figuras 15.2 a 15.10 podemos visualizar la gran precisión del proceso de seguimiento para el caso de las tres series de precios de cierre de las acciones consideradas, iniciamos con las figuras 15.2 a 15.4 que corresponden al proceso de descripción de los precios de las acciones de TELMEXL. Las figuras 15.5 a 15.7 corresponden a la descripción de los precios de las acciones de GMODELOC y en las figuras 15.8 a 15.10 se presenta el seguimiento de los precios de cierre de CEMEXCPO. En todos los casos es evidente la gran proximidad entre los valores reales de las series y los valores generados mediante la RND.

Figura 15.2. Primer acercamiento de los precios de las acciones de TELMEXL y los precios estimados por la RND

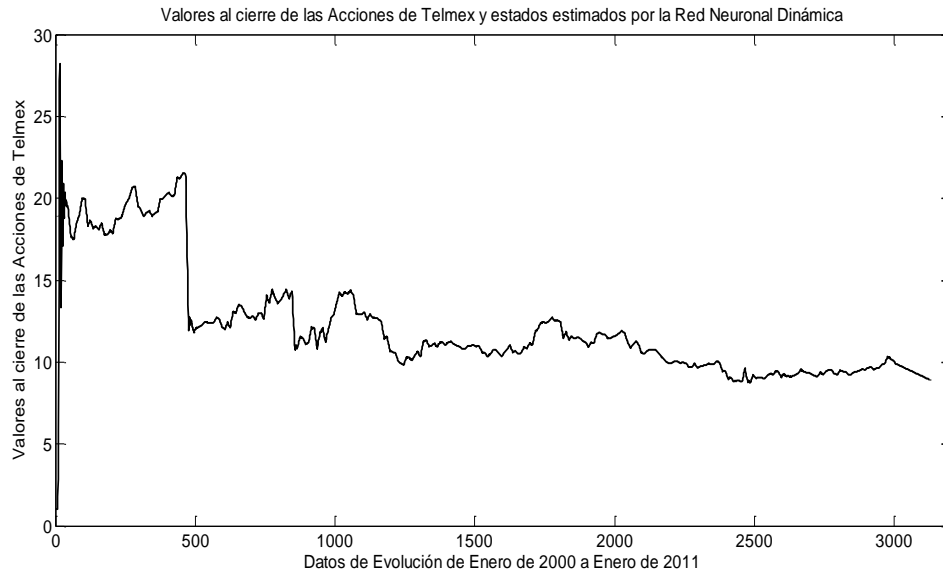


Figura 15.3. Segundo acercamiento de los precios de las acciones de TELMEXL y los precios estimados por la RND

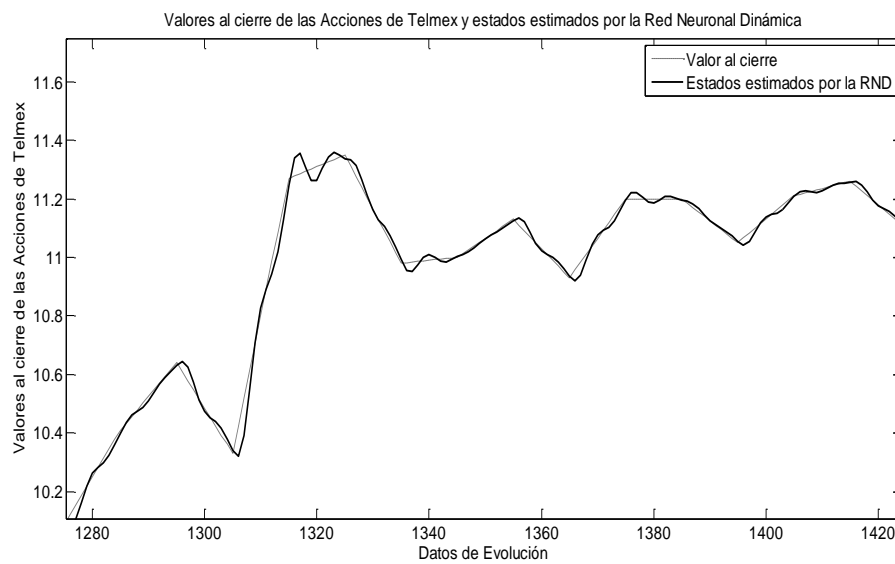


Figura 15.4. Tercer acercamiento de los precios de las acciones de TELMEXL y los precios estimados por la RND

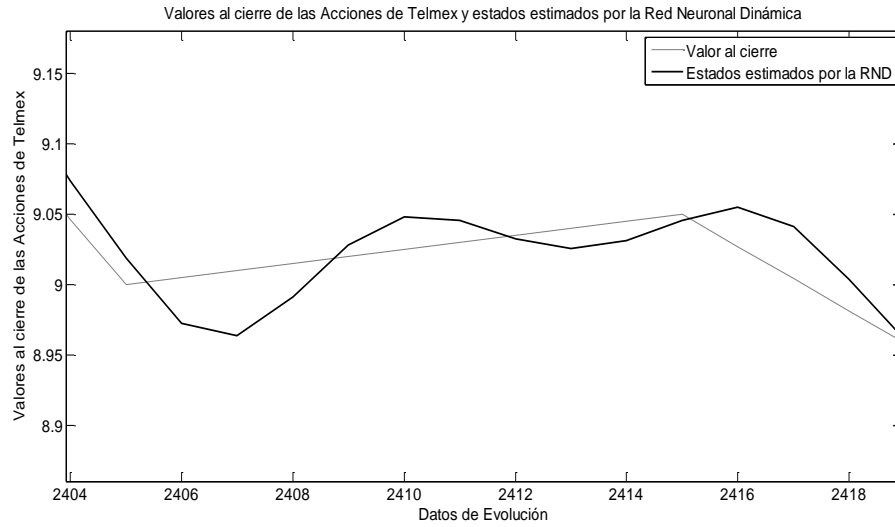


Figura 15.5. Primer acercamiento de los precios de las acciones de GMODELOC y los precios estimados por la RND

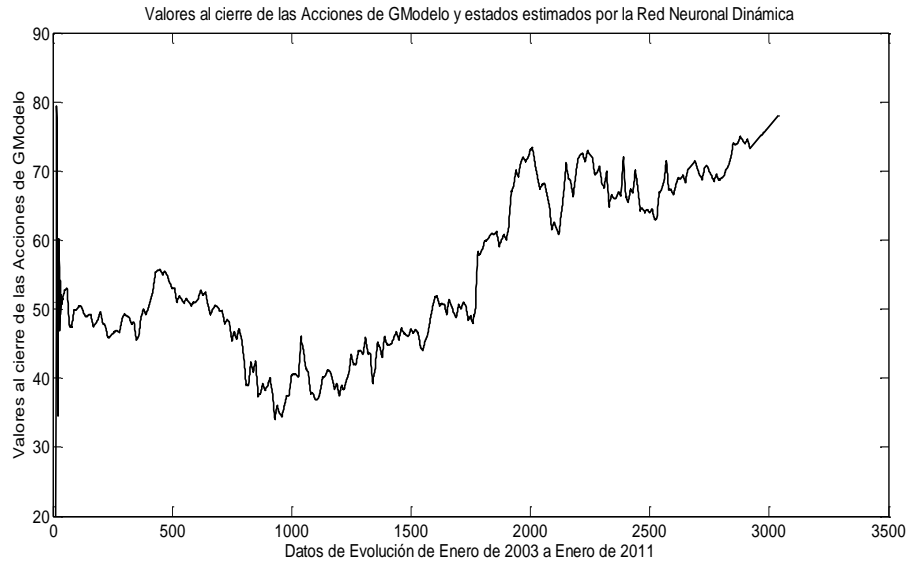


Figura 15.6. Segundo acercamiento de los precios de las acciones de GMODELLOC y los precios estimados por la RND

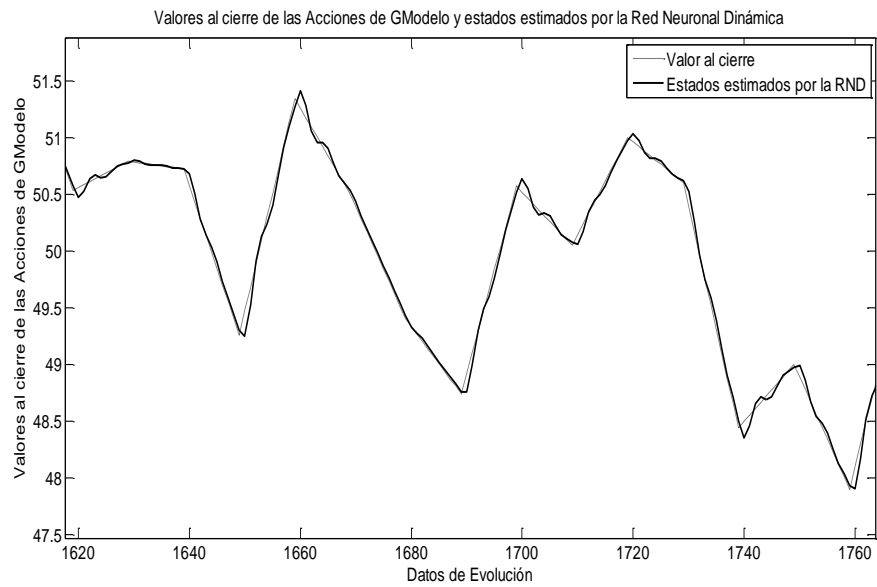


Figura 15.7. Tercer acercamiento de los precios de las acciones de GMODELLOC y los precios estimados por la RND

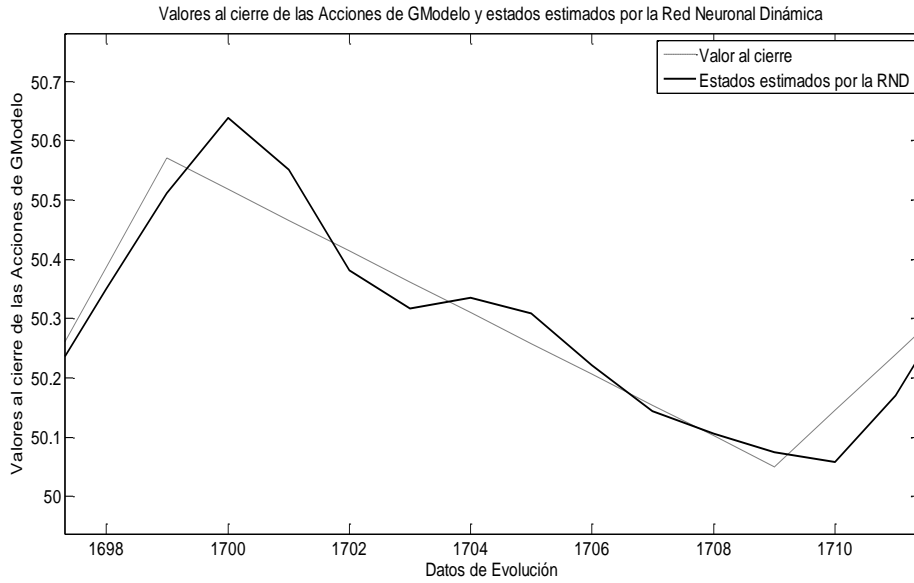


Figura 15.8. Primer acercamiento de los precios de las acciones de CEMEXCPO y los precios estimados por la RND



Figura 15.9. Segundo acercamiento de los precios de las acciones de CEMEXCPO y los precios estimados por la RND

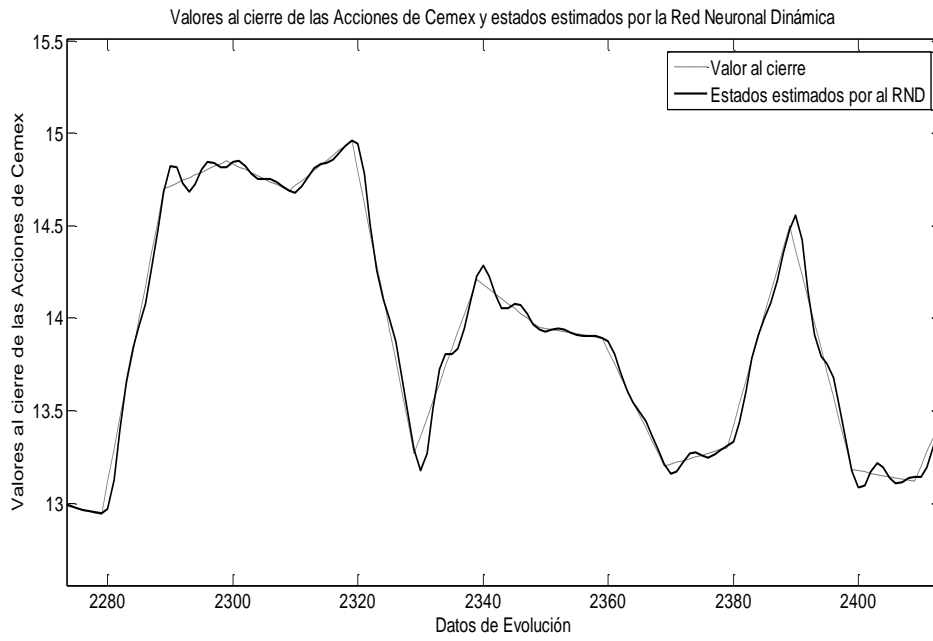
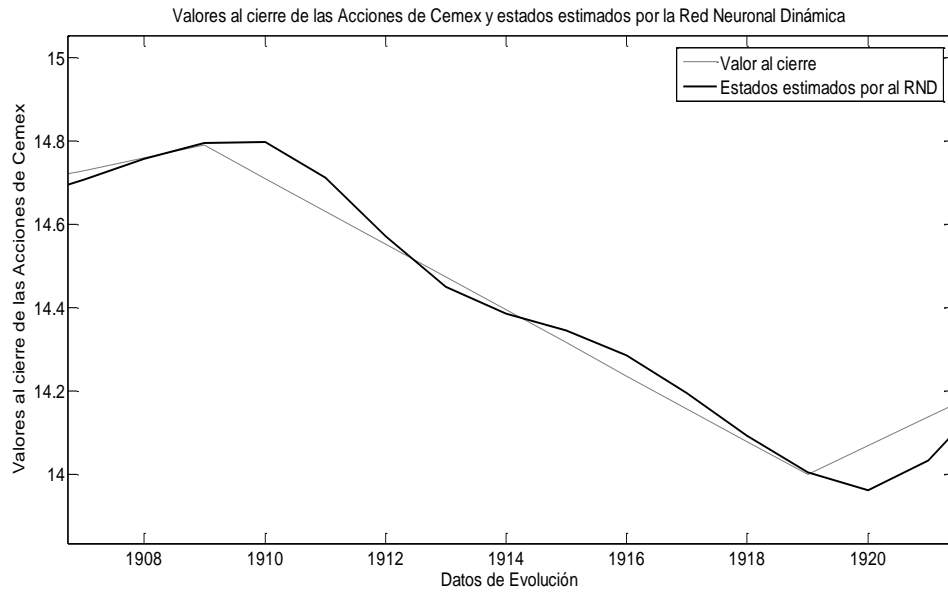
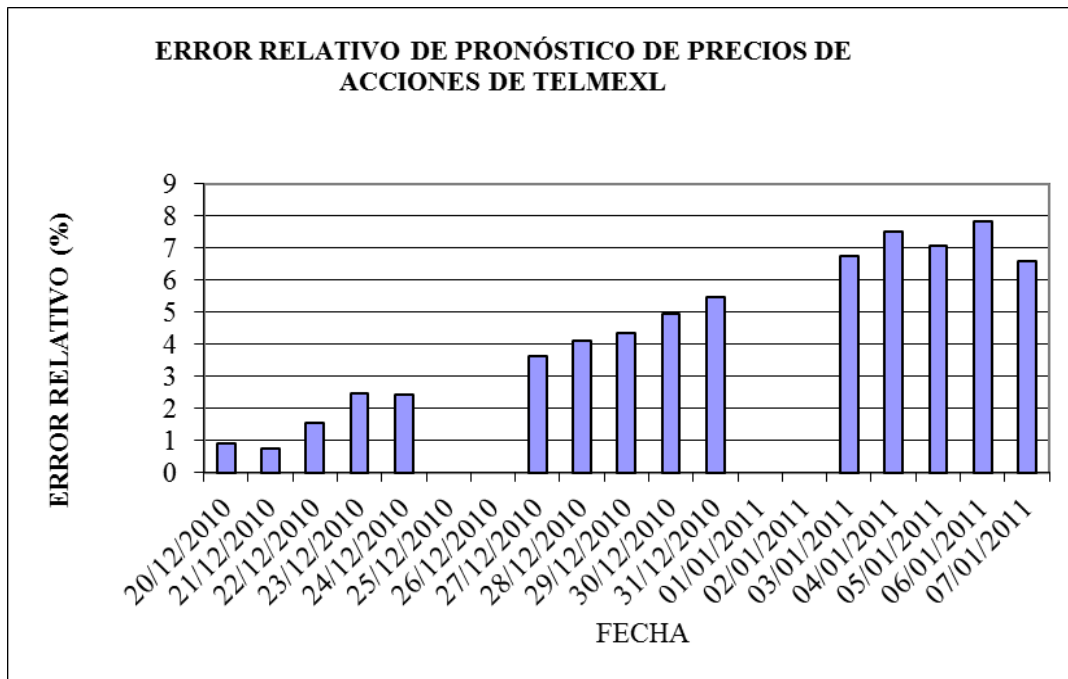


Figura 15.10. Tercer acercamiento de los precios de las acciones de CEMEXCPO y los precios estimados por la RND

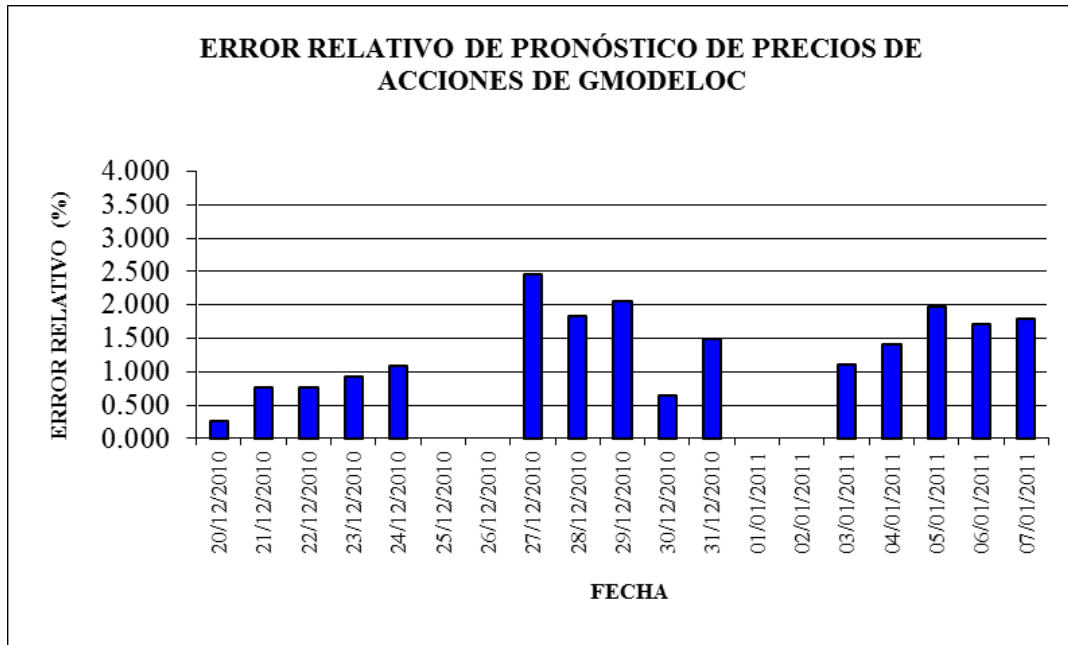


A continuación se presentan las gráficas de los errores relativos entre los precios de cierre reales de las tres acciones comparados contra los precios de cierre pronosticados por la red neuronal diferencial a lo largo de tres semanas, veremos la gran precisión de la red para llevar a cabo este tipo de pronósticos sobre todo durante las primeras dos semanas de los pronósticos.

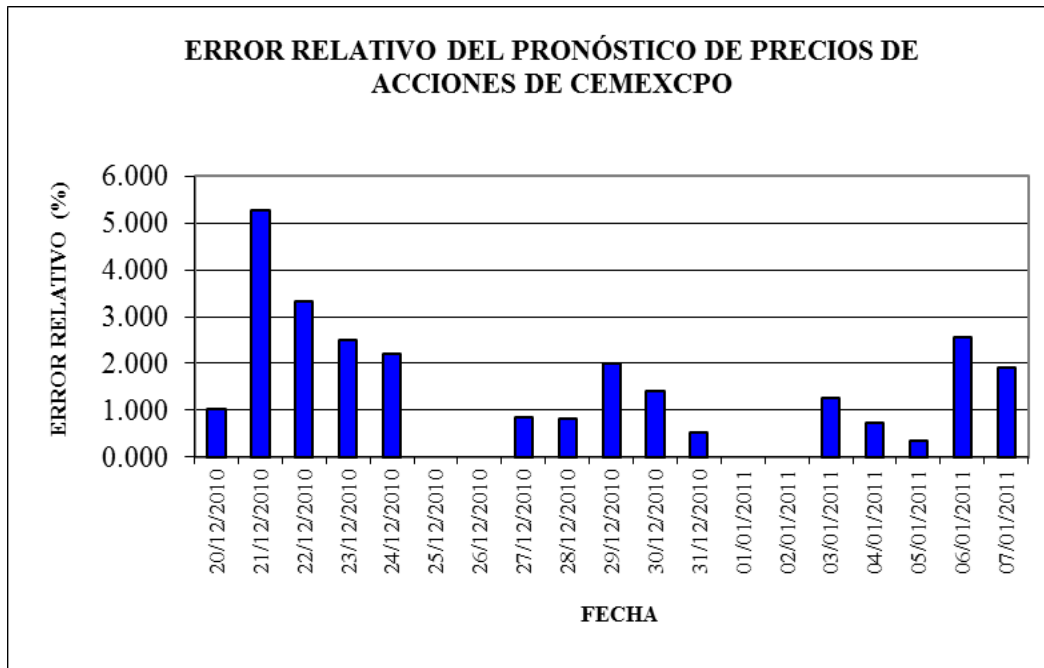
Gráfica 15.1. Error relativo de los errores en el pronóstico del precio de acciones de TELMEXL con respecto a los datos reales.



Gráfica 15.2. Error relativo de los errores en el pronóstico del precio de acciones de GMODELOC con respecto a los datos reales.



Gráfica 15.3. Error relativo de los errores en el pronóstico del precio de acciones de CEMEXCPO con respecto a los datos reales.



15.4 Conclusiones

El presente trabajo muestra la gran valía del uso de las redes neuronales diferenciales como metodología alternativa a los modelos econométricos, sobre todo en el caso de realizar análisis de sistemas dinámicos con comportamientos no lineales (el cual es el caso más común en Finanzas y Economía).

Se puede observar que los errores relativos, en el pronóstico, obtenidos para el caso de GMODELLOC no fue más allá del 2.5% en el periodo analizado. Mientras que para TELMEXL y CEMEXCPO fue un error relativo menor al 8% y 6% respectivamente. Esto permite a inversionistas tener en el corto plazo un mejor conocimiento de los precios de las acciones y en consecuencia tomar decisiones de inversión más adecuadas al armar sus portafolios de inversión.

Es muy importante destacar el hecho de que las redes neuronales diferenciales requieren de una considerable menor cantidad de datos para llevar a cabo el periodo de entrenamiento de la red, y además proporcionan resultados más precisos que las redes neuronales artificiales discretas.

Las redes neuronales artificiales en general tienen la gran ventaja sobre la gran mayoría de métodos de análisis de comportamientos dinámicos de variables de que pueden procesar e identificar patrones de comportamiento con gran precisión y relativamente rápido de series de datos aparentemente incoherentes, al grado de darles algún sentido a través de identificar algún patrón o secuencia en los datos que alimentan a la red.

Asimismo, este documento da lugar a poder realizar un extensión que permita a los inversionistas conocer no sólo un pronóstico de los precios de las acciones sino de conocer los rendimientos potenciales en sus portafolios de inversión o al menos de cada una de las acciones. Lo cual se realizará en una futura extensión de este documento de investigación.

15.5 Bibliografía

- Aguilar, N. C., A. y Chairez, I. (2006). "Neuro Tracking Control for Immunotherapy Cancer Treatment". *IJCNN '06 International Joint Conference on Neural Networks. IEEE*. Vancouver, BC, pp. 5316 - 5323.
- Bilbrey, J. y Riley, Neil. (2011). Neural Network Based Object-Oriented Framework for Simulating Stock Market Trading Strategies. *International Journal of Business and Social Science*. Vol. 2 No. 15, 33-38.
- Cabrera-Llanos, A.I., Poznyak A. P. y Aranda, T., J. S. (2002). Identification of a Fedbatch Fermentation Process: Computational and Laboratory Experiments. *Bioprocess and Biosystems Engineering*, 24, pp. 319-327.
- Cabrera-Llanos, A.I., Aranda, T., J. S. (2007). Estimating the Trehalose Cyto-plasmatic Content. *10th International Symposium on Computer Applications on Biotechnology*, Cancún, México, IFAC press.
- Cabrera-Llanos, A.I. y Ramírez-Sotelo, M. Guadalupe (2007). Sensores suaves en procesos biotecnológicos. *Superficies y Vacío*, 20(2), pp. 1-5.
- Cabrera-Llanos, A.I., Ortiz-Arango, F. (2012). Pronóstico de rendimiento del IPC mediante el uso de redes neuronales diferenciales. *Revista Contaduría y Administración*. Vol. 57, No. 2, pp. 63-81.
- Dunis, C. L., Laws, J. y Sermpinis, G. (2010). Modeling and Trading the EUR/USD Exchange Rate at the ECB Fixing. *European Journal of Finance*, No. 16/6, pp. 541- 560.
- Fonseca, H., Ortiz, V. H. y Cabrera, A. I.. Redes Neuronales Estocásticas Aplicadas al Modelo Dinámico de Glucosa para Pacientes Diabéticos. *XXVI SOMIB Congreso Nacional de Ingeniería Biomédica*, (CINVESTAV), México, noviembre 2003.
- Hanias, M., Curtis, P. y Thalassinos, J. (2007). Prediction with Neural Networks: The Athens Stock Exchange Price Indicator. *European Journal of Economics, Finance and Administrative Sciences*, Issue 9: 21-27.

- Haykin, S. (1999). *Neural Networks a comprehensive foundation*, Prentice Hall.
- Hutchinson, J. M., Andrew W. Lo y Tomaso P. (1994). A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks. *Journal of Finance*, 49, pp. 851–889.
- Isidori, A. *Nonlinear control systems 2*. (1999) Springer-Verlag, London.
- Jagielska, I. y Jaworski, J. (1996). Neural network for predicting the performance of credit card accounts. *Computational Economics*, 9(1), pp. 77-82.
- Kishore, P. y Prasad, A. (2011). Development of Improved Artificial Neural Network Model for Stock Market Prediction. *International Journal of Engineering Science and Technology*, Vol. 3, No. 2 Feb. 2011, 1576-1581.
- Kuan, C. y White, H. (1994). Artificial Neural Networks: An Econometric Perspective. *Econometric Reviews*, 13, pp. 1–91.
- Lean Yu, Shouyang Wang y Kin Keung Lai (2007). *Foreign Exchange Rate Forecasting with Artificial Neural Networks*. Springer.
- Mehrara, M., Moeini, A., Ahrari, M. y Ghafari, A. (2010). Using Technical Analysis with Neural Network for Forecasting Stock Price Index in Teheran Stock Exchange. *Middle Eastern Finance and Economics*, Issue 6: 50-61
- McCulloch, W.S y Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 5(4), pp. 115-133.
- McNelis, P. D. (2005). *Neural Networks in Finance: Gaining Predictive Edge in the Market*. Academic Press.
- Ortiz-Arango, F.; Cabrera-Llanos, A.I. y Cruz-Aranda, F. Modelado del comportamiento del tipo de cambio peso-dólar mediante redes neuronales diferenciales. *Revista Estocástica. Finanzas y Riesgo*. 2(1), pp. 49-63. January-June 2012a.
- Ortiz-Arango, F.; Cabrera-Llanos, A.I. y Danvila, I. Historical Identification and Forecast Values of IBEX 35 & IPC Financial Indices using Differential Neural Networks. *European Journal of Economics, Finance and Administrative Sciences*. 54, pp. 161-173. October, 2012b.
- Pozniak, A. S., Yu, W. y Sánchez. E. N. Nonlinear Adaptive Trajectory Tracking Using Dynamic Neural Networks. Identification via dynamic neural control. *IEEE Transactions on Neural Networks*, 10 (6) (1999) 1402-1411.

Pozniak, A. S., Sanchez. E. y Yu, W. *Differential Neural Networks for Robust Nonlinear Control*, 2001. World Scientific Publishing: Singapore.

Refenes, A. P. N. y Holt, W. T. (2001). Forecasting volatility with neural regression: A contribution to model adequacy. *IEEE Transactions on Neural Networks*, 12(4), pp. 850-865.

Sing, K. y Kumar, B. (2010). An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting. *Journal on Computer Science and Engineering*. Vol. 02, No. 06, 2104-2109.

Smith, K. y Gupta J. (2000). Neural networks in business: Techniques and applications for the operations researcher. *Computer & Operation Research*, 27, pp.1023-1044.

Zhang G. P. y Berardi V. L. (2001). Time series forecasting with neural network ensembles: an application or exchange rate prediction. *Journal of the Operational Research Society*, Vol. 52, pp. 652-664.

Capítulo 16

Uso de un modelo de Redes Bayesianas para Medir el Riesgo Operacional de una Empresa Financiera

Griselda Dávila Aragón

ECEE, Universidad Panamericana

16.1 Introducción

Los métodos comúnmente utilizados para medir el riesgo operacional están basados en su mayoría en modelos lineales, en forma de modelos de series de tiempo, modelos econométricos, modelos empíricos actuariales y en algunos casos en la teoría de valores extremos. El riesgo operacional en particular se encuentra con la limitante de no contar con bases de datos confiables que contemplen los diferentes aspectos del riesgo operacional aunado a esto se ha observado que existe una compleja interacción entre las variables de riesgo que lo caracterizan, razón por la cual diversos métodos no lineales han sido propuestos, siendo uno de ellos el que está basado en las Redes Bayesianas.

El objetivo fundamental del presente documento es aplicar la metodología de Redes Bayesianas para modelar el riesgo operacional en una Institución Financiera mexicana que administra una base de datos de nivel nacional y provee información en línea de carácter financiero, se pretende corroborar que ésta es una metodología adecuada para establecer el capital de riesgo operativo de acuerdo a las necesidades y requerimientos de la organización considerando mejores prácticas y estándares internacionales que satisfagan la normatividad vigente.

La Institución se encuentra en proceso de revisión de su Plan Estratégico de Implementación y del Manual de Políticas y Procedimientos para la Administración del Riesgo

Operacional de acuerdo a las nuevas disposiciones publicadas en 2011 para este tipo de Instituciones Financieras en México; aprovechando esta actualización se ha decidido implementar un modelo de Redes Bayesianas para tener una evaluación alternativa del riesgo operacional.

El modelo aplicado refleja las características particulares de las operaciones de la Institución, incorporando en el proceso de cuantificación del riesgo operativo la opinión de los expertos. La Dirección de Riesgos podrá a partir la implementación del modelo, obtener entre otras cosas, ideas para modificar el proceso de operación, establecer controles que promuevan la mitigación del riesgo y adecuar el programa de operación existente a las nuevas necesidades y requerimientos de la organización. Permitiendo con ello la gestión activa del riesgo operacional y la toma de decisiones del día a día; conocer las pérdidas reales asociadas a este riesgo así como la identificación y priorización de riesgos potenciales y reales.

La trascendencia de este trabajo radica en que esta empresa sería una de las primeras Instituciones de índole financiera en nuestro país que contarían con un sistema de evaluación de riesgo operacional basado no solamente en los “métodos estándar” sino que tendría como elemento adicional de verificación una metodología basada en redes bayesianas. Esto sin duda alguna situaría a esta empresa a la vanguardia de las técnicas de medición de riesgo a nivel nacional y la equipararía con instituciones similares a nivel internacional de países desarrollados.

16.2 Riesgo Operacional

16.2.1 Objetivo e Importancia de medir el riesgo operacional.

El riesgo operacional es el más antiguo de todos los riesgos que enfrentan las instituciones financieras; es inherente a toda actividad en que intervengan personas procesos y plataformas tecnológicas y es el único riesgo que no surge de la toma de una posición de riesgo como es el caso del riesgo de crédito y de mercado.

La posibilidad de sufrir un fraude interno o externo, fallas en los sistemas con consecuencias de inoperatividad de la empresa, ha existido siempre pero es hasta hace pocos años como consecuencia de la quiebra de importantes instituciones en el sector financiero originadas por fallas en sus controles operativos que se considera la posible ocurrencia de estos eventos como una clase de riesgo y tanto las instituciones como los reguladores han reconocido el desconocimiento que de él se tiene y la falta de herramientas para gestionarlo. A partir de entonces se ha puesto especial énfasis en definir medidas que cuantifiquen el riesgo operacional considerando el complejo proceso que ello implica dada la gran diversidad de factores que lo originan.

El cambio más importante se dio después de 2007 cuando por primera vez en la historia de la industria bancaria se obligó a los bancos a establecer un capital específico para el riesgo operacional. El Comité de Basilea y la industria bancaria definieron formalmente el riesgo operacional y con ello se facilitó su medición con una medida definida por los reguladores y se hizo posible comparar los perfiles de riesgo operacional de las empresas.

El Riesgo Operacional fue definido por el Comité de Basilea II como: “el riesgo de pérdida resultante de procesos internos inadecuados o fallidos, personas y sistemas, o de acontecimientos externos” El comité indica que esta definición incluye el riesgo legal, pero no se limita a, la exposición a multas, sanciones o daños punitivos derivados acciones de supervisión, así como de acuerdos privados. (BCBS 2006)

La Circular Única de Bancos (CUB) publicada en el Diario Oficial de la Federación el 2 de diciembre de 2005 define el Riesgo Operacional como “ la pérdida potencial por fallas o deficiencias en los controles internos, por errores en el procesamiento y almacenamiento de las Operaciones o en la transmisión de información, así como por resoluciones administrativas y judiciales adversas, fraudes o robos y comprende, entre otros, al riesgo tecnológico y al riesgo legal, en el entendido de que: El riesgo tecnológico se define como la pérdida potencial por daños, interrupción, alteración o fallas derivadas del uso del hardware, software, sistemas, aplicaciones, redes y cualquier otro canal de transmisión de información en la prestación de servicios bancarios a los clientes de la Institución. El riesgo legal se define como la pérdida potencial por el incumplimiento de las disposiciones legales y administrativas aplicables, la emisión de resoluciones administrativas y judiciales desfavorables y la aplicación de sanciones, en relación con las Operaciones que la Institución lleva a cabo” (CUB 2005)

El Comité de Basilea publicó diez principios para la gestión interna y supervisión del riesgo operacional, organizado en torno a cuatro temas:

16.2.2 Principios para la gestión interna y supervisión del riesgo internacional.

16.2.2.1 El desarrollo de un marco adecuado para la administración de riesgos.

Principio 1: El Consejo de administración⁴ deberá conocer cuáles son los principales aspectos de los riesgos operativos para el banco, en tanto que categoría de riesgo diferenciada, y deberá aprobar y revisar periódicamente el marco que utiliza el banco para la gestión de este riesgo. Este marco deberá ofrecer una definición de riesgo operativo válida para toda la empresa y establecer los principios para definir, evaluar, seguir y controlar o mitigar este tipo de riesgos.

Principio 2: El consejo de administración deberá asegurar que el marco para la gestión del riesgo operativo en el banco esté sujeto a un proceso de auditoría interna eficaz e integral por parte de personal independiente, capacitado y competente. La función de auditoría interna no deberá ser directamente responsable de la gestión del riesgo operativo.

Principio 3: La alta gerencia deberá ser la responsable de poner en práctica el marco para la gestión del riesgo operativo aprobado por el consejo de administración. Dicho marco deberá ser aplicado de forma consistente en toda la organización bancaria y todas las categorías laborales deberán comprender sus responsabilidades al respecto. La alta gerencia también deberá ser

responsable del desarrollo de políticas, procesos y procedimientos destinados a la gestión de estos riesgos para todos los productos, actividades, procesos y sistemas relevantes para el banco.

16.2.2.2 Identificación, evaluación, seguimiento y cobertura/control.

Principio 4: los bancos deberán identificar y evaluar el riesgo operativo inherente a todos sus productos, actividades, procesos y sistemas relevantes. Además, también deberán comprobar que antes de lanzar o presentar nuevos productos, actividades, procesos o sistemas, se evalúa adecuadamente su riesgo operativo inherente.

Principio 5: Los bancos deberán vigilar periódicamente los perfiles de riesgo operativo y las exposiciones sustanciales a pérdidas. La alta gerencia y el consejo de administración deberán recibir información pertinente de forma periódica que complemente la gestión activa del riesgo operativo.

Principio 6: Los bancos deberán contar con políticas, procesos y procedimientos para controlar y cubrir los riesgos operativos más relevantes. Además, deberán reexaminar periódicamente sus estrategias de control y reducción de riesgos y ajustar su perfil de riesgo operativo según corresponda, utilizando para ello las estrategias que mejor se adapten a su apetito por el riesgo y a su perfil de riesgo.

Principio 7: Los bancos deberán contar con planes de contingencia y de continuidad de la actividad, que aseguren su capacidad operativa continua y que reduzcan las pérdidas en caso de interrupción grave de la actividad.

16.2.2.3 Función de los supervisores

Principio 8: Los supervisores bancarios deberán exigir a todos los bancos, sea cual sea su tamaño, que mantengan un marco eficaz para identificar, evaluar, seguir y controlar o mitigar sus riesgos operativos más relevantes, como parte de su aproximación general a la gestión de riesgos.

Principio 9: Los supervisores deberán realizar, directa o indirectamente, una evaluación periódica independiente de las políticas, prácticas y procedimientos con los que cuentan los bancos para gestionar sus riesgos operativos. Además, deberán cerciorarse de que se han puesto en marcha los mecanismos necesarios para estar al tanto de cualquier novedad que se produzca en un banco.

16.2.2.4 La función de la divulgación de la información.

Principio 10: Los bancos deberán proporcionar información pública suficiente para que los participantes del mercado puedan evaluar sus estrategias de gestión del riesgo operativo.

Y estableció una división de las pérdidas operacionales (BSBC, 2006) para los bancos en ocho líneas de negocio

1. Finanzas Corporativas
2. Negociación y Ventas
3. Banca de Menudeo
4. Banca Comercial
5. Pagos y Liquidación
6. Servicios de Agencia
7. Administración de Activos
8. Intermediación Minorista

y siete tipos de eventos:

1. Fraude interno.
2. Fraude externo.
3. Relaciones laborales y seguridad en el puesto de trabajo
4. Clientes, productos y prácticas empresariales.
5. Desastres naturales y otros acontecimientos.
6. Incidencias en el negocio y fallos en los sistemas
7. Ejecución, entrega y gestión de procesos.

Una vez identificadas y clasificadas las posibles fuentes de riesgo operacional se debe cuantificar la posible pérdida. Los mecanismos para la atribución de capital para el riesgo operacional deben estar basados en el riesgo, ser transparentes, escalables y justos. Los requisitos de capital se deben modificar directamente con los niveles de riesgo controlables, y deben ofrecer incentivos para la gestión del riesgo operacional a fin de mejorar las decisiones operativas y aumentar la rentabilidad ajustada al riesgo sobre el capital.

16.2.3 Capital de Riesgo para Riesgo Operacional

No toda pérdida de riesgo operacional debe ser atribuida un capital de riesgo. Considerando que existen pérdidas esperadas (aquellas que resultan de la operación “normal” del negocio y que la institución está dispuesta a absorber como un costo del negocio dado su bajo grado de severidad) y no esperadas. Las primeras deberán estar consideradas en el presupuesto de la Institución, incluidas en el precio del producto o servicio y con ello no asignarles capital de riesgo.

En este sentido, la asignación del capital de riesgo operacional deberá estar basada en aquellas pérdidas no esperadas (aquellas pérdidas no usuales que ocurren rara vez y su ocurrencia es normalmente de alta severidad), las cuales pueden ser a su vez divididas en:

- Pérdidas severas pero no catastróficas. Inesperados fallos de funcionamiento que producen importantes pérdidas económicas a la institución.
- Pérdidas catastróficas. Un solo evento pueden poner fin a la institución.

Las instituciones deben siempre tratar de minimizar o mitigar el riesgo operacional considerando las implicaciones en costos que ello implica.

16.2.4 Métodos de medición de riesgo operacional más usados.

La Comisión Nacional Bancaria y de Valores, basada en los acuerdos del Comité de Basilea II sobre el requerimiento de capital por riesgo operacional, en el Art. 2 Bis 111 de la Circular Única de Bancos establece que: “Las Instituciones para calcular el requerimiento de capital por su exposición al Riesgo Operacional, deberán utilizar alguno de los métodos siguientes: (CUB, 2005)

16.2.4.1 Método del Indicador Básico

Es una medida cero-dimensional cuyo objetivo es cubrir el riesgo operacional con un capital mínimo equivalente al 15% del promedio de los tres últimos años de sus ingresos netos anuales positivos. Estos son el resultado de sumar los ingresos netos por concepto de intereses más otros ingresos netos ajenos a intereses y deberá ser calculado antes de cualquier deducción de reservas y gastos operativos.

Una vez calculado el requerimiento de capital para riesgo operativo (RCRO) se compara con el capital requerido para riesgo de mercado (RCRM) y crédito (RCRC) del mismo periodo de cálculo bajo la siguiente regla:

- RCRO es menor que el 5% del promedio de los últimos 36 meses de la suma de los RCRM + RCRC, el RCRO será igual al 5% del promedio de los últimos 36 meses de la suma de los RCRM + RCRC
- RCRO es superior al 15% del promedio de los últimos 36 meses de la suma de los RCRM + RCRC, el RCRO será igual al 15% del promedio de los últimos 36 meses de la suma de los RCRM + RCRC

El método entró en vigor a partir del 1 de enero de 2008 y se deberá constituir en un plazo no mayor a tres años.

16.2.4.2 Método Estándar

Medida unidimensional que considera las ocho líneas de negocio. El requerimiento de capital para riesgo operacional es la suma del promedio de los ingresos netos positivos de los últimos 3 años ponderados por un factor β .

Figura 16.1. Factores para ponderar el riesgo operativo por línea de negocio. (BSBC, 2006)

Línea de negocio	β_i Estándar
Finanzas corporativas	18%
Negociación y ventas	18%
Banca minorista	12%
Banca comercial	15%
Pago y liquidación	18%
Servicios de agencia	15%
Administración de activos	12%
Intermediación minorista	12%

En la Figura 16.1 se señalan los factores β que deberán aplicarse al ingreso bruto por línea de negocio

$$RCRO = \sum_{i=1}^N EI_i * \beta_i \quad (16.1)$$

Donde:

EI_i es el indicador de exposición en cada línea de negocio.

Este método puede dar una doble contabilidad para altas tasas de incumplimiento en las líneas de negocio, por ejemplo en préstamos de alto riesgo donde dada la alta probabilidad de incumplimiento por parte del prestatario eleva el capital requerido para el riesgo crédito y segundo eleva el capital requerido para riesgo operativo por los altos márgenes de la línea de negocio considerada .

16.2.4.3 Método Estándar Alternativo

El marco es el mismo que en el Método estándar, salvo que las ocho líneas de negocio son reagrupadas en:

- Banca tradicional: banca minorista y banca comercial
- Resto actividades: financiamiento empresarial, negociación y ventas; pagos y liquidación; servicios de agencia; administración de activos e intermediación minorista estándar

Para la Banca tradicional el indicador de exposición es reemplazado por:

$$EI_{Bt} = 0.035 * TP \quad (16.2)$$

Donde:

TP es el promedio de los últimos tres años del total de préstamos minoristas pendientes de pago y anticipos

Para el resto de actividades se calcula la β sobre cada línea de negocio

La Circular emitida por la comisión no contempla un Método que si está incluido en el documento del Comité de Basilea II:

16.2.4.4 Método de Medición Avanzada (AMA)

Este método es un modelo actuarial que utiliza el enfoque estadístico de distribución de pérdidas que captura la sensibilidad del riesgo enfrentado por la entidad financiera y donde el requerimiento de capital es la medida del riesgo producido por el modelo interno de riesgo operacional de la institución. (Chernobai, Jorion and Yu, 2011)

En este método se genera una medida de riesgo haciendo uso del sistema interno de la institución, utilizando criterios cualitativos y cuantitativos.

Los elementos a considerar para este método son (Fuentes relevantes):

- Datos internos
- Datos externos
- Análisis de escenarios
- Factores relacionados con el entorno de negocio y con el control interno

Las metodologías AMA se engloban dentro de los llamados enfoques bottom-up; ya que calculan el capital económico a partir de datos internos de pérdidas distribuidos por su tipología y unidad de negocio; tras este cálculo específico, se computa el capital para la institución en su conjunto.

Hasta el momento, el comité no ha definido claramente cómo estos elementos deben ser empleados. Los reguladores sólo han establecido que la institución necesita “un enfoque creíble, transparente, bien documentado y verificable en su sistema global de cálculo del riesgo operacional para la ponderación de estos elementos fundamentales”

El método considera tres enfoques:

- a. **Medición Interna.** El regulador determina el Índice de Exposición y un múltiplo, en el cual convierte la Pérdida Esperada en Pérdida no Esperada, en forma análoga para todo el gremio, y cada Entidad obtiene, solamente, estimaciones de la Probabilidad de Fallo y de la Proporción de Pérdida dado el Fallo.
- b. **Distribución de Pérdidas.** Es un modelo de VaR por Línea de Negocio y Tipo de Evento de Pérdida, en donde el VaR total es la suma de los VaR de todas las combinaciones sin considerar correlaciones.
- c. **Tarjetas de Puntaje.** Se calcula un nivel de riesgo tomando como base toda la estadística de eventos de pérdida disponible para la Entidad (habida cuenta de que es estadísticamente significativo), y se redistribuye por Línea de Negocio, en función de una tarjeta de puntaje diseñada exprofeso y que contiene el seguimiento de ciertas medidas de control.

El capital regulatorio del AMA está dado por la suma de la Perdida Esperada y la no esperada. En caso que la Institución pueda demostrar que se está registrando apropiadamente la Perdida Esperada, entonces podrá disminuir este monto a únicamente la Perdida no Esperada.

Para poder utilizar el método AMA se requiere de autorización del regulador local, que dependerá del cumplimiento de los siguientes requisitos: (BASILEA II)

- La Administración debe estar fuertemente involucrada con el marco de administración de riesgo operacional.
- Se debe contar con un sistema de administración de riesgo operacional íntegro.
- Se debe contar con recursos suficientes en el uso del enfoque en las líneas de negocio más importantes así como en las áreas de control y auditoría.
- Estándares cualitativos y cuantitativos.

En una primera fase de los modelos AMA se usan modelos que tratan de ajustar una distribución de pérdidas (Loss Distribution Approach) de la cual se puede extraer una cifra de VaR.

16.2.5 VaR Operacional (OpVaR)

El Valor en Riesgo (VaR) se define como la cantidad máxima de dinero que se puede perder en

una cartera durante un período determinado de tiempo, con un nivel de confianza dado. (Jorion, 2001).

El VaR ofrece un enunciado de probabilidad acerca del cambio potencial en el valor del portafolio que resulta de cambios en los factores de mercado en un periodo de tiempo específico. El VaR simplemente indica lo probable (o improbable) que es que la medida VaR se supere, pero no indica qué ocurrirá si la pérdida es superior o cuándo ello puede acontecer.

Existen dos conceptos clave en el cálculo del VaR:

1. Derivan de la distribución esperada del valor la cartera o los rendimientos del portafolio en el horizonte elegido. La distribución se puede crear utilizando alguno de los métodos siguientes: Suponiendo una distribución específica (generalmente la normal, método paramétrico). Usando la distribución empírica de los factores de riesgo (VaR por simulación histórica) Simulando un modelo paramétrico para los cambios de los factores de riesgo (MonteCarlo).
2. Identificar el nivel de confianza requerido. Ello se traduce en encontrar el percentil adecuado de la distribución de utilidades y pérdidas.

El VaR de una entidad X dado un nivel de confianza α se define como:

$$VaR_{\alpha}(X) = \inf\{x \in R | P[-X \leq x] \geq \alpha\} \quad (16.3)$$

El enfoque amplio para la creación del requerimiento de capital por exposición al riesgo operacional (OpVaR) gira en torno a la obtención de una sola figura que se define a menudo como un cuantil alto de una distribución de pérdidas¹ agregadas que se proyecta sobre un periodo de tiempo definido tomando como base datos históricos.

Para cada categoría de riesgo, se determina el percentil 99.9% de la distribución de pérdidas totales. Este percentil representará el VaR Operativo para cada categoría de riesgo.

El VaR Operativo de la distribución de pérdidas globales se obtiene agregando las pérdidas totales de todas las categorías de riesgo bajo cierto supuesto de correlación. (Supuesto conservador $\rho = 1$)

El VaR de riesgo de mercado y riesgo de crédito se calcula asumiendo que los datos siguen una distribución normal (Jorion 2001); sin embargo para riesgo operacional las pérdidas no tienen normalidad, las caracteriza un sesgo positivo y tienen colas pesadas. Es este último punto la principal preocupación en el requerimiento de capital económico. En el modelado de la severidad de las pérdidas, una combinación de enfoques ha sido empleado, incluyendo la Teoría de Valores Extremos.

Conceptualmente es razonable creer que la ocurrencia de un evento de riesgo operacional está relacionada con el volumen de transacciones de la línea de negocio por lo que es necesario

¹ Definido en 99.9% por el Comité de Basilea

descomponer la línea en las partes que la componen y después modelar el proceso de cada componente sin perder de vista el número de transacciones y los eventos ocurridos de riesgo operacional.

16.2.6 Modelos para medir el riesgo operacional bajo el AMA

El riesgo operacional necesita ser analizado desde las causas y no los efectos (Chonawee, Chris y Lucas, 2006); para asegurar una apropiada clasificación del riesgo, es vital comprender la forma en que el riesgo de pérdidas operacionales surge dentro de la estructura y funcionamiento de la organización. Los modelos de causa proporcionan una base sobre la cual la administración puede intervenir para lograr modificar el perfil de riesgo.

El principal problema en los modelos para medir riesgo operacional es la falta de datos. Incluso aún si se cuenta con una base de datos definir cuáles de ellos son importantes en materia de riesgo operacional no es una discusión trivial (Power, 2005). Los datos internos registrados de eventos de menor frecuencia son pocas veces suficientes para modelar la distribución de pérdidas con la precisión requerida. El uso de distribuciones paramétricas de pérdidas requiere de estimación de parámetros y toman como punto de partida la base de datos existente. Incluso para la teoría de valores extremos, una muestra pequeña hace que la cola de la distribución sea muy sensible a las inclusiones o exclusiones de eventos con baja frecuencia pero de alto impacto.

En el afán de superar la falta de información interna, se ha intentado utilizar datos externos, lo cual no es un ejercicio simple, en particular en el caso del riesgo operacional. Existen problemas cualitativos tales como la calidad de la información de las partes externas, la diferencia de las prácticas de administración en las empresas y la falta de actualización de datos entre otros.

La falta de datos internos y la complejidad de incluir datos externos conlleva a sugerir la inclusión de la opinión de expertos. Un experto es una persona cuya experiencia y conocimiento le permite hacer conjeturas suficientemente creíbles acerca de cómo las operaciones de la institución afectan el perfil de riesgo de la misma. En un inicio su opinión puede ser utilizada como un sustituto de los datos y puede proporcionar información valiosa de las operaciones de la institución que es difícil de capturar de los datos por sí solos. El reto del modelador es incorporar estas aportaciones en el modelo de riesgo operacional.

Usando métodos no lineales es posible incorporar información cualitativa como decisiones gerenciales, competencias y preferencias a la medición. Los métodos más empleados son:

- Modelos multifactoriales. Es esencialmente un modelo de regresión en la que la variable a explicar es un atributo del riesgo operacional y las variables explicativas son los factores de riesgo. En este modelo la variable por explicar son pérdidas de operación y las variables independientes, son factores de entorno de control, es decir, el tiempo de inactividad del sistema, número de empleados, calidad de los datos y número total de transacciones.
- Lógica difusa, Utiliza un conjunto lógico multivariante que reconoce que las decisiones humanas no son binarias y permite grados en las formulaciones. (Hoffman 2002)

- **Redes neuronales.** El método es útil modelando relaciones complejas entre variables que sería muy difícil establecer utilizando métodos lineales. Las redes consisten en nodos con valores de entrada, salida y variables intermedias. Técnicas de minería de datos son usadas para entrenar al modelo usando complejos algoritmos que aprendan la relación entre las variables. El modelo es calibrado de forma tal que sus salidas son casi tan cercanas a las salidas reales de la información disponible.
- **Sistemas dinámicos.** Este enfoque implica el uso de aportaciones de los expertos para trazar una red de relaciones de causa y efecto entre las variables que afectan el riesgo operacional de una unidad de negocio. La relación entre cada conjunto de causa y efecto de las variables se cuantifica mediante la combinación de datos y aportaciones de expertos para obtener una gráfica de dos ejes. (Miccolis and Shah, 2000)
- **Las Redes Bayesianas.** Una clase de modelo grafico que puede ser usado para modelar dependencias causales entre las variables aleatorias para la determinación de factores de riesgo operacional y asignar el capital de riesgo operacional (Neil, Fenton y Tailor, 2005) Al utilizar redes bayesianas es posible considerar los cuatro elementos esenciales de Método de Medición Avanzada (datos internos, datos externos, análisis de escenarios y ambiente de control) de manera integral. La ventaja principal es estar en posibilidades de determinar las relaciones causales.

Pueden parecer complejas para el modelado pero son una estructura que representa el entendimiento de un proceso y sus dependencias (causa-efecto). Una representación gráfica de un proceso con sus dependencias no debe ser una tarea ardua. El problema de calcular las probabilidades numéricas es esencialmente el mismo que el que se tiene al realizar una autoevaluación de control de riesgo operacional.

Las redes bayesianas han surgido como una herramienta para integrar criterios y aspectos cualitativos subjetivos con datos cuantitativos internos y externos facilitando la comprensión causal dentro de las estimaciones (Guidici , 2004; Neil, Fenton y Tailor, 2005)

16.3 Redes Bayesianas²

16.3.1 Introducción a la estadística bayesiana

En teoría de la probabilidad clásica, se asume que las estadísticas de la muestra pertenecen a cierta población con una distribución específica, la cual es definida por el conjunto de parámetros con un valor fijo. La tarea para el estadístico es estimar los parámetros lo mejor posible basado en los datos disponibles. Cuando es posible realizar experimentos varias veces y obtener así una muestra suficientemente grande para asignar valores a estos parámetros.

Thomas Bayes (1702-1761) estudió e investigó el problema de la determinación de la probabilidad de las causas a través de los efectos que se observan.

² Las ideas presentadas en este capítulo siguen de cerca se basan en la teoría presentada por Madsen y Kjaerulff (2008) y Cowell (1999)

Los estadísticos bayesianos permiten a los parámetros ser variables aleatorias. Las afirmaciones hechas sobre las características de una población son necesariamente dependientes no sólo de las observaciones empíricas o de los datos (información objetiva), sino también de cualquier conocimiento disponible para el estadístico, previo al inicio de las observaciones (información subjetiva). Este conocimiento puede venir en forma de datos de un lugar diferente y se considera que tiene un cierto grado de relevancia para la población observada. Más importante aún, también puede venir de información obtenida de partes interesadas y expertos, es decir, cuya familiaridad con el tema los hace una fuente creíble.

Como consecuencia, en donde la probabilidad clásica trata principalmente con la evaluación de declaraciones incondicionales de probabilidad, como: “la probabilidad del evento A es x ” denotada por $P(A) = x$, el vocabulario bayesiano se expande a utilizar probabilidades condicionales en sus afirmaciones como: “la probabilidad del evento A dado el que el evento B ha ocurrido es y ” y se denota por $P(A|B) = y$

La manipulación de tales probabilidades consiste en tratarlas como funciones de variables que usan las reglas del cálculo de probabilidades. La regla fundamental del cálculo de probabilidades y piedra angular en la estadística Bayesiana es el teorema de probabilidad condicional de los eventos A y B :

$$P(A \cap B) = P(A|B)P(B) \quad (16.4)$$

Dado que la función $P(A \cap B)$ es simétrica, el teorema se puede expresar también de la siguiente manera:

$$P(A \cap B) = P(B|A)P(A) \quad (16.5)$$

Igualando ambas ecuaciones y despejando $P(B|A)$ se obtiene el Teorema de Bayes

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (16.6)$$

El teorema anterior es interpretado como: la probabilidad “posteriori” $P(B|A)$ es igual a la probabilidad a priori $P(B)$ multiplicada por la razón $\frac{P(A|B)}{P(A)}$; es decir, la información previa acerca de B puede ser utilizada para revisar la probabilidad de B .

Al aplicar el teorema de Bayes a las distribuciones sobre los parámetros del modelo, por ejemplo, donde las observaciones \underline{x} tomadas de la muestra de la población con una distribución de probabilidad definida por el parámetro θ se obtienen la siguiente forma:

$$P(\theta|\underline{x}) = \frac{P(\underline{x}|\theta)P(\theta)}{P(\underline{x})} \quad (16.7)$$

Esta expresión permite entender el teorema de Bayes en términos de la información objetiva y subjetiva. $P(\theta)$ es la información “a priori” de la población y es subjetiva. En este caso es una probabilidad incondicional que representa la incertidumbre acerca del parámetro θ . La función $P(\underline{x}|\theta)$ es comúnmente llamada la probabilidad del conjunto de datos \underline{x} , denotada también como $L(\underline{x}|\theta)$ y se interpreta como la probabilidad de observar un conjunto de ciertos datos \underline{x} dado que ciertas características de la población θ son ciertas.

Al combinar la información subjetiva con las observaciones empíricas (la información “a priori” y las verosimilitudes), se obtiene la probabilidad posteriori, es decir, la probabilidad de que el parámetro θ tome ciertos valores dado que las observaciones \underline{x} se han hecho, lo cual se denota $P(\theta|\underline{x})$. Si se suma $P(\theta|\underline{x})P(\underline{x})$ sobre los valores de \underline{x} , se obtienen la probabilidad marginal del parámetro θ .

La probabilidad incondicional de los datos $P(\underline{x})$, sirve como una constante de escala la regla puede ser reexpresada de la siguiente manera:

$$P(\theta|\underline{x}) = P(\underline{x}|\theta)P(\theta) \approx L(\underline{x}|\theta)P(\theta) \quad (16.8)$$

O en otras palabras:

$$\text{Posteriori} \approx \text{Verosimilitud} * \text{A priori}$$

Intuitivamente la idea bayesiana dice que las conjeturas acerca de la población son una combinación del conocimiento previo de la población y las observaciones hechas con respecto a esa misma población.

No es sólo la información “a priori” la que tienen influencia; el grado de confianza en la información del experto también tiene un efecto. El análisis bayesiano simplemente formaliza este concepto: Las densidades posteriores toman en cuenta en menor o mayor grado la información de las creencias dependiendo de la confianza que sobre ésta se tenga.

No hay una respuesta simple a la pregunta sobre cuánta información de la muestra debe ser usada y que tanta confianza se debe tener en las creencias previas. La única manera en que se puede saber cuáles son las mejores elecciones para el problema en análisis, como en cualquier modelo estadístico, es llevar a cabo un análisis de “back testing”.

La información sobre el perfil del riesgo operacional de una empresa es a menudo incompleta; se deriva de fuentes cuantitativas y cualitativas y se actualiza continuamente con nuevos datos. La estadística bayesiana por lo tanto, ofrece una forma metodológica práctica para evaluar este tipo de riesgo.

16.3.2 Definición de Redes Bayesianas

Las redes de creencias, las redes causales, los modelos gráficos dirigidos o modelos generativos son genéricamente tipos de modelos gráficos. Los modelos gráficos son una combinación de la teoría de la probabilidad y la teoría de grafos. Es el resultado de la convergencia de la evolución de la modelación estadística, ingeniería e inteligencia artificial que inició en la década de los 80's. (Alexander, 2002)

Inicialmente sus cálculos excesivos basados en la teoría de probabilidad hacían inviable su uso; sin embargo, la utilización de independencia condicional en la teoría de grafos y el reciente desarrollo de algoritmos eficientes para la propagación de evidencia a través de estructuras gráficas han hecho que este campo sea mucho más factible computacionalmente.

Las redes bayesianas han sido exploradas como una herramienta potencial para varias aplicaciones en la gestión de riesgos. Sus principales características que incluyen la combinación de la opinión subjetiva de expertos, de datos observados y los modelos de causa y efecto las

hacen especialmente adecuadas para la investigación y captura del funcionamiento de las instituciones financieras. Aunque su uso hasta ahora ha sido limitado a áreas específicas, su aplicación a los riesgos empresariales más amplios está siendo cada vez más documentada, especialmente en el área de riesgo operacional.

Las redes bayesianas permiten combinar los datos estadísticos que están disponibles con los datos cualitativos de expertos de manera que refleja la estructura causal subyacente del mismo proceso; de este modo se facilita entender y comunicar a los usuarios del negocio los riesgos operativos implicados.

Una red bayesiana es un grafo dirigido acíclico; los nodos representan las variables de interés y las aristas son los enlaces causales o de influencia entre las variables. Asociado con cada nodo está una tabla de probabilidad del nodo, una distribución estadística o una función parametrizada. En el caso de una tabla de probabilidad del nodo, la relación se rige por un conjunto de valores de probabilidad condicionales que modelan la relación incierta entre el nodo y sus padres junto con cualquier incertidumbre presente en esa relación.

La teoría detrás de las redes bayesianas combina la teoría de probabilidad bayesiana y la noción de independencia condicional para representar dependencias entre las variables.

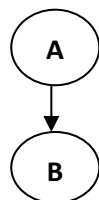
Las redes bayesianas permiten razonar bajo incertidumbre y combinar las ventajas de la representación visual intuitiva con una base matemática de probabilidad bayesiana. Con las redes bayesianas es posible articular dependencias entre diferentes variables y propagar sistemáticamente el impacto de la evidencia sobre las probabilidades de resultados inciertos.

La clave para un diseño exitoso de una red bayesiana es la descomposición significativa de un dominio del problema en un conjunto de proposiciones causales o condicionales sobre el dominio. En lugar de llevar a cabo la laboriosa y difícil tarea de calcular la distribución de probabilidad conjunta de todas las variables de interés, se aplica el principio “divide y vencerás” y tomar las especificaciones parciales del modelo que son en sí significativos en el dominio del experto.

Posteriormente se modela la tabla de probabilidad del nodo para cada variable (nodo), lo cual puede ser hecho usando datos históricos o solamente pidiéndole a un experto que nos proporcione una serie de estimaciones subjetivas, las cuales estarán idealmente basadas en el conocimiento y la experiencia.

Una vez construida la red bayesiana, puede ser ejecutada usando un algoritmo apropiado de propagación. Cuando la red bayesiana se ejecuta, el efecto de los datos ingresados dentro de uno o más nodos son propagados por toda la red, en cualquier dirección y la distribución marginal de los nodos es actualizada. Esto hace al modelo ideal para un análisis de escenarios. En la figura 16.2 se ejemplifica el uso de modelos gráficos.

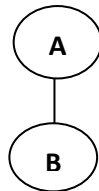
Figura 16.2 Gráfico dirigido de modelos gráficos



A y B se llaman nodos y representan a las variables A y B. El borde dirigido de A hacia B implica una relación de causalidad entre A y B. Indica que un cambio en lo que se sabe de A causa un cambio en lo que se sabe de B. Este cambio es normalmente el resultado de nueva información que llega sobre A. A esta nueva información se le llama evidencia.

Esta relación de causalidad entre las variables le da nombre a los nodos, de esta forma el nodo A se llama “padre” y el nodo B “hijo”.

Figura 16.3. Gráfico no dirigido de modelos gráficos.



Otra forma de conexión se muestra en la Figura 16.3, cuando los bordes no están dirigidos, en este caso no existe relación de causalidad. Esta es una forma más débil de asociación entre el nodo A y el B. A ésta relación se le llama “vecindad”. Si no existe dirección entre A y B, se dice que son sucesos independientes entre sí lo que significa que la ocurrencia de eventos A no tienen relación con la ocurrencia de eventos B y viceversa. Sin embargo esto dependerá de si las variables que intervienen existen entre A y B

En ambos diagramas se representa la probabilidad conjunta de A y B pero de manera diferente. En la Figura 16.2, la relación causal que existe entre A y B significa que la distribución conjunta puede ser expresada como el producto de la probabilidad de A y de la probabilidad de B condicionada por A, es decir la probabilidad expresada en la ecuación 16.5 que no es más que la relación básica de probabilidad en forma grafica. Esta relación no está definida en la Figura 16.3 por lo tanto esta gráfica sólo representa la distribución conjunta $P(A \cap B)$

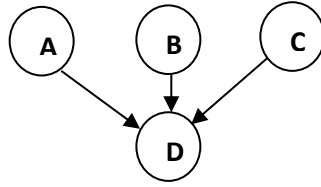
En general, los modelos gráficos forman parte de una red de nodos que conectan variables con algún tipo de relación, ya sea de correlación o de causalidad.

16.3.2.1 Gráfico Dirigido Acíclico (GDA)

Las redes bayesianas contienen relaciones de causalidad y por lo tanto su nodos están conectados por bordes dirigidos. Son parte de un subconjunto de modelos gráficos conocidos como Grafico Dirigido Acíclico (GDA). Los GDA son construidos con relaciones como las de la Figura 16.2 como su bloque básico. Estos bloques están dispuestos de tal manera que son acíclicos, es decir, se mueven a lo largo de los bordes en las direcciones implicadas, es imposible volver a un nodo anterior (Alexander, 2000).

Asociado con cada GDA se tiene asociado un conjunto de probabilidades condicionales que describe el comportamiento del nodo condicionado a sus “padres”.

Figura 16.4. Gráfico dirigido acíclico de cuatro nodos (GDA)



El “padre” del nodo D en la Figura 16.4 es el conjunto de nodos $\{A, B, C\}$.

Una propiedad útil del GDA es que ilustra el supuesto de que la probabilidad conjunta sobre un conjunto de variables $U = \{X_i: i = 1, \dots, n\}$ puede ser expresada como el producto de las densidades condicionales de cada nodo.

$$P(U) = \prod_x P(X_i | pa(X_i)) \quad (16.9)$$

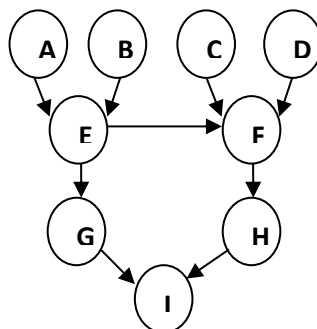
Esta factorización proviene de la propiedad de independencia condicional inherente a la estructura de los GDA. Sin embargo aún esta propiedad puede hacer las manipulaciones de los GDA muy complejas, sobre todo si los nodos representan variables con muchos estados.

16.3.2.1 Arbol de Derivación.

Una factorización que simplifica los cálculos es la estructura llamada “Arboles de Derivación”. Esta estructura realiza cálculos modulares localizados que se ejecutan utilizando un algoritmo de “paso de mensajes”. Estos “Arboles de Derivación” son gráficas no direccionales que consisten en una colección de gráficos conocidos como cliques. Estos cliques son también llamados universos de creencias y son grupos de nodos en los que cada nodo del grupo está conectado con cada otro nodo en el grupo. Cuando un grupo de nodos está conectado de esta manera. Diferentes cliques dentro de un árbol de derivación están conectados a través de separadores. Esto son nodos comunes compartidos por dos cliques adyacentes (Cowell 1999).

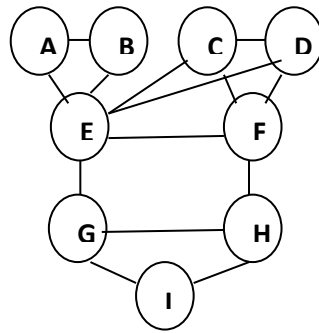
El proceso de transformar un GDA en un árbol de derivación se llama triangulación; aun cuando el nombre hace referencia estrictamente a sólo uno de los pasos. El proceso consta de tres fases:

Figura 16.5. Gráfico Dirigido Acíclico (GDA)



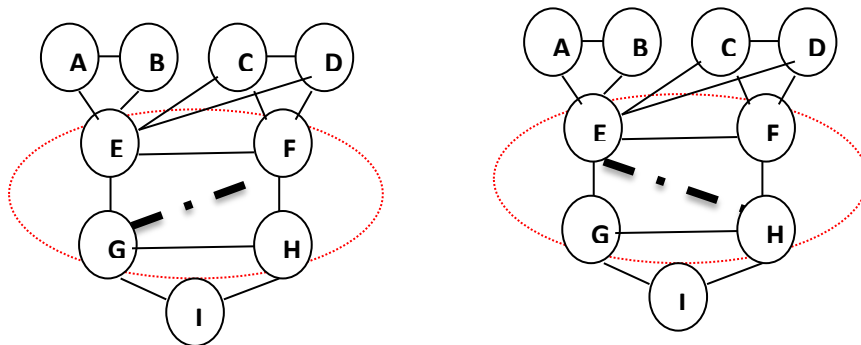
La primera fase se conoce como Moralización. En este paso, todos los nodos padre de un nodo hijo en común que no están conectados, son unidos con un borde no dirigido, una vez hecho esto, a todos los bordes se les quita la dirección (Figura 16.6).

Figura 16.6. Gráfico no dirigido. Moralización



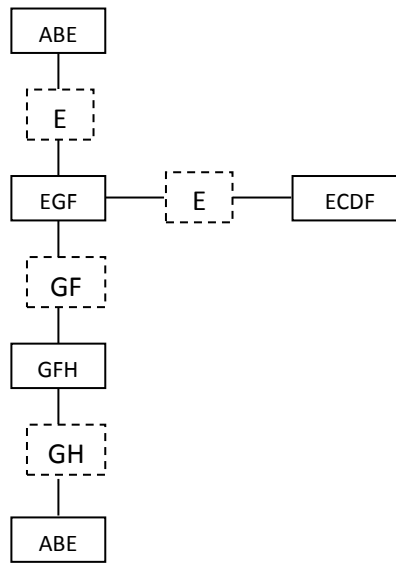
La segunda fase le da nombre al proceso y se le conoce como Triangulación. Un ciclo es una secuencia de nodos conectados por bordes que comienzan y terminan en el mismo nodo. Un ciclo de longitud n consiste de una secuencia de n bordes consecutivos. La triangulación es el proceso de agregar bordes no dirigidos de tal manera que cualquier ciclo que tiene una longitud no mayor a tres. (Figura 16.7)

Figura 16.7. Triangulación



Una vez obtenido el gráfico triangular, un árbol de derivación puede ser especificado. Esto implica la identificación de cliques dentro de la gráfica y los separadores que los conectan (Figura 16.8). Esta es la última fase del proceso y se le llama Especificación del árbol de derivación.

Figura 16.8. Árbol de derivación



16.3.2.1 Inferencia a partir del árbol de Derivación.

Una vez que se tiene el árbol de derivación es posible hacer inferencia a partir de él. Los pasos para la inferencia son tres:

El primer paso consiste en definir las distribuciones a priori. Esta es la distribución no condicional a priori de los nodos, sin “padres” y la distribución condicional a priori para los nodos “hijos”.

Para cada a priori se necesitaran las probabilidades de cada configuración de la combinación de estados de las variables involucradas, es decir, los datos para cada estado, lo cual puede ser determinado por:

- Opinión de un experto.
Los expertos son entrevistados para llegar a conclusiones cuantificables en probabilidades. La confianza en la exactitud de los consejos de los expertos es un requisito para utilizar este método.
- Estimación por máxima verosimilitud.
Para condicionales a priori este método implica tomar la razón de frecuencia del evento con la frecuencia de configuración principal.
Para a priori no condicionales será solo la proporción de ocurrencias entre los varios estados de la variable.
El método de máxima verosimilitud asume que la información pasada es relevante y completa.

En la práctica, no habrá una clara selección del método; los expertos también se basarán en los datos disponibles pero templados con su experiencia y conocimiento respecto a su proyección para futuros eventos.

Una vez definidas las distribuciones a priori se debe inicializar el árbol con ellas. Para cada clique y separador, existe una tabla con celdas que corresponden a cada combinación de estados de las variables en el clique o separador. Los valores numéricos en estas celdas se llamarán potenciales. Estos cambiarán a medida que se pasa la información a lo largo del árbol y que cada clique o separador se actualice.

Al iniciar, todos los potenciales se establecen en la unidad. Las distribuciones a priori factorizadas son multiplicadas por los cliques que contienen las variables en el conjunto de a priori factorizados.

Algunos cliques pueden tener más variables que cualquier conjunto disponible de los a priori factorizados, en estos casos, los a priori se multiplican por los cliques, siempre y cuando las variables en las a priori sean subconjuntos de las variables clique.

Los cliques para los cuales esta operación no es posible se quedan con todos los potenciales como la unidad. Los potenciales en los separadores también permanecen como la unidad. Una vez hecho lo anterior, el árbol está listo para la propagación.

El último paso es la propagación de la información que se da en dos fases. Esta etapa inicia con mensajes originados de las hojas del árbol (los cliques en los extremos del árbol) y se pasan a lo largo del árbol hasta la raíz del clique para luego salir de nuevo por las hojas. En el camino, los potenciales de cada clique se actualizarán.

Los mensajes que pasan entre dos cliques se llevan a cabo de la siguiente manera:

- Marginalización en el clique fuente.
Suma vertical de las celdas del clique de configuraciones comunes de las variables de interés (los que están en el separador).
- Actualizar el nodo del separador.
El nodo separador es actualizado con la suma marginal obtenida por la marginalización del clique fuente. Esto se convierte en el potencial actualizado del separador.
- Actualizar clique
El clique que recibe el mensaje es actualizado mediante la “razón de actualización”. Esta es la razón de los potenciales del separador después de que el mensaje es transmitido entre los potenciales antes de la transmisión del mensaje. La actualización se realiza multiplicando la razón de actualización a las celdas correspondientes en el clique. Estas son las únicas con variables correspondientes a las variables de separación.
Cliques que reciben actualizaciones de más de una fuente, envían un mensaje una vez que este se ha producido. El flujo hacia la raíz clique se llama pase de recolección

Todo el proceso se repite pero con los mensajes saliendo de las hojas del clique raíz. El flujo hacia el exterior es llamado pase de distribución. Por lo tanto el proceso es en ocasiones llamado “Dos fases de propagación”. Una vez que las dos fases han sido completadas, el árbol ha alcanzado el equilibrio. En esta etapa, los potenciales de cada clique necesitan ser normalizados, es decir, escalados de modo que todos los potenciales sumen uno. Posteriormente pueden ser marginalizados para obtener las distribuciones posteriori marginales de cada variable del nodo, lo cual es el resultado buscado.

16.3.2.2 Incorporación de evidencia al árbol de derivación.

Los arboles de derivación pueden incorporar la evidencia de manera muy eficiente y el método de propagación de dos fases puede ser utilizado con la única diferencia de que la evidencia se incorpora en el estado de inicialización. La evidencia implicaría que una variable nodo sea conocida y se asuma uno de los posibles estados, implicando con certeza que la variable no ha tomado ninguno de los otros posibles estados. Esto se incorpora en los potenciales del clique inicial estableciendo todos los potenciales que afectan los otros estados de la misma variable igual a cero. Los otros potenciales permanecen tal como estaban.

Una vez hecho lo anterior, el resto de la propagación continúa. Las probabilidades marginales que se obtienen reflejan el impacto de la evidencia introducida.

Una útil aplicación del árbol de derivación es la generación de muestras aleatorias de configuraciones para las diferentes distribuciones de probabilidad en el GDA. Hay dos métodos de muestreo.

i. Muestreo lógico probabilístico.

Este método se muestrea directamente desde el propio GDA. Las muestras se toman primero de los nodos incondicionales basados en las distribuciones de probabilidad marginal en estos nodos. Después, el nodo hijo de estos nodos son muestreados basándose en las distribuciones de probabilidad condicional de estos nodos, condicionados al resultado de la configuración de la muestra de sus respectivos padres.

Lo anterior continúa hasta que todos los nodos han sido muestreados. Lo cual será contado como una muestra para todo el GDA. El proceso se realiza hasta obtener todas las muestras necesarias. Las frecuencias en cada nodo pueden ser usadas para expresar las probabilidades marginales muestreadas.

ii. Muestreo del árbol de derivación.

Una mejor manera de hacer una muestra considerando la evidencia es utilizar el árbol de derivación. El proceso es similar a la propagación de dos fases. El árbol de derivación es primero inicializado con la evidencia incorporada. Después se lleva a cabo la fase de recolección. Cuando la clique raíz ha sido actualizada, una configuración es muestreada de la raíz clique. Esta muestra es inmediatamente introducida en la raíz clique como evidencia.

La fase de distribución es llevada a cabo pero con la modificación en este paso de muestrear inmediatamente para introducir la evidencia en el clique antes de pasar el mensaje a los cliques adyacentes. El algoritmo continúa hasta que todo el árbol tiene la configuración de la muestra en cada clique.

Este método es mucho más eficiente porque ninguna muestra es rechazada en comparación con el método de la lógica probabilista; la razón es que en el árbol de derivación la evidencia ya está incorporada en el árbol, por lo tanto el proceso de muestreo no dará como resultado muestra con valores que discrepen con la evidencia.

Una vez construida la red bayesiana, su actualización puede darse al incorporar la opinión de los expertos en ella.

16.3.3 Inferencia a partir de la Red Bayesiana

La inferencia bayesiana puede ser expresada por medio de la ecuación (16.8). Si se supone que la variable aleatoria X en una red bayesiana puede tomar los valores x_1, x_2, \dots, x_n con probabilidades $\theta_1, \theta_2, \dots, \theta_n$. La verosimilitud para una secuencia de n muestras independientes, donde n_1 muestras tienen el valor x_1 , n_2 muestras tienen el valor x_2 y sucesivamente pueden ser representadas con la distribución multinomial:

$$p(n_1, n_2, \dots, n_k | \theta) = \frac{n!}{\prod_{i=1}^k n_i!} \prod_{i=1}^k \theta_i^{n_i} \quad (16.10)$$

Un candidato para la θ_i es la distribución Dirichlet con función de densidad:

$$p(\theta) = \frac{\sum_i \alpha_i}{\prod_{i=1}^k \alpha_i} \prod_{i=1}^k \theta_i^{\alpha_i - 1} \quad (16.11)$$

Donde $\sum_i \alpha_i$ es la precisión. La distribución Beta es un caso específico de la distribución Dirichlet de dos parámetros; dado que los nodos binarios son comunes en las redes bayesianas, la distribución beta es usada a menudo.

Los parámetros de la distribución Dirichlet $(\alpha_1, \alpha_2, \dots, \alpha_k)$ se pueden ver como el tamaño implícito de la muestra para θ_i . En la práctica, la opinión de los expertos acerca de θ_i puede ser usada para trabajar retrospectivamente y llegar a la distribución de Dirichlet dado que se conocen los estadísticos:

$$E(\theta_j) = \frac{\alpha_j}{\sum_i \alpha_i} \quad (16.12)$$

$$Var(\theta_j) = \frac{\alpha_j (\sum_i \alpha_i - \alpha_j)}{(\sum_i \alpha_i)^2 (\sum_i \alpha_i + 1)} \quad (16.13)$$

A partir de los cuales se puede obtener α_i .

La distribución posteriori es entonces:

$$p(\theta|n_1, n_2, \dots, n_k) \approx \prod_{i=1}^k \theta_i^{\alpha_i+n_i-1} \tag{16.14}$$

Que es una distribución Dirichlet con nuevos parámetros, $D(\alpha_i + n_1, \alpha_2 + n_2, \dots, \alpha_k + n_k)$

Si se considerará una variable C y su probabilidad condicional $P(C|A, B)$ donde C es una variable binaria, la distribución a priori sería $B(\alpha_1, \alpha_2)$. La estimación de los parámetros se realiza primero recopilando la opinión de los expertos; respecto a su mejor estimación de la media y un rango de los valores probables. El mejor estimador se toma como la media y el rango para abarcar dos desviaciones estándar sobre la media de donde se obtiene la varianza. En la Figura 16.9 se muestra la forma de almacenar los datos.

Figura 16.9. Almacenamiento de los resultados

Padre	C_s		C_n	
	$E(\theta_s)$	$Var(\theta_s)$	$E(\theta_n)$	$Var(\theta_n)$
A_s, B_s				
A_n, B_s				
A_s, B_n				
A_n, B_n				

Para cada configuración padre se tienen un par de ecuaciones simultaneas³, ecuaciones (16.15) y (16.16) que pueden ser resueltas para α_s y α_n .

Para la esperanza

$C_s:$	$E(\theta_s) = \frac{\alpha_s}{\alpha_s + \alpha_n}$	(16.15)
$C_n:$	$E(\theta_n) = \frac{\alpha_n}{\alpha_s + \alpha_n}$	

Para la varianza

$Var(\theta_s) = \frac{\alpha_s \alpha_n}{(\alpha_s + \alpha_n)^2 (\alpha_s + \alpha_n + 1)}$	(16.16)
$Var(\theta_n) = \frac{\alpha_s \alpha_n}{(\alpha_s + \alpha_n)^2 (\alpha_s + \alpha_n + 1)}$	

³ Caso de dos parámetros

Habrán dos conjuntos de parámetros (α_1, α_2) a elegir del cuál se sugiere tomar el que da la menor precisión y por lo tanto la mayor varianza por ser el más conservador.

Para llegar a la distribución posterior de Dirichlet se requiere añadir los datos al parámetro a priori para obtener el parámetro posteriori. Suponiendo la distribución a priori para $P(C|A_S, B_S)$ es $B(\alpha_s, \alpha_n)$ y se obtendrá el siguiente conjunto de datos relevantes:

$$\{(A_S, B_S, C_S, \dots), (A_S, B_S, C_n, \dots), (A_S, B_S, C_n, \dots), (A_S, B_S, C_n, \dots), (A_S, B_S, C_S, \dots)\}$$

Entonces la función de distribución posteriori es: $B(\alpha_s + 2, \alpha_n + 3)$.

Las estimaciones de probabilidad actualizadas se pueden obtener tomando la media de la función de distribución posteriori. Para $P(C_S|A_S, B_S)$ la posteriori es $\alpha_s + 2 / \alpha_s + \alpha_n + 5$ y para

$$P(C_n|A_S, B_S) \text{ es } \alpha_n + 3 / \alpha_s + \alpha_n + 5$$

El proceso de actualización bayesiana se llama “aprendizaje”. Cuando las redes bayesiana se realizan sin el aprendizaje, las probabilidades permanecen sin cambio con los datos.

16.4 Aplicación de redes bayesianas para medir el riesgo operativo en una empresa financiera.

La oportunidad de incorporar la opinión de expertos en el proceso de seleccionar las variables de interés, definir la estructura del modelo por medio de sus dependencias de causalidad y la especificación de las distribuciones a priori y por consiguiente las probabilidades condicionales de cada nodo es una aplicación de un modelo de Redes Bayesianas a la medición del Riesgo Operacional de una empresa.

La representación gráfica de las redes bayesianas facilita el entendimiento de la estructura causal y el perfil de riesgo de la empresa se presenta de forma intuitiva (Neil, Fenton y Marquez, 2004).

La teoría definida en las secciones anteriores se aplica a medir el Riesgo Operacional de una empresa cuyo flujo de información se realiza totalmente en línea. Es una empresa privada que opera desde 2006 con una concesión del Gobierno Federal (de México) para administrar una Base de Datos Nacional.

La consulta de la información por parte de los interesados se realiza en línea por medio de un portal en Internet; razón por la cual se tiene un gran capital invertido en software y hardware que soporten la operación. El volumen de las transacciones realizadas es sumamente alto por lo cual el consejo directivo requiere contar con los recursos financieros y humanos necesarios para hacer frente a una contingencia.

El objetivo del análisis es:

Identificar las causas de eventos de Riesgo Operativo que ocasionan inactividad en las transacciones y/o pérdidas a la empresa.

- Apoyar a la administración de la empresa a decidir cuánto capital destinar para cubrir las contingencias por eventos adversos.

Se asume que la base de datos utilizada para el análisis está completa; en caso contrario algunos métodos para trabajar con bases de datos incompletas son sugeridos por Cowell (1999) y funciones de distribuciones de probabilidad discretas; modelos con distribuciones continuas conocidas como distribuciones gaussianas pueden ser empleadas como tema de una investigación posterior.

16.4.1 Construcción de la red bayesiana

Las principales causas de riesgo operativo por consultas realizadas en línea son:

- Tiempo de inactividad en las transacciones y del servidor así como pérdida de datos.
- Severidad que representa el tiempo de inactividad en las transacciones y del servidor serán catalogadas en función del tiempo que se tarde en “levantar” (reestablecer la operación) la red o el servidor, siendo este de un día completo, medio día o sin inactividad. En el caso de la pérdida de datos su severidad será considerada como la pérdida del 100%, el 50% o el 0% de los datos.

Durante el tiempo que la empresa lleva realizando las transacciones en línea, se ha estimado el costo total de eventos adversos considerando el costo de reparaciones, salarios pagados al personal inactivo, costo de recuperación de la información y daño en la reputación de la empresa.

Las fallas en la red tienen como consecuencia inmediata tiempo de inactividad de las transacciones que puede ser mitigado con una red de alta velocidad para operar. Fallas en el servidor originan que este quede inactivo lo cual puede deberse a problemas de energía eléctrica y la calidad del servidor, este riesgo puede ser aminorado con una planta de energía eléctrica y el uso de un servidor de alta velocidad.

La pérdida de datos es consecuencia de fallas en el servidor o fallas en el software. Las fallas en el software son originadas por modificaciones, intencionales o no, hechas por los empleados de la empresa, virus en el sistema que pueden ser controladas implementando restricciones de acceso al software. Un riesgo más que amenaza la pérdida de datos es que la empresa sea víctima de piratas informáticos para lo cual será necesario contar con software que prevenga esta exposición. Un ataque de piratas informáticos puede también producir fallas de la red.

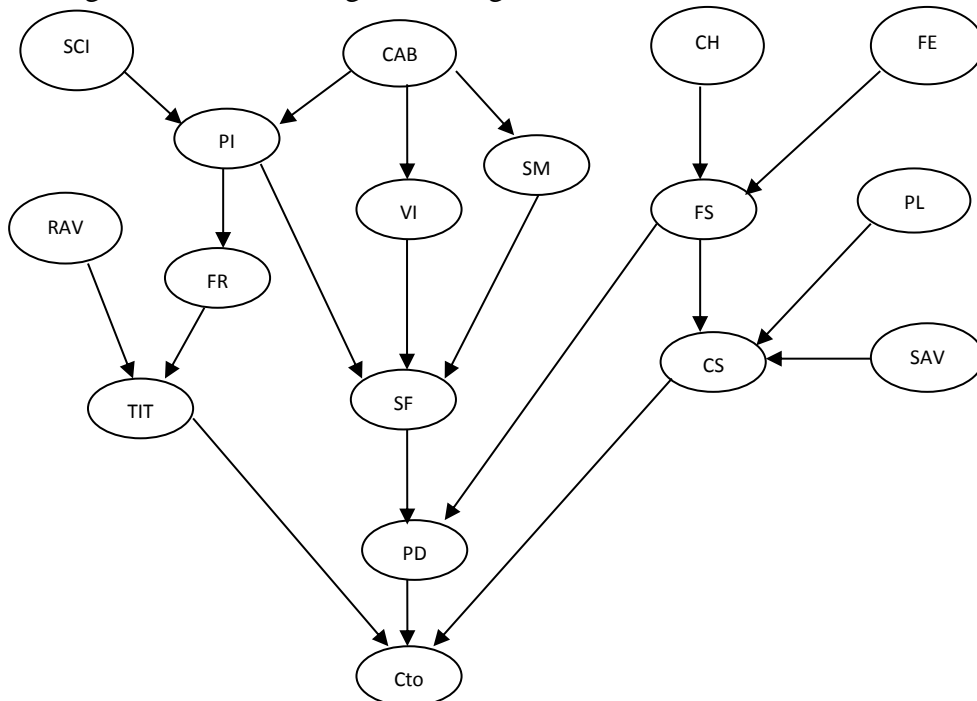
Consultando con expertos en Tecnologías de la Información de la empresa y considerando las causas de riesgo y su origen se definieron las siguientes variables para el diseño del modelo que se muestran en la figura 16.10

Figura 16.10. Definición de las variables aleatorias

	Variable	Abreviatura	Valores
1.	Software contra intrusos	SCI	Aplicación o filtro
2.	Control de acceso a la base de datos	CABD	Alto y bajo control
3.	Calidad del hardware	CH	Alta y Baja
4.	Fuente de energía	FE	Sí y No
5.	Planta de luz	PL	Sí y No
6.	Servidor de alta velocidad	SAV	Sí y No
7.	Red de alta velocidad	RAV	Sí y No
8.	Piratas Informáticos	PI	Sí y No
9.	Virus informático	VI	Sí y No
10.	Modificación del software por los usuarios	SM	Sí y No
11.	Fallas de la red	FR	Sí y No
12.	Fallas del software	SF	Total, Parcial y Sin Falla
13.	Fallas del servidor	FS	Sí y No
14.	Tiempo de inactividad en las transacciones	TIT	Cero, medio día o todo el día
15.	Pérdida de la base de datos	PD	0%, 50% y 100%
16.	Servidor caído	CS	Cero, medio día o todo el día
17.	Costo total	Cto	Millones de pesos

Con base la definición de variables aleatorias anterior, se establecen las relaciones de causalidad entre las variables. La figura 16.11 muestra el gráfico dirigido acíclico de la empresa.

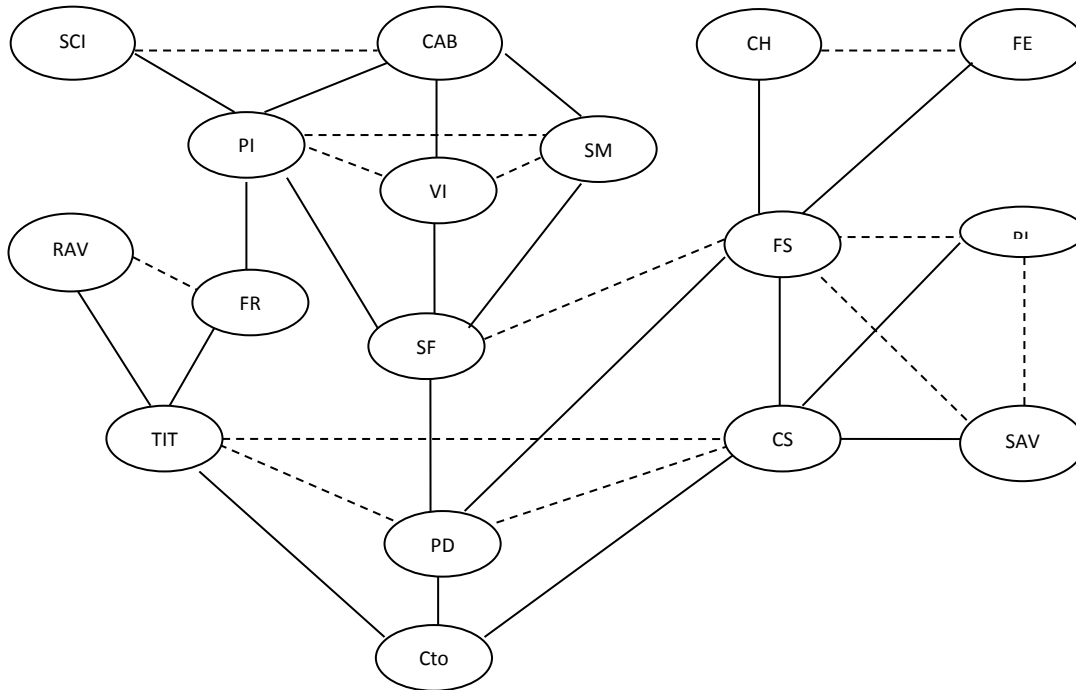
Figura 16.11. Modelo gráfico dirigido acíclico del caso de análisis.



El siguiente paso es obtener el árbol de derivación tomando la teoría desarrollada en la sección anterior aplicando para ello las tres fases necesarias.

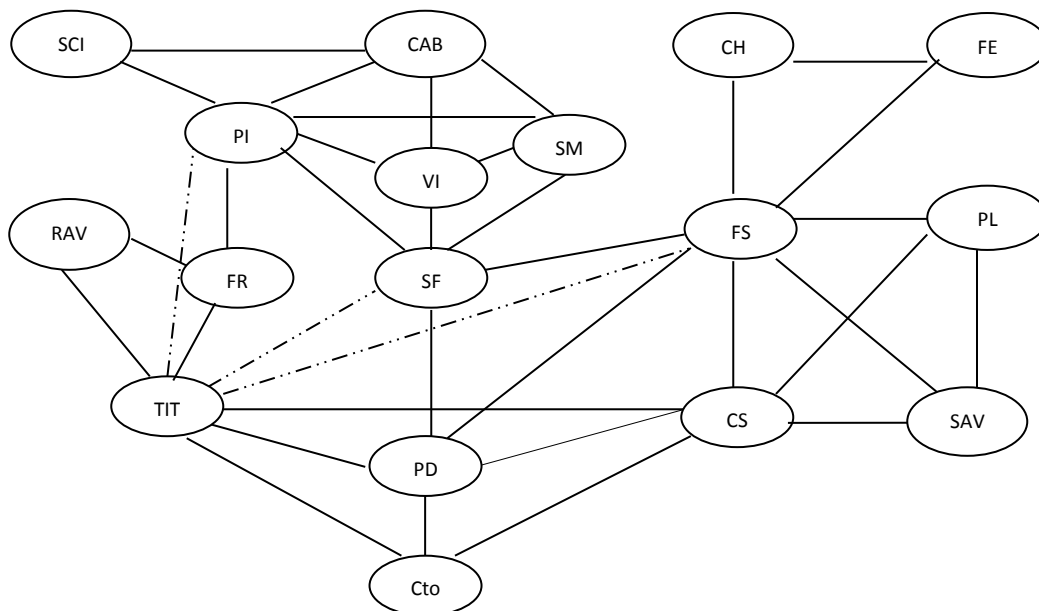
Moralización. Todos los nodos padre con un nodo hijo en común que no están conectados son unidos con un borde no dirigido y a todos los bordes se les quita la dirección (Figura 10).

Figura 16.12. Moralización



Triangulación. Agregar bordes no dirigidos de manera tal que cualquier ciclo quede de longitud tres. Para el caso en análisis la figura 16.13 muestra el proceso de triangulación.

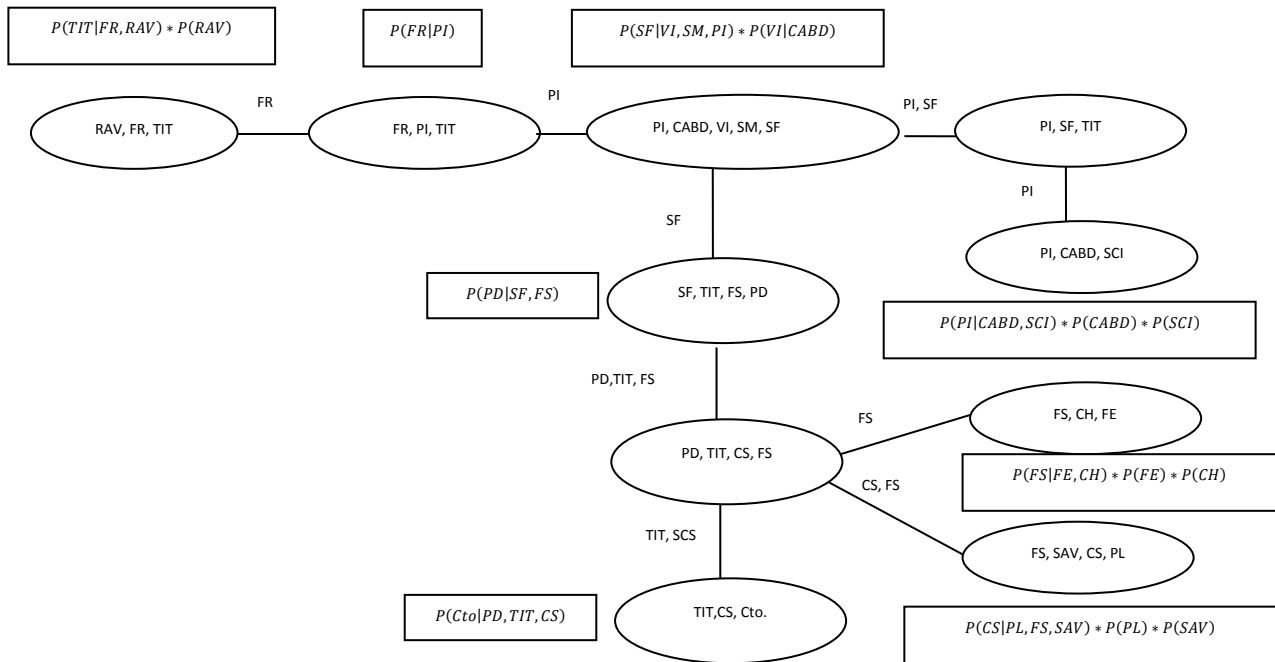
Figura 16.13. Triangulación



Especificación del árbol de derivación. Una vez realizado el proceso de triangulación se puede especificar el árbol de derivación; lo anterior implica la identificación de los cliques dentro de la gráfica y los separadores que los conecta.

Un árbol de derivación no es único y para seleccionarlo es útil evitar cliques grandes. El seleccionado para el presente trabajo se muestra en la Figura 16.14

Figura 16.14. Especificación del árbol de derivación.



Una vez definido el árbol de derivación con el que se va a trabajar, es factible hacer inferencia a partir de él acerca de las distribuciones a posteriori.

La fase de inicialización del árbol de derivación implica ajustar todos los cliques potenciales en 1, las combinaciones de a priori que factorizan como producto de las densidades condicionales de cada nodo⁴ se multiplican en los cliques que contienen el conjunto correspondiente de variables. La fase de recolección implica normalizar los cliques potenciales, la suma de las marginales correspondientes a diferentes combinaciones de los estados de las variables separadores se utilizan para actualizar los separadores.

La empresa nos proporcionó las probabilidades marginales apriori (Anexo I); por flexibilidad en el modelo, a algunas de las probabilidades condicionales a priori se les asignó distribución uniforme.

⁴ $\prod_x P(X|pa(X))$ Factorización que proviene de la propiedad de independencia condicional inherente a la estructura de los gráficos dirigidos acíclicos (GDA)

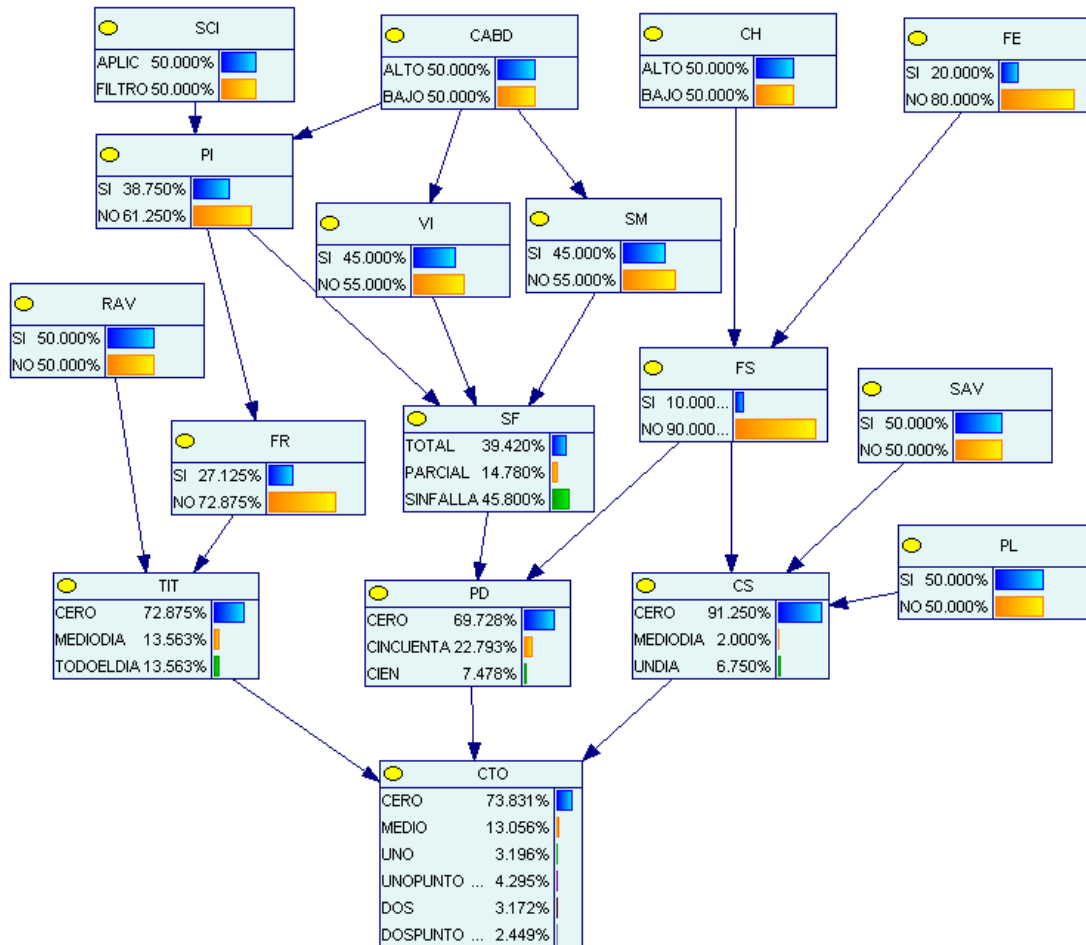
Teóricamente, cualquier nodo puede ser tratado como una variable de entrada, existen nodos donde el control es interno y otros como el nodo “Fuente de Energía” donde evidentemente el control esta fuera de la empresa y donde será necesario utilizar evidencia histórica para obtener las probabilidades; este nodo no será tratado como variable de entrada.

Al alimentar la Red Bayesiana con información actualizada (evidencia) en nodos donde la opinión de los expertos fue utilizada para el cálculo de las probabilidades apriori, se hablará de análisis de escenarios o pruebas de estrés en variables de entrada.

Existe software disponible en el mercado⁵ ya sea en versión gratuita o versiones comerciales para realizar las fases, para el presente trabajo se utilizó una versión libre del software GeNIe.

Si no se introduce evidencia en la Red Bayesiana, el árbol inicia con las probabilidades marginales a priori para todos los nodos. El árbol inicializado se muestra en la Figura 16.15 con las distribuciones marginales a priori.

Figura 16.15. Inicialización del árbol



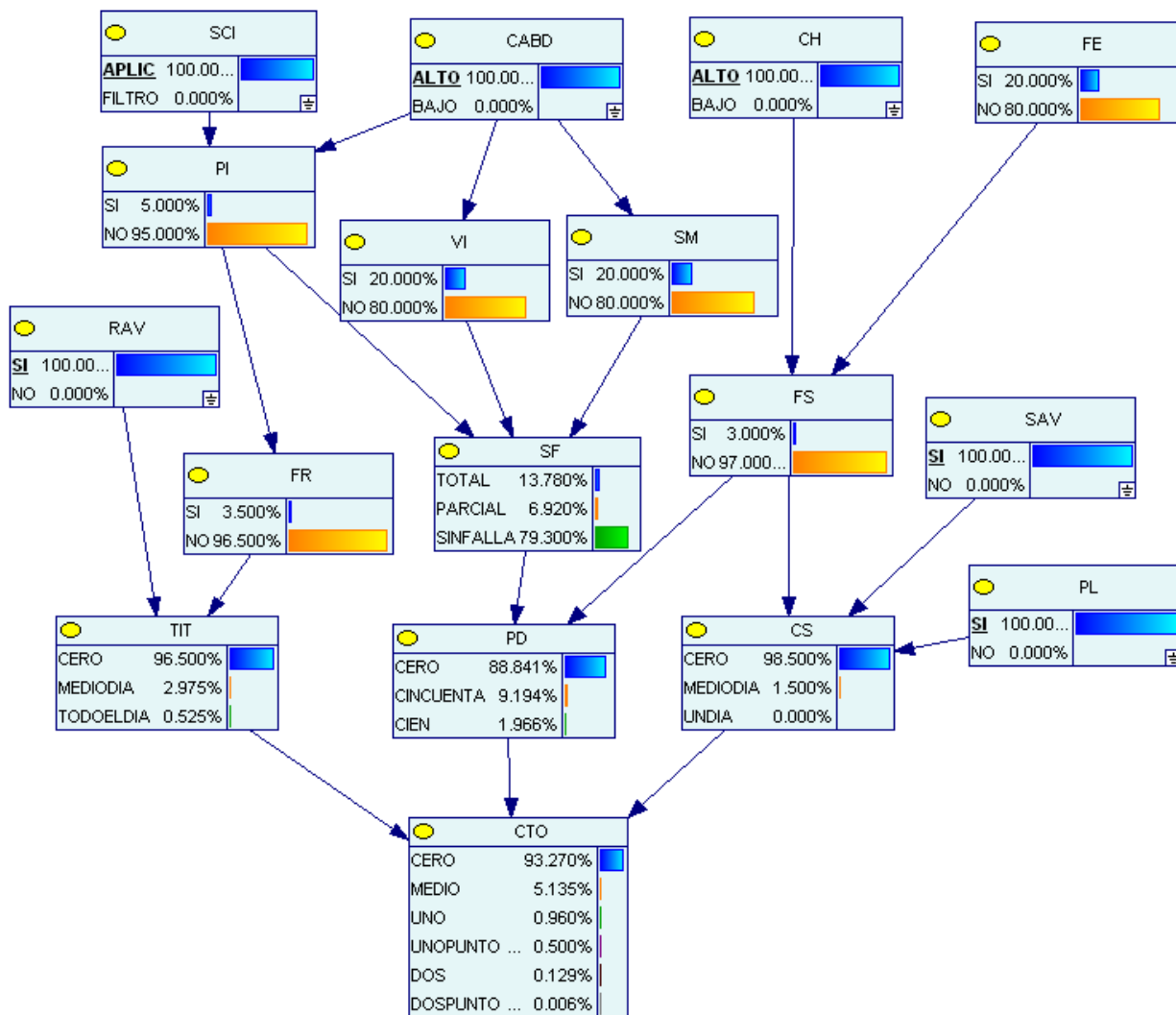
⁵ GeNIe <http://genie.sis.pitt.edu/downloads.html>

16.4.2 Inferencia a partir de la Red Bayesiana

La Red Bayesiana está lista para modelar el riesgo operacional; en particular para decidir el monto de capital que debe destinar para estar cubierta en un evento extremo de fallas de la red.

Al conocer el estado actual de ciertas variables, el modelo puede alimentarse de estos datos para conocer el estado real de la empresa por medio del Muestreo del árbol de derivación. Si se sabe que la empresa ha implementado un software contra intrusos, existe un alto control sobre la base de datos, se cuenta con una red y un servidor de alta velocidad, la calidad del hardware con el que se opera es alta y se cuenta con una planta de energía en el caso de suspensión de energía eléctrica; el modelo se alimenta en la fase de inicialización con la evidencia proporcionada y se reinicia el árbol para obtener las nuevas distribuciones marginales posteriori. El resultado se observa en la Figura 16.16

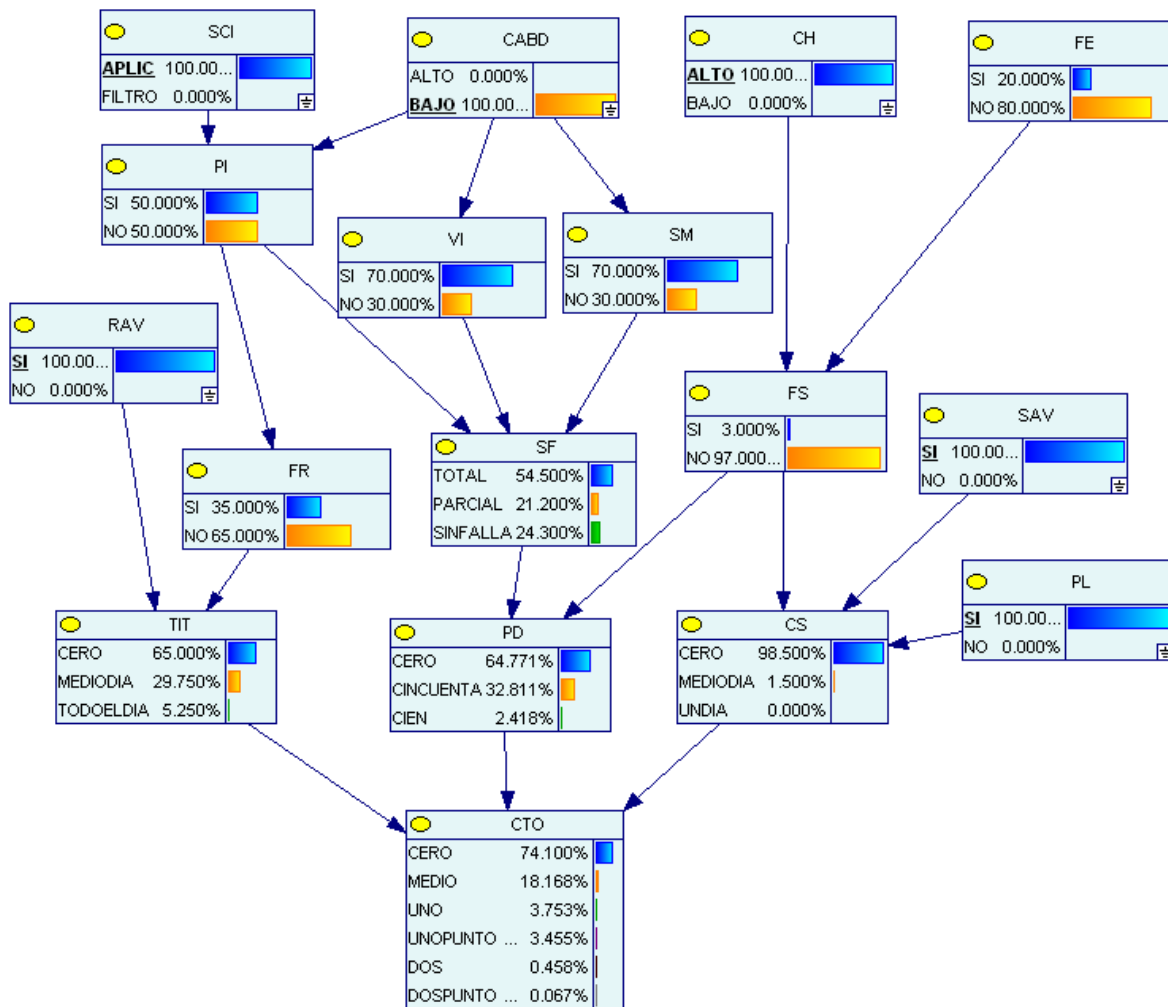
Figura 16.16. Distribuciones marginales posteriori basadas en la evidencia de la empresa.



Para determinar el capital de riesgo que debe designarse a cubrir una falla en la red utilizando la técnica de árboles bayesianos se establece un nivel de confianza y un intervalo de tiempo. En un periodo de tiempo de una semana con una confianza del 95% el costo es de 0.17 millones de pesos y con un nivel de confianza del 99% el costo es de 0.63 millones de pesos⁶.

Una aplicación del modelo es apoyar a la administración a optimizar su perfil de riesgo. Si desea reducir costos de operación disminuyendo el control de acceso a la base de datos (CABD); la decisión se incorpora al árbol y se propaga para obtener las densidades marginales que se observan en la Figura 16.17.

Figura 16.17. Distribuciones marginales posteriores con pruebas de escenario.



La estrategia en reducción de costos repercute directamente en el capital de riesgo que debe asignarse a cubrir una falla en la red. Este deberá ser ahora de 0.86 millones de pesos para una confianza del 95% y de 1.43 millones para una confianza del 99%. El reducir costos por medio del control de acceso a la base de datos implica considerar un incremento de capital de

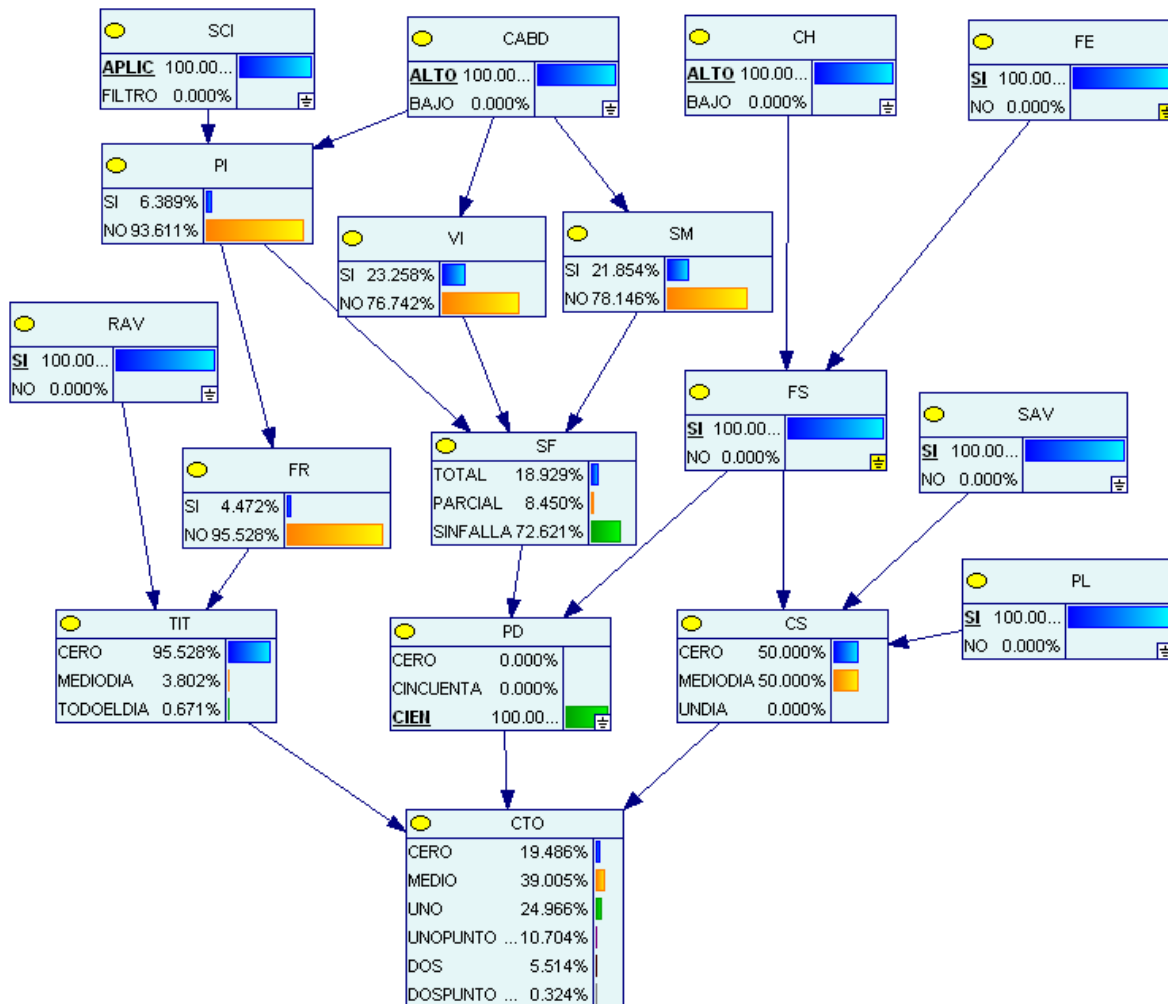
⁶ Se calcula mediante el uso de interpolación lineal.

riesgo operacional de 0.69 millones con una confianza del 95% y de 0.8 millones con una confianza del 99%.

Si un evento adverso se presenta en la institución, el modelo de Redes Bayesianas permite investigar el impacto del mismo por medio de pruebas de estrés⁷.

Una pérdida del 100% de la base de datos hará incurrir a la compañía en un costo de 1.58 millones de pesos con un nivel de confianza del 95% y en 1.94 millones de pesos con un nivel de confianza del 99% tal como se observa en la Figura 16.18

Figura 16.18 Distribuciones marginales posteriori modelando causa-efecto. Perder 100% de la base de datos.



Al analizar la pérdida total de la base de datos, la probabilidad de piratería informática, virus y modificaciones del software por los usuarios disminuye considerablemente; esto conlleva a una disminución de la probabilidad de que el software falle; con lo que la fuente principal de pérdida de información sería una falla del servidor causada por falla en la energía eléctrica y baja calidad del hardware.

⁷ Análisis de causas.

Cada semana se tendrá nueva información de los variables incluidas en el árbol; una vez que el modelo ha sido implementado, esta información puede y debe ser incorporada a la Red Bayesiana para mantener al modelo dinámico.

16.4.3 Actualización de la Red Bayesiana

Los datos observados y la opinión de los expertos se introducen al modelo de Redes Bayesianas definido. La opinión de los expertos es cuantificada de forma matemática por medio de la distribución de Dirichlet, distribución que permite actualizar la información usando inferencia bayesiana.

La media y la desviación estándar de las probabilidades de los valores en la variable se obtiene de la opinión de los expertos. Para expertos no familiarizados con conceptos de estadística se les solicita dar un valor estimado y un rango de los valores. La mejor estimación es tomada como la media y el rango de valores como dos veces la desviación estándar con respecto a la media, de lo cual se estima la varianza.

Se analiza el caso de fallas en la red (FR) sabiendo la ocurrencia de piratería informática (PI) ($P(FR|PI_{si})$). Los expertos en Tecnologías de la Información estiman que la media de fallas en la red dado la ocurrencia de piratería informática es de 0.7 en un rango de 0.65 y 0.85; la media de que no ocurra una falla de red aún habiendo sufrido un ataque de piratería informática es de 0.3 con un rango de valores de 0.2 y 0.5.

Conociendo los valores de media y desviación estándar es posible obtener los parámetros de la distribución Beta⁸ mostrados en la Figura 16.19 por medio de la resolución de las ecuaciones simultaneas (16.15) y (16.16):

Figura 16.19. Valores de media y rango de $P(FR|PI_{si})$.

	Media	Rango		Varianza	α_{si}	α_{no}
PI=si FR=si	0.7	0.65	0.85	0.01	14	3.5
PI=si FR=no	0.3	0.2	0.5	0.0225	5.8333	1.4583

Se toman como válidos los parámetros $\alpha_{si} = 5.83$ y $\alpha_{no} = 1.46$ por ser los de menor precisión.

Todo lo que se requiere para llegar a la distribución posteriori de Dirichlet es añadir los datos al parámetro a priori para obtener el parámetro posteriori.

⁸Es un caso específico de dos parámetros de la distribución Dirichlet y comúnmente usada en nodos de variables binarias (SI/NO).

Se sabe que en el transcurso de un año, el sistema fue víctima de ocho ataques de piratería informática de los cuales tres provocaron fallas en la red, los parámetros de la distribución Beta se actualizan con las últimas probabilidades estimadas para obtener la media posteriori de la distribución Beta.

La función de distribución posteriori es: $B(\alpha_s + 3, \alpha_n + 5) = B(8.83, 6.46)$.

Las probabilidades estimadas con los datos actualizados se obtienen de la media de la distribución Beta posteriori $B(8.83, 6.46)$ y se observan en la Figura 16.20

$$P(FR_{si}|PI_{si}) = \frac{\alpha_{si} + 3}{\alpha_{si} + \alpha_{no} + 8} = \frac{8.83}{5.83 + 1.46 + 8} = 0.58$$

$$P(FR_{no}|PI_{si}) = \frac{\alpha_{no} + 5}{\alpha_{si} + \alpha_{no} + 8} = \frac{6.46}{5.83 + 1.46 + 8} = 0.42$$

Figura 16.20 Valores de media con datos actualizados.

	Media apriori	Eventos	α_{si}	α_{no}	$\alpha_{si} + \text{eventos}$	$\alpha_{no} + \text{eventos}$	Media Posteriori
PI=si FR=si	0.7	3	5.83		8.83		0.58
PI=si FR=no	0.3	5		1.46		6.46	0.42

El estimador a priori de fallas en la red causadas por piratería informática se redujo de 70% a 58% con la actualización de la información al incorporar el hecho de que 3 de 8 eventos ataques de piratería informática derivaron en fallas de la red.

El proceso de actualización de la red se conoce como “aprendizaje”. Cuando las redes bayesiana se realizan sin aprendizaje, las probabilidades permanecen sin cambio.

16.5 Conclusiones

El uso de Redes Bayesianas para modelar el Riesgo Operacional incorpora la opinión de expertos mediante la:

1. Selección de las variables de interés.
2. Definición de la estructura del modelo por medio de las relaciones de dependencia.
3. Especificación de las distribuciones a priori y las probabilidades condicionales de cada nodo.

La actualización de la red con la nueva información disponible permite garantizar que ésta no es estática y el modelo no pierde valides con el tiempo. La red se adapta rápidamente a la nueva información, incorporando la opinión de los expertos de forma matemática.

Las pruebas de estrés que pueden ser fácilmente desarrolladas en las redes bayesianas es una característica para un sistema de alertas tempranas en regímenes regulatorios de instituciones

financieras por lo que existe un tremendo potencial para el ajuste del capital de riesgo operacional con el uso de modelos internos.

El modelo puede volverse complejo si existen muchos nodos por especificar, especialmente si los nodos provienen de muchos padres. En estos casos, existirán varias probabilidades condicionales que definir que requerirán de un gran volumen de información si se usa el método de máxima verosimilitud para especificarlas, lo cual reduce una de las principales ventajas del uso de las redes bayesianas. Uno de los mayores retos es lidiar con la opinión de los expertos que por lo general no están familiarizados con términos estadísticos.

Existe también la problemática de no tener una estructura de relación causal única del modelo, por lo que escoger el modelo que mejor ajuste se vuelve subjetivo y por consiguiente es sujeto a debate. (Alexander, 2000)

Las instituciones financieras están a la búsqueda continua de modelos cuantitativos cada vez más sofisticados que sirvan para satisfacer los requerimientos de las autoridades de regulatorias en materia de requerimientos de capital y que al mismo tiempo demuestren una sólida administración del riesgo.

16.6 Bibliografía

Basel Committee on Banking Supervision (BCBS), 2006. International Convergence of Capital Measurement and Capital Standards, Bank for International Settlements.

Alexander, C. (2002) "Operational Risk Measurement: Advanced Approaches" ISMA Centre, University of Reading, UK

Circular Unica de Bancos (CUB),2005. Secretaría de Hacienda y Crédito Público.

Chernobai A., Jorion P. y Yu F (2011) "The Determinants of Operational Risk in U.S. Financial Institutions". *Jornal of Financial and Quantitative Analysis*, Vol. 46 No. 6 pp. 1683-1725

Chonawee S. Chris K. And Lucas H.,2006. "Cause to Effect Operational Risk Quantification and Management", Risk Management.

Cowell R.G., A.P. Dawid, S.L. LAuritzen y D.J. Spiegelhalter, 1999, "Probabilistic Networks and expert Systems", Sinpinger-Verlag

Hoffman, D., 2002, "Managing Operational Risk: 20 Firmwide Best Practice Strategies". Wiley
Hull J. C. "Risk Management and Financial Institutions" Second Edition, 2007

Jorion, P., 2001, "Value at Risk: The New Benchmark for Managin Financial Risk", McGraw Hill.

Madsen A. and Kjaerulff U, 2008. "Bayesian Networks and Influence Diagrams" Springer

Miccolis, J.A., and S. Shah, 2000, "Getting a Handle on Operational Risk", Tillinghast-Towers Perrin

Neil, M.N. Fenton and M. Tailor, 2005. "Using Bayesian Networks to model expected and unexpected operational losses" Risk Analysis Journal

Neil, M.N. Fenton and D. Marquez, 2004 "Using Bayesian Networks to model the operational risk to information technology infrastructure in financial institutions" Journal of financial transformation.

Power M., 2005 "The Invention of Operational Risk", Review of International Political Economy

ANEXO I

Incondicionales

Control de acceso a los archivos

CABD	
Alto	Bajo
0.5	0.5

Red de alta velocidad

RAV	
Si	No
0.5	0.5

Servidor de alta velocidad

SAV	
Si	No
0.5	0.5

Software contra intrusos

SCI	
Aplicación	Filtro
0.5	0.5

Fuente de energia

FE	
Si	No
0.2	0.8

Calidad Hardware

CH	
Alto	Bajo
0.5	0.5

Siministro ininterrumpido de energia

PL	
Si	No
0.5	0.5

Condicionales

FS-FE,CH

CH	FE		
	SI	NO	0.85
ALTO	SI	0.15	0.85
	NO	0	1
BAJO	SI	0.85	0.15
	NO	0	1

FS

PI-SCI,CABD

SCI	CABD		PI	
	Alto	NO	SI	NO
AP	Alto	0.05	0.95	
	Bajo	0.5	0.5	
FILTRO	Alto	0.2	0.8	
	Bajo	0.8	0.2	

TIT-FR,RAV

RAV	TIT			
	FR	NO FALLA	MEDIO DIA	1 DIA
SI	SI	0	0.85	0.15
	NO	1	0	0
NO	SI	0	0.15	0.85
	NO	1	0	0

SF-PI,VI,SM

PI	SM	SF			
		VI	AP	FILTRO	OK
Y	Y	Y	0.9	0.1	0
		N	0.6	0.4	0
	N	Y	0.7	0.3	0
		N	0.6	0.2	0.2
N	Y	Y	0.4	0.3	0.3
		N	0.1	0.3	0.6
	N	Y	0.5	0	0.5
		N	0	0	1

CS-SAV,FS,PL

SAV	FS	CS			
		PL	NO FALLA	MEDIO DIA	1 DIA
Y	Y	Y	0.5	0.5	0
		N	0	0.2	0.8
	N	Y	1	0	0
		N	1	0	0
N	Y	Y	0	0.1	0.9
		N	0	0	1
	N	Y	1	0	0
		N	1	0	0

VI-CABD

CABD	VI	
	Si	No
Alto	0.2	0.8
Bajo	0.7	0.3

SM-CABD

CABD	SM	
	Si	No
Alto	0.2	0.8
Bajo	0.7	0.3

FR-PI

PI	FR	
	Si	No
SI	0.7	0.3
NO	0	1

PD-FS,SF

FS	SF	PD		
		0%	50%	100%
Y	Total	1	0	0.9
	Parcial	0.1	0.1	0.8
	Sin Falla	0.2	0.2	0.6
N	Total	0.5	0.5	0
	Parcial	0.7	0.3	0
	Sin Falla	1	0	0

Cto-TIT,PD,CS

PD	TIT	Cto.							
		SC	0	0.5	1	1.5	2	2.5	
0%	No SFlla	No SFlla	1	0	0	0	0	0	
		Medio día	0	0.6	0.4	0	0	0	
		Un Día	0	0.1	0.2	0.3	0.2	0.2	
	Medio Día	No SFlla	0.7	0.3	0	0	0	0	
		Medio día	0	0.5	0.3	0.2	0	0	
		Un Día	0	0	0.2	0.4	0.2	0.2	
	Un Día	No SFlla	0.6	0.4	0	0	0	0	
		Medio día	0	0.4	0.3	0.2	0.1	0	
		Un Día	0	0	0.1	0.2	0.4	0.3	
5.00%	No SFlla	No SFlla	0.6	0.4	0	0	0	0	
		Medio día	0	0.5	0.4	0.1	0	0	
		Un Día	0	0	0.1	0.3	0.4	0.2	
	Medio Día	No SFlla	0.3	0.3	0.2	0.2	0	0	
		Medio día	0	0.3	0.3	0.2	0.2	0	
		Un Día	0	0	0.1	0.2	0.5	0.2	
	Un Día	No SFlla	0.3	0.2	0.2	0.2	0.1	0	
		Medio día	0	0.3	0.2	0.2	0.2	0.1	
		Un Día	0	0	0	0.1	0.4	0.5	
100% 362	No SFlla	No SFlla	0.4	0.4	0.2	0	0	0	
		Medio día	0	0.4	0.3		0.1	0	
		Un Día	0	0	0	0.3	0.4	0.3	
	Medio Día	No SFlla	0.2	0.2	0.3	0.2	0.1	0	
		Medio día	0	0.2	0.2	0.3	0.2	0.1	
		Un Día	0	0	0	0.2	0.4	0.4	
	Un Día	No SFlla	0	0.1	0.2	0.3	0.2	0.2	
		Medio día	0	0	0.2	0.3	0.3	0.2	
		Un Día	0	0	0	0.2	0.3	0.8	