

ANÁLISIS Y COMPARACIÓN DE ALTERNATIVAS PARA CUANTIFICAR EL RIESGO OPERACIONAL

LUIS CEFERINO FRANCO ARBELÁEZ



**ESCUELA DE CIENCIAS Y HUMANIDADES
MAESTRÍA EN MATEMÁTICAS APLICADAS
2009**

**ANÁLISIS Y COMPARACIÓN DE ALTERNATIVAS PARA CUANTIFICAR EL
RIESGO OPERACIONAL**

LUIS CEFERINO FRANCO ARBELÁEZ

**TRABAJO DE GRADO PRESENTADO COMO REQUISITO PARCIAL PARA
OPTAR AL TÍTULO DE MAESTRÍA EN MATEMÁTICAS APLICADAS**

DIRECTOR

ERMILSON VELÁSQUEZ CEBALLOS

Doctor en Ciencias Matemáticas



**ESCUELA DE CIENCIAS Y HUMANIDADES
DEPARTAMENTO DE CIENCIAS BÁSICAS
MAESTRÍA EN MATEMÁTICAS APLICADAS
2009**

Nota de aceptación

Coordinador de la Maestría

Director del proyecto

Ciudad y fecha (día, mes, año):

AGRADECIMIENTOS

Agradezco a la Universidad EAFIT por el apoyo financiero para el desarrollo de este trabajo, mediante proyecto de investigación con código: 103-000023 de 2009.

Una gratitud muy especial a DIOS y a mi familia.

CONTENIDO

INTRODUCCIÓN	6
1. PRELIMINARES	10
2. MÉTODO DE DISTRIBUCIÓN DE PÉRDIDAS (LDA)	15
2.1 DISTRIBUCIÓN DE FRECUENCIAS	20
2.2 DISTRIBUCIÓN DE SEVERIDAD	22
2.3 CUANTIFICACIÓN DE LAS PÉRDIDAS AGREGADAS	24
2.3.1 Simulación Montecarlo (SMC)	25
2.3.1.1 SMC. Algoritmo 1.	25
2.3.1.2 SMC. Algoritmo 2.	26
2.3.2 Algoritmo de recursión de Panjer	27
2.3.2.1 Algoritmo de Panjer	30
2.3.2.1.1 Lema	30
2.3.2.1.2 Teorema de Panjer	31
2.3.2.2 Comentarios sobre el Algoritmo de Panjer	32
2.3.2.3 Implementación de Algoritmos de Panjer	36
2.3.3 Aproximación en forma cerrada de Böcker y Klüppelberg.	36
2.3.3.1 Distribuciones de severidad subexponenciales	37
2.3.3.2 Lema de Kesten	39
2.3.3.3 Teorema EKM. (Embrechts, Klüppelberg and Mikosch)	39
2.3.3.4 Teorema de la Fórmula analítica para OpVaR	41
3. APLICACIÓN	44
3.1 ANÁLISIS DE LOS DATOS	44
3.2 APLICACIÓN DE SIMULACIÓN MONTECARLO	48
3.3 APLICACIÓN DEL ALGORITMO DE RECURSIÓN DE PANJER	50
3.4 APLICACIÓN DE APROXIMACIÓN ANALÍTICA DE BÖCKER Y KLÜPPELBERG	52

CONCLUSIONES	55
APÉNDICE	57
CÓDIGOS DE MATLAB	57
BIBLIOGRAFÍA	66

TABLAS

Tabla 2.1. Distribuciones de cola pesada.

Tabla 2.2. Constantes y Valor Inicial.

Tabla 2.3. Algunas Distribuciones Subexponenciales.

Tabla 2.4. Aproximaciones asintóticas para el $\text{OpVaR}_t(\alpha)$, para las distribuciones de severidad más usuales.

Tabla 3.1. Estadísticas relevantes de los datos.

Tabla 3.2. Ajuste de la Distribución de Frecuencias.

Tabla 3.3. Ajuste de la Distribución de Severidad.

Tabla 3.4. Resumen de Resultados.

INTRODUCCIÓN

Aunque ha sido una preocupación histórica, la cuantificación del riesgo operacional se ha convertido en una actividad obligada en las instituciones financieras, desde el surgimiento de la Convergencia Internacional de Medidas y Estándares de Capital, o Nuevo Acuerdo de Basilea, emitido por el Banco de Pagos Internacionales (*Bank for International Settlements-BIS*) en el año 2004, que incorporó ese riesgo en el cálculo de la relación de solvencia, para estimar los requerimientos de capital. Se debe cuantificar el riesgo para satisfacer los estándares regulatorios; pero el objetivo fundamental debe ser robustecer los procesos de control y la disminución de pérdidas potenciales, y en general, fortalecer la toma de decisiones tendientes a la generación de valor. Según ese acuerdo, se define el riesgo operacional como la pérdida potencial de una entidad por fallas o deficiencias en los sistemas internos, en los procesos, en las personas, o algunos factores externos. Se incluye el riesgo legal, pero se excluye el riesgo de reputación. La norma colombiana, de la Superintendencia Financiera Colombiana, incluye además el riesgo de reputación.

El riesgo legal es la pérdida posible por sanciones o indemnizaciones como consecuencia del incumplimiento de normas, regulaciones u obligaciones contractuales o por fallas en los contratos o en las transacciones. El riesgo de reputación corresponde a pérdidas posibles por desprestigio, mala imagen o publicidad negativa.

El procesamiento sistemático de cualquier tipo de riesgo en finanzas, cubre tres etapas básicas: Identificación, cuantificación y gestión. Entre los retos relativos al riesgo operacional se incluyen desde la carencia de una estandarización

conceptual, y la prevalencia de una cultura reactiva, hasta la falta de una metodología unificada para la cuantificación. En la etapa central de cuantificación, la modelación matemática y estadística entran a ser las herramientas inevitables.

Para el proceso de modelación y cuantificación del riesgo operacional, en el ámbito mundial, especialmente durante los últimos años, se han desarrollado múltiples estudios. Entre los autores que han generado investigaciones relacionadas con riesgo operacional en los últimos años se tiene: Frachot, P. Georges, T. Roncalli (2001), analizan el modelo LDA como una adecuación de técnicas actuariales; Marcelo Cruz (2002) aporta un análisis cuantitativo para el riesgo operacional y propone estrategias para modelar, gestionar y cubrir ese riesgo. Antoine Frachot, Olivier Moudoulaud, Thierry Roncalli (2003), plantean la implementación de métodos de medición avanzada para riesgo operacional, de tal forma que los estándares cualitativos y cuantitativos establecidos por Basilea puedan ser reconciliados. Antoine Frachot, Thierry Roncalli, Eric Salomon (2004) analizan el problema de correlación en riesgo operacional. Pavel V. Shevchenko y M. V. Wüthrich (2006) proponen modelar el riesgo operacional mediante inferencia bayesiana, y la combinación de datos de pérdidas con opiniones de expertos para tratar de superar la dificultad habitual de escasez de datos. Johanna Nešlehová, Paul Embrechts, Valérie Chavez-Demoulin (2006), investigan algunos efectos de las colas pesadas en riesgo operacional, y en especial aspectos relacionadas con los modelos de media infinita. Otros estudios reconocidos en este campo son Ioannis S. Akkizidis, Vivianne Bouchereau (2006); Klaus Böcker (2006); Kabir Dutta, Jason Perry (2006); Matthias Degen, Paul Embrechts, Dominik D. Lambrigger (2006); Falko Aue, Michael Kalkbrener (2007); Anna Chernobai, Svetlozar T. Rachev (2007), Shevchenko, P. y G. Temnov (2008), Peters, G. W. P. V. Shevchenko and M. V. Wüthrich (2009).

La diversidad de métodos refleja diferentes niveles de sofisticación y sensibilidad al riesgo. Los AMA (Advanced Measurement Approaches), término que se ha

convertido en genérico para representar diversos modelos de medición avanzada, según lo establecido por Basilea, admiten flexibilidad en la cuantificación del riesgo operacional, y permiten a las entidades elaborar su propio sistema de modelación y medición del riesgo operacional. Análogo a lo que ocurre con riesgo de mercado, Basilea no establece un tipo particular de modelo, sino que determina un conjunto de estándares cualitativos y cuantitativos que deben ser satisfechos por cualquier modelo interno que se decida implementar. Entre los estándares cuantitativos hay dos fundamentales: la medida del riesgo operacional es un OpVaR- *Operational Value at Risk*) a un nivel de confianza del 99.9%, y el método de cuantificación debe capturar eventos potenciales de pérdidas severas en la cola. El Valor en Riesgo Operacional a un nivel de confianza α , denotado OpVaR (α), significa el nivel de pérdidas que sólo es excedido con una probabilidad $1 - \alpha$, en el horizonte de tiempo considerado.

Entre los métodos de medición avanzada AMA, la técnica más utilizada es el método de distribución de pérdidas, conocido como el LDA (Loss Distribution Approach).

En este trabajo se desarrollan formalmente tres de las alternativas de modelación y cuantificación asociadas a los modelos LDA: El método de Simulación Montecarlo, el algoritmo de Recursión de Panjer y la Aproximación Analítica de Böcker y Klüppelberg. Luego, mediante algoritmos implementados en MatLab, cuyos códigos se anexan en el Apéndice, se aplican esas técnicas a una base de datos de riesgo operacional. Finalmente, se analizan los resultados obtenidos para hacer comparaciones y obtener conclusiones sobre su desempeño relativo.

CAPÍTULO 1. PRELIMINARES

Se consideran en este capítulo, sin pretender ser exhaustivos, algunos conceptos relevantes, que se utilizan en el contenido posterior del presente trabajo. Algunos elementos teóricos adicionales son explicados en el contexto.

- **Banco de Pagos Internacionales (BPI):** El Banco de Pagos Internacionales (BPI) (inglés: Bank for International Settlements BIS) Es una organización internacional con sede en Basilea (Suiza), que es considerada como el banco central de bancos centrales. Tiene como misión fomentar la cooperación financiera y monetaria internacionales entre los bancos centrales y facilitar las operaciones financieras, generando estudios y buscando soluciones, para propiciar la estabilidad financiera en el ámbito internacional.
- **Comité de Basilea:** Es la denominación abreviada del **Comité de Supervisión Bancaria de Basilea (BCBS, sigla de *Basel Committee on Banking Supervision*)**, organización mundial, con sede en el Banco de Pagos Internacionales, que reúne a las autoridades de supervisión bancaria, y tiene como función fortalecer la estabilidad de los sistemas financieros. Entre las normas fundamentales emitidas por ese comité, se tienen los conocidos Acuerdos de Basilea. El primero de ellos publicado en 1988, y el segundo, conocido como Basilea II, emitido en el año 2004.

El Comité de Basilea es un organismo que no tiene capacidad normativa directa, sino que emite directrices y recomendaciones que luego pueden ser convertidas en normatividad interna por los reguladores de cada país, lo que en Colombia corresponde a la Superintendencia Financiera. Aunque la

adopción de esas recomendaciones es voluntaria la mayoría de los países las acogen.

- **Acuerdo de Basilea II:** En el año 2004, el Comité de Basilea publicó un conjunto de recomendaciones para establecer un capital mínimo que debe tener una entidad bancaria en función de los riesgos que afronta. Este capital debe ser suficiente para hacer frente a las pérdidas potenciales por los diferentes tipos de riesgo. El documento definitivo *Convergencia Internacional de Medidas y Normas de Capital*, fue publicado en el año 2006, y es conocido como Basilea II.
- **Patrimonio Técnico:** Nivel o porcentaje mínimo de patrimonio adecuado o relación de solvencia que deben cumplir las entidades financieras, con respecto a sus activos ponderados por nivel de riesgo. El patrimonio técnico es una definición especial dada por la reglamentación, que incluye sólo algunos de los rubros del patrimonio contable con específicas ponderaciones. Este patrimonio técnico se clasifica en capital primario y secundario. En el capital primario se contemplan los rubros que son realizables en el corto plazo, mientras que en el capital secundario, se ubican los rubros realizables en el mediano y largo plazo.
- **Relación de solvencia (Índice de Basilea):** Entre los estándares cuantitativos planteados por Basilea II se estable que las entidades deben tener fondos propios, o carga de capital, de por lo menos el 8% del total de activos ponderados por riesgo. Las ponderaciones deben efectuarse sobre los riesgos de crédito, riesgo de negociación, riesgo de tipo de cambio y riesgo operacional. La relación de solvencia mide el nivel de respaldo patrimonial de las operaciones activas de las entidades financieras. Está dada por la relación entre el patrimonio técnico y los activos ponderados por su nivel de riesgo. De manera formal se tiene:

$$\frac{C}{APRC + 12.5RM + 12.5RO} \geq 0.08.$$

C: Capital mínimo requerido.

APRC: Activos ponderados por riesgo crediticio.

RM: Carga de capital por riesgo de mercado.

$12.5 * RM = APRM =$ Activos ponderados por riesgo de mercado.

RO = Carga de capital por riesgo operacional.

$APRO = 12.5 * RO =$ Activos ponderados por riesgo operacional.

$APR = APRC + APRM + APRO =$ Total de activos ponderados por riesgo.

Aunque a nivel internacional el Comité de Basilea plantea que la relación de solvencia debe ser por menos el 8%, en Colombia se exige como mínimo el 9%.

- **Capital regulatorio:** El capital regulatorio es el establecido por el regulador con el objeto de minimizar el riesgo de quiebra y los problemas de riesgo sistémico.
- **Capital económico:** Se refiere al nivel de capital acorde con los riesgos que la entidad asume -con independencia de la existencia de activos- Es la cuantificación probabilística del importe de pérdidas futuras potenciales. La utilización de un buen modelo de capital económico permite a la alta dirección estar preparada para anticipar problemas potenciales.

El concepto del capital económico se diferencia de "capital regulador" en el sentido que el "capital regulador" es el capital obligatorio que los reguladores requieren que se debe tener, mientras que el capital económico es la mejor estimación del capital requerido que las instituciones financieras utilizan internamente para manejar su propio riesgo y asignar el

coste de mantener el capital regulador entre diversas unidades dentro de la organización

- **Riesgo operacional:** Riesgo operacional u operativo, es la pérdida posible por deficiencias, fallas o inadecuaciones, en el recurso humano, los procesos, la tecnología, la infraestructura o por la ocurrencia de acontecimientos externos.
- **Ínfimo de un conjunto:** Se dice que un conjunto A está acotado inferiormente si y sólo si existe un número real c que es menor o igual que cualquier elemento de A . La mayor de las cotas inferiores se denomina ínfimo de A , y se denota $\inf(A)$.
- **Función cuantil:** La función cuantil de una distribución de probabilidad es la inversa de la función de distribución. $F^{-1}(p) = \inf\{x \in R: F(x) \geq p\}$, para una probabilidad $0 < p < 1$, generando el valor mínimo de x para el cual se acumula esa probabilidad. Entre los cuantiles se tienen los percentiles.
- **Cola de una distribución:** Sea X una variable aleatoria no negativa con función de distribución F . La cola de la distribución F se define como: $\overline{F(x)} = 1 - F(x)$, para $x > 0$.
- **Función generadora de momentos:** La función generadora de momentos de la variable aleatoria X es $E(e^{tx})$ y se denota por $M_X(t)$. Así que: $M_X(t) = E(e^{tx})$.
- **Distribución de cola pesada:** Si X es una variable aleatoria, con función de distribución F , y para su función generadora de momentos se cumple que $E(e^{sX}) = \infty$, para todo $s > 0$, entonces se dice que F es de cola pesada.

- **Distribución degenerada:** Una variable aleatoria X se dice que tiene una distribución degenerada en un punto h si su función de probabilidad es

$$p_X(x) = \begin{cases} 1, & \text{si } x = h \\ 0, & \text{si } x \neq h \end{cases}$$

La función de distribución de una variable aleatoria

degenerada es: $F(x) = \begin{cases} 0, & \text{si } x < h \\ 1, & \text{si } x \geq h \end{cases}$

- **Convolución:** Si X_1, X_2, \dots, X_n son variables aleatorias idénticamente distribuidas, con función de distribución F , la n -ésima convolución de la distribución F , se denota por F^{*n} , o por F^{n*} y está dada por:

$$F^{*n}(x) = P(X_1 + X_2 + \dots + X_n < x).$$

- **Distribución sub-exponencial:** Una función de distribución F con soporte $(0, \infty)$ es sub-exponencial, si para todo $n \geq 2$ se cumple que $\lim_{n \rightarrow \infty} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)} =$

$$n. \text{ O en forma equivalente si } \lim_{x \rightarrow \infty} \frac{P(X_1 + X_2 + \dots + X_n > x)}{P(\max\{X_1, X_2, \dots, X_n\} > x)} = 1.$$

CAPÍTULO 2. MÉTODO DE DISTRIBUCIÓN DE PÉRDIDAS (LDA)

La investigación en riesgo operacional ha llevado a muchas entidades a usar modelos heredados de las técnicas generales de estimación en seguros, los llamados “Métodos actuariales” o también referidos como “Métodos estadísticos”. Todos los métodos estadísticos se apoyan en bases de datos de eventos de pérdidas para estimar el riesgo operacional. La principal dificultad con estos métodos es la escasez de datos de buena calidad que reflejen las operaciones reales de la entidad, teniendo en cuenta que, según Basilea II, las estimaciones de riesgo operativo generadas internamente y utilizadas para efectos de capital regulatorio, deberán basarse en un período de observación de por lo menos cinco años, excepto cuando el banco utilice por primera vez el modelo AMA, caso en el cual se aceptará un período de observación de tres años.

La mayoría de esos modelos actuariales estiman una función de probabilidad para el riesgo operacional y, a partir de ella, deducen una medida de la carga de capital. Estos métodos varían en complejidad. Algunos se limitan a ajustar una curva para agregar las pérdidas experimentadas y esa curva es usada para estimar percentiles extremos; otros utilizan la frecuencia y severidad de los eventos de pérdidas, ajustándolas a distribuciones de probabilidad separadas, que luego son modeladas estocásticamente para producir una función de probabilidad combinada que se usa para deducir medidas de capital. Este es un procedimiento estándar aplicado en el campo actuarial.

El método de distribución de pérdidas, conocido como LDA (Loss Distribution Approach) es la alternativa más comúnmente utilizada entre los llamados modelos de medición avanzada (AMA).

El método LDA incluye la modelación separada de la distribución de probabilidad de la severidad y la distribución de probabilidad de la frecuencia de las pérdidas, y

luego las combina mediante simulación Montecarlo u otra técnica estadística para generar una distribución de pérdidas agregadas para cada combinación línea de negocio/tipo de riesgo para un horizonte temporal específico.

El principal aspecto consiste en ajustar los puntos de pérdidas totales observadas a una curva de ocurrencias de pérdidas. Esa curva será la que permita extrapolar a partir de puntos de datos para determinar el monto probable de pérdida máxima total o mínimo capital requerido para un nivel de confianza dado.

Los orígenes del LDA se ubican en las aplicaciones actuariales, desarrolladas por la industria de seguros durante muchos años (Bühlmann, 1970). Es una de las ideas más naturales importadas de técnicas actuariales, para ser aplicadas en situaciones específicas que caracterizan el riesgo operacional, y en particular cuando hay escasez de datos. Diversos estudios que se han hecho acerca de datos empíricos, confluyen en que esa escasez de datos tiene un impacto dramático en la carga de capital, por lo que, en metodologías diferentes al LDA, se impone la necesidad de usar para el tratamiento de ellos, procedimientos muy sofisticados.

El principal objetivo del modelo LDA es proporcionar un estimativo del riesgo operacional para una entidad y sus unidades de negocio, a partir de una distribución de pérdida que refleja los datos de pérdidas subyacentes. El LDA se soporta en la recopilación de datos de pérdidas históricas internas (frecuencia y severidad), que pueden ser complementados con datos externos, adecuadamente escalados.

Según Basilea II, en una entidad financiera la exposición al riesgo operacional se divide en una serie de líneas de negocio y eventos. En general, se consideran ocho líneas de negocio (finanzas corporativas, negociación y ventas, banca minorista, banca comercial, pagos y liquidación, servicios de agencia, administración de activos e intermediación minorista) y siete eventos de pérdidas (fraude externo, fraude interno, clientes, ejecución y administración de procesos, fallas tecnológicas, daños a activos físicos, relaciones laborales).

Las pérdidas son clasificadas en una matriz que relaciona las líneas de negocio de la organización y los eventos operacionales de pérdida. Aunque Basilea propone ocho líneas de negocio y siete tipos de eventos estándar para la construcción de esa matriz, cada firma tiene la libertad de considerar sus propias líneas y eventos, de tal forma que se adecuen a la estructura y necesidades de la organización, pero debe estar en capacidad de adaptar su estructura matricial de pérdidas a la forma matricial de Basilea II (Shevchenko, 2008).

El modelo LDA proporciona estimaciones para la pérdida agregada, tanto por línea de negocio como por evento, que luego son combinadas para estimar la pérdida operacional total.

En el LDA la pérdida total se define como una suma aleatoria de las distintas pérdidas:

$$L = \sum_{i=1}^7 \sum_{j=1}^8 S_{ij} \quad (2.1)$$

donde S_{ij} es la pérdida total en la celda i, j de la matriz de pérdidas. Las S_{ij} se calculan como:

$$S_{ij} = \sum_{N=1}^n X_{N_{ij}} \quad (2.2)$$

Donde N_{ij} es la variable aleatoria que representa el número de eventos de riesgo en la celda i, j (frecuencia de los eventos) y X_N es el monto de la pérdida en la celda i, j (severidad del evento). En consecuencia, las pérdidas son resultado de por lo menos dos diferentes fuentes de aleatoriedad, la frecuencia y la severidad.

Los modelos generalmente utilizados en la industria para analizar y cuantificar las pérdidas por riesgo operacional, aunque difieren en su implementación, la mayoría están basados esencialmente en la misma metodología.

El cálculo del riesgo operacional para la celda i, j , será representado por un percentil α determinado (por ejemplo, el 99,9%) de la distribución de pérdidas agregadas por período en esa celda, que en lo sucesivo se denota simplemente

como $S(x)$, porque, en primera instancia, el análisis se hace para cada celda específica.

La distribución de $S(x)$ se obtiene mediante el estudio por separado de la distribución frecuencias de pérdidas $p_n = P(N = n)$ y la distribución de severidad de las pérdidas $f_X(x)$. Estas dos distribuciones se asume que son independientes y estables sobre el tiempo.

La distribución de las pérdidas agregadas es resultado de una composición entre la variable aleatoria discreta asociada a la frecuencia, y la variable aleatoria continua asociada a la severidad de los eventos de riesgo. Esto es, la distribución de pérdida agregada por período puede ser formalmente obtenida como la media ponderada de la enésima convolución de la severidad, donde los pesos son las probabilidades de masa de las frecuencias.

La enésima convolución de la distribución de severidad es la probabilidad de ocurrencia del agregado de n pérdidas individuales.

Entonces, si las pérdidas agregadas para la celda específica, están dadas por: $S = X_1 + X_2 + \dots + X_N = \sum_{i=1}^N X_i$, donde N es la variable aleatoria de conteo del evento, y X_i es la variable aleatoria severidad, por ocurrencia del evento, y las X_i se asumen independientes e idénticamente distribuidas con función de distribución común dada por $F_X(x) = P(X \leq x)$, entonces la enésima convolución de la distribución de severidad, denotada por $F_X^{*n}(x)$, está dada por: $P(X_1 + X_2 + \dots + X_n \leq x) = F * F * \dots * F(x) = F_X^{*n}(x)$, y por lo tanto la función de distribución de las pérdidas agregadas está dada por: $G_S(x) = P(S \leq x) = \sum_{n=0}^{\infty} P(N = n) F_X^{*n}(x)$.

Diversos enfoques han sido desarrollados para calcular la distribución de pérdidas agregadas. La mayoría de ellos se basan en la suposición de que se tienen disponibles tanto la distribución de frecuencias, como la distribución de severidad o impacto.

El punto inicial del método LDA consiste en identificar tanto la distribución de frecuencias como la de severidad de las pérdidas, para obtener la distribución de

pérdidas agregadas mediante la composición de ellas. En general, no existe una forma analítica para expresar la distribución compuesta de pérdidas, y se hace necesario aplicar algoritmos numéricos para calcularla. En el contexto de la modelación del riesgo operacional algunos métodos muy usuales son la simulación Montecarlo, el algoritmo de recursión de Panjer y técnicas de inversión mediante transformadas; también es posible utilizar la aproximación en forma cerrada de Böcker y Klüppelberg, que es aplicable a situaciones específicas, como se demostrará más adelante.

A partir de la función de pérdidas agregadas $S(x)$ que se haya determinado, la carga de capital por riesgo operacional, $OpVaR(\alpha)$, como está definido previamente, se obtiene como:

a. $OpVaR(\alpha) = G_S^{-1}(\alpha)$, para un nivel de confianza α determinado, que Basilea II ha fijado en 99.9%, donde G_S es la función de distribución de las pérdidas agregadas.

b. Cuando la entidad demuestra hacer provisiones para las pérdidas esperadas, la carga de capital se obtiene restando las pérdidas esperadas del percentil α . Así que: $OpVaR(\alpha) = G_S^{-1}(\alpha) - E(S)$.

Como G_S es la función de distribución de las pérdidas agregadas, es claro que el $OpVaR$, con nivel de confianza α , significa el nivel de pérdidas por riesgo operacional que sólo es excedido con una probabilidad $1 - \alpha$, en el horizonte de tiempo considerado.

El VaR operacional así calculado corresponde al requerimiento o carga de capital por riesgo operacional para la institución financiera.

La aplicación de este método, requiere estimar para cada mixtura línea de negocio/evento de riesgo, las funciones de distribución de probabilidad de la frecuencia del evento y de su impacto, para calcular, a partir de ellas, la función de distribución de probabilidad de la pérdida acumulada por riesgo operacional.

2.1 DISTRIBUCIÓN DE FRECUENCIAS

Sea N una variable aleatoria que representa el número de eventos de riesgo operacional por período, en la celda (i, j) considerada (frecuencia de los eventos), y que sigue una distribución de probabilidad $p(n)$. Entonces $p(n) = P(N = n)$, $n \in N \cup \{0\}$.

Por lo tanto, la distribución de frecuencias de ocurrencia de eventos de riesgo operacional es modelada mediante una variable aleatoria discreta que representa el número de eventos observados durante un período de tiempo establecido. En este sentido, autores como Marcelo Cruz (2002), Mariano González (2004), Pavel Shevchenko (2006) y Christopher Lee Marshall (2001), muestran que la distribución de Poisson se ajusta a muchas situaciones reales de riesgo operacional. Sin embargo, recomiendan considerar, además, otras alternativas como la Binomial o la Binomial Negativa.

Para la especificación de la distribución de frecuencias, se ha argumentado sobre la conveniencia de utilizar sólo datos internos (Aue y Kalkbrener, 2007). Estos autores plantean que los datos internos de pérdidas reflejan más adecuadamente el perfil de riesgo de la institución; que es difícil garantizar la completitud de los datos de pérdidas de otras instituciones financieras, y que esa completitud es esencial para la calibración de la frecuencia. Así mismo, los requerimientos de datos para calibrar las distribuciones de frecuencias, son más bajos que para calibrar las distribuciones de severidad, en particular si se utiliza la distribución de Poisson.

En principio, cualquier distribución con soporte los enteros no negativos, podría ser seleccionada para modelar la frecuencia. Sin embargo, para los modelos LDA, las familias de distribuciones más utilizadas han sido la Poisson, y la binomial negativa (Johnson et al. 1993), Klugman et al. (2004). Una forma de decidir entre esas alternativas consiste en analizar la dispersión de la serie de tiempo empírica,

mediante la comparación de la media y la varianza. Si la serie de tiempo es equidispersa (tiene media igual a la varianza), se ajustará mejor una distribución de Poisson. Si se presenta sobredispersión (media menor que varianza), se modela la frecuencia con una distribución binomial negativa; y en el caso de subdispersión (media mayor que varianza) se ajusta mediante una distribución binomial. Como un análisis de dispersión no siempre es evidente con respecto a cuáles combinaciones media/varianza deben considerarse equidispersas, los resultados deben corroborarse con pruebas de bondad de ajuste como la chi-cuadrado. Otra prueba aplicable en este contexto, consiste en analizar el tiempo entre ocurrencias de eventos de pérdidas. Si los datos son extraídos de un proceso de Poisson, el tiempo entre ocurrencias de pérdidas sigue una distribución exponencial.

Según lo anterior, la distribución binomial negativa, permite modelar la sobredispersión, que significa que para todo $t > 0$, la varianza es mayor que su media, mientras que la distribución de Poisson es equidispersa, ya que se cumple la igualdad. Sin embargo, como se verá más adelante, en lo relacionado con riesgo operacional, la sobredispersión es de menor importancia, por lo cual se justifica frecuentemente la escogencia de la distribución de Poisson para describir las frecuencias de eventos de pérdidas operacionales.

En resumen, en esta etapa del proceso, el objetivo es encontrar la mejor distribución que pueda describir las ocurrencias aleatorias de los eventos de pérdida. Utilizando las bases de datos disponibles se puede determinar la frecuencia de los eventos, y usar métodos estadísticos para ajustar varias distribuciones sobre los datos de los eventos de pérdida. En la práctica, muchas entidades que aplican el LDA asumen que las frecuencias de pérdidas operacionales siguen una distribución de Poisson. La distribución de Poisson se caracteriza por un solo parámetro λ que es la media y la varianza de la distribución.

2.2 DISTRIBUCIÓN DE SEVERIDAD

La siguiente etapa consiste en ajustar distintos modelos de distribución probabilística a la serie de datos históricos de pérdidas operacionales desglosadas por su tipología para una determinada línea de negocio y evento de pérdida. Se trata de encontrar la distribución de probabilidad que mejor se ajuste a los datos observados y estimar sus parámetros. Christopher Lee Marshall (2001), Marcelo Cruz (2002), Mariano González (2004), Pavel Shevchenko y J. Donnelly (2005) y Carrillo (2006), proponen la distribución Lognormal o la de Weibull como las más recomendables para modelar la severidad, aunque en la práctica puede ocurrir que ninguna distribución simple se ajuste satisfactoriamente a los datos, y que pueda ser necesario recurrir a una mixtura de distribuciones para variables aleatorias continuas.

Sea X el monto de la pérdida en la celda (i, j) de la matriz de pérdidas (Severidad o impacto económico por evento). Esta variable tendrá una función de densidad $f_X(x)$, y una función de distribución $F_X(x)$, dada por $F_X(x) = P(X \leq x)$.

La distribución de severidad de pérdidas describe el tamaño de los montos de pérdidas causadas por un evento dado. Modelar la severidad puede ser más complicado que modelar la frecuencia, debido al tamaño impredecible de altos eventos de severidad. En este sentido, existen tres clases de severidades operacionales que se deben tener en cuenta: baja severidad, alta frecuencia (HFLI); alta severidad, baja frecuencia (LFHI); y severidad catastrófica, que ocurre muy pocas veces.

Para ajustar los datos de severidad operacional se deben tener en cuenta esas tres clases de severidades. En este punto el principal aspecto es escoger una distribución que pueda cubrir todas las pérdidas que puedan ocurrir en un determinado período. Un ajuste adecuado de los datos históricos puede no ser suficiente, ya que se puede asumir que hay una gran cantidad de pérdidas por riesgo operacional que no son reportadas, y por lo tanto no son registradas en la

base de datos. La incertidumbre en el registro de pérdidas por riesgo operacional puede ser explicada por diversas causas como el intento de las secciones dentro de la institución, de evitar una mala reputación, por lo cual tratarán de resolver internamente los problemas. Igualmente puede ocurrir pérdida de datos para alguna línea de negocio o eventos, generando bases de datos incompletas, o eventos extremos escasamente representados en la base de datos interna. Esto, a su vez, genera incertidumbre acerca de la severidad estimada por pérdidas operacionales.

Por lo anterior, para seleccionar la distribución que mejor se ajuste a la severidad de las pérdidas, puede ser más adecuado escoger una distribución de cola pesada para representar esos posibles montos de pérdida, esperando que también considere pérdidas reales no tenidas en cuenta en la base de datos.

Para las distribuciones de severidad, entre las distribuciones de cola pesada con soporte $(0, \infty)$, (Klugman, Panjer and Willmot (2004)), las familias generalmente utilizadas en modelación de riesgo operacional son: La distribución log-normal, la distribución de valor extremo generalizada, la distribución generalizada de Pareto y la distribución de Weibull, que se describen en la Tabla 2.1.

Tabla 2.1. Distribuciones de cola pesada

Distribución	Función de distribución	Parámetros
Log-Normal	$F(x) = \theta \left(\frac{\ln x - \mu}{\sigma} \right)$	$(\mu, \sigma > 0)$
Weibull	$F(x) = 1 - \exp \left(- \left(\frac{x - \alpha}{\beta} \right)_+^\zeta \right)$	$(\alpha, \beta > 0, \zeta)$
Pareto Generalizada	$F(x) = 1 - \left(1 + \zeta \frac{x - \alpha}{\beta} \right)_+^{-\frac{1}{\zeta}}$	$(\alpha, \beta > 0, \zeta)$
Valor extremo Generalizada	$F(x) = \exp \left(- \left(1 + \zeta \frac{x - \sigma}{\beta} \right)_+^{-\frac{1}{\zeta}} \right)$	$(\alpha, \beta > 0, \zeta)$

Un criterio para seleccionar la mejor distribución para ajustar la severidad de las pérdidas podría ser escoger la distribución que mejor se ajuste a la cola superior. Esto significa que la mejor distribución debería sobreestimar los eventos de riesgos extremos, para ser tenidos en cuenta en la carga de capital.

Es comúnmente aceptado que las distribuciones Lognormal y Weibull se ajustan razonablemente a datos de pérdidas operacionales sobre una gran parte de los datos de riesgo operacional pero tienen un desempeño relativamente débil en la cola, ya que las pérdidas operacionales tienden a tener colas más pesadas que esas distribuciones, por lo cual se pueden producir subestimaciones de grandes pérdidas. Por el contrario, la distribución de Pareto produce un buen ajuste en la cola, cuando existen suficientes datos para permitir este análisis, pero un ajuste débil en el cuerpo de la distribución.

Es ineludible resaltar que el problema de encontrar una distribución de severidad que describa debidamente los datos empíricos de pérdidas es una tarea no trivial, y que la parametrización de distribuciones apropiadas para la frecuencia y la severidad, es parte primordial de cualquier modelo AMA. Klugman, Panjer and Willmot (2004), desarrollan un análisis minucioso de aspectos estadísticos, pruebas de hipótesis y estimación de parámetros en el contexto de modelos de pérdidas.

2.3 CUANTIFICACIÓN DE LAS PÉRDIDAS AGREGADAS

En esta sección se plantean tres alternativas para cuantificar las pérdidas agregadas en una celda específica (i, j) de la matriz de pérdidas: El método de simulación Monetcarlo, el Algoritmo de Recursión de Panjer, y la Aproximación Analítica de Böcker y Klüppelberg.

Se hace la presentación formal de cada uno de los métodos, y se describen los pasos del algoritmo correspondiente. Además, los algoritmos fueron

implementados en Matlab, para diversas combinaciones de distribuciones de frecuencia y severidad, y los códigos respectivos se muestran en el Apéndice.

2.3.1 Simulación Montecarlo (SMC)

Una de las alternativas más utilizadas para generar la distribución de pérdidas agregadas mediante composición entre frecuencia y severidad es la Simulación Montecarlo.

Mediante el enfoque de Simulación Montecarlo se estima la distribución de pérdidas agregada utilizando un número suficiente de escenarios hipotéticos, generados aleatoriamente, a partir de las distribuciones de severidad y frecuencia.

Entre los autores que plantean esta metodología están Chapelle, Crama, Hübner y Peters (2005), que describen el procedimiento de Simulación Montecarlo, modelando la distribución de frecuencias mediante una distribución de Poisson con parámetro igual al número de pérdidas observadas durante el período completo.

El procedimiento tendría los siguientes pasos:

1. Determinar la combinación línea/evento que se quiere simular.
2. Generar una muestra aleatoria de la distribución de frecuencias asumida.
3. Generar una muestra aleatoria de la distribución de severidad.
4. Generar la distribución de pérdidas agregadas.

A continuación se presentan dos formas alternativas, que difieren ligeramente en su interpretación, para implementar el algoritmo de Simulación Montecarlo con el fin de generar una distribución de pérdidas agregadas por período.

2.3.1.1 SMC. Algoritmo 1.

1. Generar un gran número M de variables aleatorias Poisson (λ) (por ejemplo, 10000). Esos M valores representan el número de eventos para cada uno de los M períodos simulados.

2. Para cada período generar el número requerido de variables aleatorias de severidad. Si el número simulado de eventos para el período m es k , entonces simular k severidades de pérdidas y sumarmas para obtener la pérdida agregada para el período.
3. El vector obtenido representa M períodos simulados. Si $M = 10000$, cuando se ordenan en forma creciente, el más pequeño valor representa el cuantil 0.0001, el segundo será el cuantil 0.0002..., y así sucesivamente, lo que permite calcular fácilmente el VaR Operacional.
4. Calcular el OpVaR, para el nivel de confianza α específico.

2.3.1.2 SMC. Algoritmo 2.

1. Generar un valor aleatorio n de la distribución de frecuencia.
2. Generar n valores aleatorios de la distribución de severidad. Denotar estas pérdidas simuladas por L_1, L_2, \dots, L_n .
3. Sumar las n pérdidas simuladas y obtener una pérdida para el período:
$$S = L_1 + L_2 + \dots + L_n.$$
4. Retornar al paso 1, y repetir un número muy grande M de veces. Por lo tanto se obtienen S_1, S_2, \dots, S_M ,
5. Formar el histograma de S_1, S_2, \dots, S_M , que representará la distribución de pérdidas agregadas simuladas para el período.
6. Calcular el percentil 99.9 de la distribución de pérdidas agregadas simuladas.
7. Calcular la media de la distribución de pérdidas agregadas simuladas para el período. La pérdida esperada (EL) se calcula como la media de esas pérdidas agregadas simuladas.
8. La carga de capital por riesgo operacional, el $OpVaR(99.9\%)$, será el percentil 99.9 de las pérdidas agregadas simuladas, o la diferencia entre ese percentil y la media de la distribución de pérdidas agregadas simuladas, cuando se manejan provisiones para las pérdidas esperadas.

En el Apéndice se presenta el código del algoritmo implementado en Matlab, para estimar las pérdidas agregadas aplicando simulación Montecarlo. El algoritmo diseñado es fácil de adaptar a cualquier combinación de distribuciones de frecuencia y severidad.

2.3.2 Algoritmo de recursión de Panjer

Como se ha planteado, en el método de distribución de pérdidas (LDA) se construye la distribución de pérdidas agregadas a partir de dos distribuciones asociadas a cada celda (i, j) combinación de evento de pérdida/línea de negocio.

Se tiene una variable aleatoria de conteo N con función de probabilidad $p(n) = P(N = n)$, que corresponde a la distribución de frecuencias de pérdidas. Además, una variable aleatoria continua X que representa el impacto o severidad de la pérdida, cuando ocurre el evento.

El algoritmo de Panjer está basado en el cálculo de la distribución compuesta mediante convoluciones, como se describe más adelante. Se utiliza el hecho de que la distribución de la suma de dos variables aleatorias continuas independientes puede ser calculada como una convolución.

Las pérdidas agregadas estarán dadas por:

$$S = X_1 + X_2 + \dots + X_N = \sum_{i=1}^N X_i \quad (2.3)$$

donde $X_i = 0$, cuando $N = 0$.

Usualmente se hacen dos supuestos fundamentales:

- a. X_1, X_2, \dots son variables aleatorias idénticamente distribuidas.
- b. Las variables aleatorias N, X_1, X_2, \dots son independientes.

La expresión (2.3) será una suma aleatoria, y se asume que sus distribuciones componentes satisfacen los supuestos.

Esas dos distribuciones (frecuencia y severidad), representan el núcleo del modelo LDA, y se utilizan para calcular la pérdida operacional agregada para la celda, en un horizonte temporal específico, generalmente un año.

La distribución de probabilidad de las pérdidas agregadas, está dada por $G_S(x) = P(S \leq x)$, y se puede expresar como:

$$\begin{aligned} G_S(x) &= P(S \leq x) = \sum_{n=0}^{\infty} P(S \leq x / N = n)P(N = n) \\ &= \sum_{n=0}^{\infty} P(X_1 + X_2 + \dots + X_n \leq x)P(N = n) \end{aligned} \quad (2.4)$$

Además,

$$P(X_1 + X_2 + \dots + X_n \leq x) = F * F * \dots * F(x) = F_X^{*n}(x) \quad (2.5)$$

donde $F_X^{*n}(x)$ es la n -ésima convolución de F , que es la función de distribución de X , y se tiene que:

$$\begin{aligned} F_X^{*0}(x) &= \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \\ F_X^{*1}(x) &= F_X(x) \text{ y } F_X^{*n}(x) = F_X^{*(n-1)}(x) * F_X(x). \end{aligned} \quad (2.6)$$

Por lo tanto la expresión (2.4) se transforma en:

$$G_S(x) = \sum_{n=0}^{\infty} P(N = n) F_X^{*n}(x) \quad (2.7)$$

Donde $F_X(x) = P(X \leq x)$ es la función de distribución común de las X .

La distribución de S es llamada la distribución compuesta.

Si X es una variable aleatoria continua con probabilidad cero para valores no positivos, la convolución se define recursivamente en la siguiente forma:

$$F_X^{*k}(x) = \int_0^x F_X^{*(k-1)}(x-y)f_X(y)dy, \text{ para } k = 2,3 \quad (2.8)$$

Para $k = 1$ la ecuación se reduce a $F_X^{*1}(x) = F_X(x)$.

Diferenciando en (2.8) se obtiene la función de densidad:

$$f_X^{*k}(x) = \int_0^x f_X^{*(k-1)}(x-y)f_X(y)dy, \text{ para } k = 2,3,\dots \quad (2.9)$$

Por lo tanto, si X es continua, entonces S tiene una función de densidad de probabilidad, que para $x > 0$, está dada por:

$$g_S(x) = \sum_{n=1}^{\infty} P(N = n) f_X^{*n}(x) \quad (2.10)$$

Si X tiene una distribución discreta, con probabilidades en $0,1,2, \dots$ se tendrá:

$F_X^{*k}(x) = \sum_{y=0}^{\infty} F_X^{*(k-1)}(x-y)f_Y(y)$ para $x = 0,1, \dots$ para $k = 2,3, \dots$ la correspondiente función de probabilidad es:

$$f_X^{*k}(x) = \sum_{y=0}^{\infty} f_X^{*(k-1)}(x-y)f_Y(y) \text{ para } x = 0,1, \dots k = 2,3,\dots$$

Si la distribución de severidad es discreta con función de probabilidad $p(x) = P(X = x)$, entonces la distribución de pérdidas agregadas también es discreta, y por analogía con lo anterior, la función de probabilidad de S se puede obtener directamente como:

$$g_S(x) = \sum_{n=0}^{\infty} p^{*n}(x)P(N = n) \quad (2.11)$$

donde $p^{*n}(x) = p * p * \dots * p(x) = P(X_1 + X_2 + \dots + X_n = x)$ y $p^{*0}(x) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases}$

En estas condiciones se tiene que $G_S(x)$ de la ecuación (2.4) se puede expresar como:

$$G_S(x) = \begin{cases} \sum_{n=1}^{\infty} p(n) F_X^{*n}(x), & x > 0 \\ p_0, & x = 0 \end{cases} \quad (2.12)$$

Donde, $F_X(x)$ es la probabilidad de que la cantidad agregada de n pérdidas sea x ; es decir, la función de distribución de las X . La correspondiente función de densidad está dada por:

$$g_S(x) = \begin{cases} \sum_{n=1}^{\infty} p(n) f_X^{*n}(x), & x > 0 \\ p_0, & x = 0 \end{cases} \quad (2.13)$$

Una de las alternativas para calcular $G_S(x)$ de la ecuación (2.12) o $g_S(x)$ de la ecuación (2.13) es el algoritmo de recursión de Panjer que se demuestra a continuación.

2.3.2.1 Algoritmo de Panjer

2.3.2.1.1 Lema

Sean X_1, X_2, \dots, X_n , variables aleatorias independientes e idénticamente distribuidas, y que toman valores en los enteros positivos. Entonces, para valores enteros positivos de x , y para la convolución $f^{*n}(x) = P(X_1 + X_2 + \dots + X_n = x)$, se tiene:

$$(a). f^{*n}(x) = \sum_{i=1}^x f(i) f^{*(n-1)}(x - i).$$

$$(b). f^{*n}(x) = \frac{n}{x} \sum_{i=1}^x i f(i) f^{*(n-1)}(x - i).$$

Demostración

Para $n = 1$, es claro que los dos resultados se reducen a $f^{*1}(x) = f(x) f^{*0}(x)$.

Para $n > 1$, el resultado (a) usando la ley de probabilidad total para evaluar: $P(X_1 + X_2 + \dots + X_n = x)$ condicionada al valor que tome X_1 , se tiene:

$$P(X_1 + X_2 + \dots + X_n = x) = \sum_{i=1}^x P(X_1 = i) P(X_2 + \dots + X_n = x - i).$$

Pero $P(X_2 + \dots + X_n = x - i)$ y $P(X_1 + X_2 + \dots + X_n = x)$ corresponden, respectivamente, a las convoluciones $(n - 1)$ y (n) , de $f(i)$.

Entonces $f^{*n}(x) = \sum_{i=1}^x f(i)f^{*(n-1)}(x - i)$.

Para $n > 1$, se obtiene el resultado (b) considerando la esperanza condicional $E(X_k|X_1 + X_2 + \dots + X_n = x)$ para $k = 1, 2, 3, \dots, n$. Por simetría esas cantidades son las mismas para todos esos valores de k . Como su suma es x , cada uno es igual a $\frac{x}{n}$.

Entonces la esperanza condicional es evaluada como:

$$E(X_1|X_1 + X_2 + \dots + X_n = x) =$$

$$\sum_{i=1}^x iP(X_1 = i)P(X_2 + \dots + X_n = x - i)/P(X_1 + X_2 + \dots + X_n = x).$$

Pero $P(X_2 + \dots + X_n = x - i)$ y $P(X_1 + X_2 + \dots + X_n = x)$ corresponden, respectivamente, a las convoluciones $(n - 1)$ y (n) , de $f(i)$. Es decir, $f^{*(n-1)}(x)$ y $f^{*n}(x)$. Despejando $f^{*n}(x)$ se tiene el resultado: $f^{*n}(x) = \frac{n}{x} \sum_{i=1}^x if(i)f^{*(n-1)}(x - i)$. \square

2.3.2.1.2 Teorema de Panjer

Para distribuciones compuestas donde la distribución de probabilidad para N , el número de eventos, satisface la condición: $\frac{P(N=n)}{P(N=n-1)} = a + \frac{b}{n}$, para $n = 1, 2, \dots$, y donde la distribución de severidad $f(i)$ está restringida a los enteros positivos, la función de probabilidad de las pérdidas agregadas cumple la siguiente fórmula recursiva:

$$g_S(x) = \sum_{i=1}^x \left[a + \frac{bi}{x} \right] f(i)g_S(x - i), \text{ para } x = 1, 2, 3, \dots, \text{ con valor inicial dado por } g_S(0) = P(N = 0).$$

Demostración

Para la distribución compuesta se tiene: $g_S(x) = \sum_{i=1}^{\infty} P(N = n) f^{*n}(x)$.

Si la distribución de frecuencias cumple la condición $P(N = n) = \left[a + \frac{b}{n} \right] P(N = n - 1)$, entonces: $g_S(x) = a \sum_{i=1}^{\infty} P(N = n - 1) f^{*n}(x) + \sum_{i=1}^{\infty} \frac{b}{n} P(N = n - 1) f^{*n}(x)$, y por el lema 2.3.2.1.1, se cumple que:

$$g_S(x) = a \sum_{n=1}^{\infty} P(N = n - 1) \sum_{i=1}^x f(i) f^{*(n-1)}(x - i) + \sum_{n=1}^{\infty} \frac{b}{n} P(N = n - 1) \sum_{i=1}^x i f(i) f^{*(n-1)}(x - i)$$

Intercambiando el orden de las sumatorias, se tiene:

$$g_S(x) =$$

$$a \sum_{i=1}^x f(i) \sum_{n=1}^{\infty} P(N = n - 1) f^{*(n-1)}(x - i) + \frac{b}{x} \sum_{i=1}^x i f(i) \sum_{n=1}^{\infty} \frac{b}{n} P(N = n - 1) f^{*(n-1)}(x - i) =$$

$$a \sum_{i=1}^x f(i) g_S(x - i) + \frac{b}{x} \sum_{i=1}^x i f(i) g_S(x - i) = \sum_{i=1}^x \left[a + \frac{bi}{x} \right] f(i) g_S(x - i).$$

□

2.3.2.2 Comentarios sobre el Algoritmo de Panjer

El algoritmo parte de una familia particular de distribuciones de frecuencias para el número de eventos, pertenecientes a la llamada Clase de Panjer $(a, b; 0)$, lo que significa que cumplen la siguiente fórmula de recurrencia:

$$p_n = \left(a + \frac{b}{n} \right) p_{n-1}, n = 1, 2, \dots \quad 2.14)$$

Donde a, b son constantes reales que determinan la distribución de probabilidad de la frecuencia y p_n es la probabilidad de que ocurran n eventos.

Sundt y Jewell (1981) demostraron que las distribuciones de la clase de Panjer $(a, b; 0)$ son las distribuciones Poisson ($a = 0$), Binomial ($a < 0$), y la binomial negativa ($a > 0$). Como caso particular de la distribución binomial negativa también verificará la fórmula de recurrencia la distribución geométrica.

En efecto, considerando la fórmula recursiva $\frac{P(N=n)}{P(N=n-1)} = a + \frac{b}{n}$, para $n = 1, 2, \dots$, se obtiene:

a) Distribución de Poisson con parámetro λ :

La función de probabilidad es $P(N = n) = \frac{e^{-\lambda} \lambda^n}{n!}$. Así que: $\frac{P(N=n)}{P(N=n-1)} = \frac{\lambda}{n}$.

Entonces, $a = 0$ y $b = \lambda$ y la recursión es: $g_S(x) = \frac{\lambda}{x} \sum_{i=1}^x i f(i) g_S(x-i)$, con $g_S(0) = e^{-\lambda}$.

b) Distribución binomial negativa:

Parámetros r y p :

La función de probabilidad es: $P(N = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$.

Entonces,

$$\frac{P(N = n)}{P(N = n - 1)} = (1 - p) \left[n + r - \frac{1}{n} \right] = (1 - p) + \frac{(1 - p)(r - 1)}{n}$$

Así que, $a = (1 - p)$ y $b = (1 - p)(r - 1)$, y la recursión es:

$g_S(x) = (1 - p) \sum_{i=1}^x (r - 1) f(i) g_S(x - i)$, con $g_S(0) = p^r$.

c) Distribución Binomial con parámetros m y p :

La función de probabilidad es:

$$P(N = n) = \binom{m}{n} p^n (1 - p)^{m-n}$$

Así que:

$$\frac{P(N = n)}{P(N = n - 1)} = \frac{(m + 1 - n)}{n} \frac{p}{1 - p} = \frac{-p}{1 - p} + \frac{(m + 1)p}{n(1 - p)}$$

Entonces $a = \frac{-p}{1-p}$ y $b = \frac{(m+1)p}{(1-p)}$, y la recursión es:

$g_S(x) = \left(\frac{p}{1-p}\right) \sum_{i=1}^x \left[(m+1) \frac{i}{x} - 1 \right] f(i) g_S(x-i)$, con $g_S(0) = p^r$, con $g_S(0) = (1-p)^m$.

Los valores de las constantes a, b , y los valores de inicio correspondientes para cada una de esas distribuciones se resumen en la Tabla 2.2.

Tabla 2.2. Constantes y Valor Inicial.

Distribución	Constante a	Constante b	Valor inicial $p_0 = P(N = 0)$
Poisson (λ)	0	λ	$e^{-\lambda}$
Binomial negativa (r, p)	$1 - p$	$(r - 1)(1 - p)$	p^r
Binomial (m, p)	$-\frac{p}{1 - p}$	$(m + 1)\frac{p}{1 - p}$	$(1 - p)^m$

El Algoritmo de Panjer (1981) también establece que si la función de probabilidad de la frecuencia puede ser escrita como: $p_n = \left(a + \frac{b}{n}\right)p_{n-1}, n = 1, 3, \dots$, donde p_n representa la probabilidad de que el número de eventos sea n , y a y b son constantes reales, y la distribución de severidad es continua, entonces para g_S , la función de densidad de las pérdidas agregadas S , se cumple que:

$$g_S(x) = p_1 f(x) + \int_0^x \left(a + \frac{by}{x}\right) f(y) g(x - y) dy, x > 0 \quad (2.15)$$

Si la distribución de severidad es discreta y definida en los enteros no negativos, la correspondiente fórmula de recurrencia es:

$$g_i = \sum_{j=1}^i \left(a + \frac{bj}{i}\right) f_j g_{i-j}, i = 1, 2, 3 \dots, \text{ con } g_0 = p_0. \quad (2.16)$$

Para el caso especial de la distribución de Poisson, se tiene el siguiente resultado:

$$g_S(x) = \lambda e^{-\lambda} f(x) + \frac{\lambda}{x} \int_0^x y f(y) g(x - y) dy, x > 0. \quad (2.17)$$

Si además la distribución de severidad es discreta, la fórmula de recurrencia toma la siguiente forma:

$$g_S(x) = \frac{\lambda}{x} \sum_{i=1}^x i f(i) g_S(x - i) \quad (2.18)$$

En este caso, el algoritmo es aplicable para distribuciones de severidad discretas. Ello implica que la severidad, al ser una variable continua, debe ser discretizada

antes de aplicar dicho procedimiento, para lo cual se puede utilizar, por ejemplo, el método de redondeo.

Sin embargo, el principal inconveniente, en la práctica, radica en la complejidad al realizar las convoluciones.

En forma análoga se definen las distribuciones de la clase de la clase de Panjer $(a, b; 1)$ como la familia de distribuciones que satisfacen la recursión:

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, n = 2, 3, \dots \quad (2.19)$$

Donde la única diferencia con la clase de Panjer $(a, b; 0)$ es el valor el valor de inicio de la recursión.

En general, se definen las distribuciones pertenecientes a la clase de Panjer (a, b, k) , con $k \in N_0$, como aquellas que satisfacen esa recursión para $n \geq k + 1$, y $p_0 = p_1 = \dots = p_{k-1} = 0$. Todas ellas fueron identificadas por Sundt and Jewell (1981) para el caso $k = 0$, Willmot (1986) para el caso $k = 1$, y finalmente, por Hess et al. (2002), en general para $k \in N_0$.

Si la distribución de frecuencias pertenece a la clase de Panjer (a, b, k) , y la distribución de severidades es discreta, entonces el procedimiento clásico para calcular la distribución de pérdidas agregadas S es aplicar la recursión de Panjer.

Panjer y Wang (1993), mostraron que para distribuciones de severidad no degeneradas, la estabilidad numérica de la recursión de Panjer para una distribución de frecuencias de eventos de pérdidas operacionales perteneciente a la clase de Panjer (a, b, k) depende únicamente de los valores de a y b . Así que, como las distribuciones de Poisson, binomial y la binomial negativa, pertenecen a esta clase, el cálculo de las pérdidas agregadas, mediante el procedimiento descrito, es numéricamente estable.

2.3.2.3 Implementación de Algoritmos de Panjer

Se implementaron algoritmos de Panjer en Matlab, para estimar las pérdidas agregadas por riesgo operacional, considerando seis combinaciones posibles de distribuciones de frecuencia y severidad: Poisson-lognormal, Binomial-lognormal, Binomial negativa-lognormal, Poisson-Weibull, Binomial-Weibull, Binomial Negativa-Weibull. Los códigos correspondientes se presentan en el Apéndice.

2.3.3 Aproximación en forma cerrada de Böcker y Klüppelberg.

Klaus Böcker y Claudia Klüppelberg (2005) investigaron un modelo de distribución de pérdidas simple para el riesgo operacional, mostrando que se puede tener una aproximación para el $OpVaR(\alpha)$, en forma cerrada, cuando los datos de pérdida son de cola pesada, como ocurre en muchas situaciones reales. Ellos aplican esta distribución en particular al modelo de severidad de Pareto.

La aproximación analítica planteada por Böcker and Klüppelberg, como se demuestra más adelante, es una fórmula directa, aplicable en casos específicos, para el cuantil de la distribución de pérdidas agregadas en la siguiente forma:

$G_S^{-1}(\alpha) = F^{-1}\left(1 - \frac{1-\alpha}{E(N)}\right)$, que relaciona directamente el percentil $\alpha \rightarrow 1$, de las pérdidas agregadas, con un alto percentil dado por: $p = 1 - \frac{(1-\alpha)}{E(N)}$, de la función de distribución simple de las severidades F , dependiendo del número esperado de eventos por período $E(N)$.

Ese resultado, cuya formulación matemática se presenta más adelante, está basado en una propiedad analítica de las distribuciones pertenecientes a la clase de distribuciones subexponenciales, que permite expresar en el límite de alta severidad la convolución como una función de la distribución de la severidad de la pérdida individual.

La aproximación analítica de Böcker y Klüppelberg, para el cálculo de la carga de capital, también está soportada en la teoría de valores extremos. De acuerdo con uno de los requerimientos de Basilea de incorporar las propiedades de colas pesadas de las distribuciones de severidad, estos autores consideran la familia de las distribuciones subexponenciales, que incluye no solo las lognormales, sino también las distribuciones con colas más pesadas, en particular las de Pareto.

El Acuerdo de Basilea, en el contexto de los modelos AMA, permite a las entidades financieras la flexibilidad para que cada una implemente su modelo interno y sistema de cuantificación del riesgo operacional. En lugar de prescribir un modelo específico, se plantean unos estándares cualitativos y cuantitativos que los modelos internos deben satisfacer. Como se mencionó anteriormente, entre los estándares cuantitativos más relevantes, y que son la base de la aproximación analítica, se tiene que la medida de riesgo operacional es un VaR con un nivel de confianza del 99.9% y para un horizonte temporal de un año, y además, el método de medición debe capturar eventos de pérdidas potencialmente severas en la cola de la distribución de pérdidas.

Se presentan a continuación los elementos teóricos básicos, y el soporte matemático de la aproximación analítica de Böcker y Klüppelberg.

2.3.3.1 Distribuciones de severidad subexponenciales

Para tener en cuenta la propiedad de cola pesada de las distribuciones de severidad, como es el requerimiento de Basilea II, se consideran las distribuciones más conocidas con esa propiedad, que son Lognormal, Weibull y Pareto, que pertenecen a la llamada clase de distribuciones subexponenciales con soporte $(0, \infty)$, y se describen en la Tabla 2.3.

Tabla 2.3. Algunas Distribuciones Subexponenciales

NOMBRE	FUNCIÓN DE DISTRIBUCIÓN	PARÁMETROS
Lognormal	$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$	$\mu \in \mathbb{R}, \sigma > 0$
Weibull	$F(x) = 1 - e^{-(x/\theta)^\tau}$	$\theta > 0, 0 < \tau < 1$
Pareto	$F(x) = 1 - \left(1 + \frac{x}{\theta}\right)^{-a}$	$a, \theta > 0$

Para las distribuciones subexponenciales se cumple que sus colas decaen más suavemente que cualquier cola exponencial. Es decir, la cola de la suma de n variables aleatorias subexponenciales tiene el mismo orden de magnitud que la cola de la variable máxima entre ellas. Más precisamente:

$$\lim_{x \rightarrow \infty} \frac{P(X_1 + X_2 + \dots + X_n > x)}{P(\max\{X_1, X_2, \dots, X_n\} > x)} = 1, \text{ para algún } n \geq 2. \quad (2.20)$$

Esto significa que la suma de n severidades independientes e idénticamente distribuidas es más probable que sea grande debido a que uno de sus términos es grande; o en relación al riesgo operacional, pérdidas totales severas son principalmente causadas por una única pérdida grande, más que la consecuencia de pequeñas pérdidas independientes acumuladas.

El modelo LDA estándar tiene las siguientes características:

- Las severidades $\{X_i\}_{i \in \mathbb{N}}$ son variables aleatorias positivas, independientes e idénticamente distribuidas.
- El número de eventos de pérdida en un intervalo $[0, t]$, para $t > 0$, es aleatorio, y el correspondiente proceso de frecuencias es $\{N(t)\}_{t \geq 0}$.
- El proceso de severidad y el proceso de frecuencia se suponen independientes.
- La pérdida agregada $S(t)$ hasta el tiempo t está dada por $S(t) = \sum_{i=0}^{N(t)} X_i, t \geq 0$.

Además, si G_t es la distribución de pérdida agregada, entonces el Valor en Riesgo Operacional en el horizonte de tiempo t , a un nivel de confianza α está dado por

$OpVaR_t(\alpha) = G_t^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : G_t(x) \geq \alpha\}$, con $0 \leq \alpha \leq 1$. En particular si G_t es estrictamente creciente y continua se tiene que $OpVaR_t(\alpha) = G_t^{-1}(\alpha)$.

2.3.3.2 Lema de Kesten

Si F es una distribución subexponencial, entonces dado $\varepsilon > 0$, existe una constante finita K tal que, para todo $n \geq 2$, $\frac{\overline{F^{n*}}(x)}{\overline{F}(x)} \leq K(1 + \varepsilon)^n, x \geq 0$.

Demostración

Sea $\alpha_n = \sup_{x \geq 0} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)}$.

Además, $\frac{\overline{F^{(n+1)*}}(x)}{\overline{F}(x)} = 1 + \frac{F(x) - F^{(n+1)*}(x)}{\overline{F}(x)} = 1 + \int_0^x \frac{\overline{F^{n*}}(x-t)}{\overline{F}(x)} dF(t)$.

Entonces para cada $T < \infty$,

$$\begin{aligned} \alpha_{n+1} &\leq 1 + \sup_{0 \leq x \leq T} \int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x)} dF(y) + \sup_{x \geq T} \int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x-y)} \frac{\overline{F}(x-y)}{\overline{F}(x)} dF(y) \\ &\leq 1 + A_T + \alpha_n \sup_{x \geq T} \frac{F(x) - F^{2*}(x)}{\overline{F}(x)}, \text{ donde } A_T = (\overline{F}(T))^{-1} < \infty. \end{aligned}$$

Como F es subexponencial, dado cualquier $\varepsilon > 0$ se puede escoger T tal que $\alpha_{n+1} \leq 1 + A_T + \alpha_n(1 + \varepsilon)$. De lo cual se tiene que: $\alpha_n \leq (1 + A_T)\varepsilon^{-1}(1 + \varepsilon)^n$, por lo tanto:

con $K = (1 + A_T)\varepsilon^{-1}$, se tiene el resultado: $\frac{\overline{F^{n*}}(x)}{\overline{F}(x)} \leq K(1 + \varepsilon)^n, x \geq 0$. □

2.3.3.3 Teorema EKM. (Embrechts, Klüppelberg and Mikosch)

Sea el LDA estándar $S(t) = \sum_{i=0}^{N(t)} X_i, t \geq 0$. Suponga que las severidades X_i son subexponenciales con función de distribución F . Fije $t > 0$ y defina la distribución de frecuencias por $P(N(t) = n) = p_t(n)$, para $n \in N_0 = N \cup \{0\}$, y la distribución de

pérdida agregada dada por: $G_t(x) = \sum_{n=0}^{\infty} p_t(n) F^{n*}(x)$, $x \geq 0$, $t \geq 0$, donde $F(\cdot) = P(X_k \leq \cdot)$ es la función de distribución de X_k y $F^{n*}(\cdot) = P(\sum_{i=1}^n X_i \leq \cdot)$ es la n -ésima convolución de F con $F^{1*} = F$ y $F^{0*} = I_{[0, \infty)}$.

Si para algún $\varepsilon > 0$, $\sum_{n=0}^{\infty} (1 + \varepsilon)^n p_t(n) < \infty$, entonces se cumple que $\lim_{x \rightarrow \infty} \frac{EN(t)\bar{F}(x)}{\bar{G}_t(x)} = 1$. O equivalentemente:

$$\bar{G}_t(x) \sim EN(t)\bar{F}(x), \quad x \rightarrow \infty. \quad (2.21)$$

donde $EN(t)$ es el valor esperado de la frecuencia de pérdidas.

$\bar{F}(x) = 1 - F(x)$ es la distribución de severidad de la cola, y $\bar{G}_t(x) = 1 - G_t(x)$ es la distribución de la pérdida agregada de la cola.

Demostración

$$\frac{G_t(x)}{\bar{F}(x)} = \sum_{n=0}^{\infty} p_t(n) \frac{F^{n*}(x)}{\bar{F}(x)}.$$

Sea $\varepsilon > 0$ tal que $\sum_{n=0}^{\infty} (1 + \varepsilon)^n p_t(n) < \infty$.

$(1 + \varepsilon) > 1 \Rightarrow (1 + \varepsilon)^{-1} < 1$, entonces existe $\rho > 0$ tal que: $(1 + \varepsilon)^{-1}(1 + \rho) < 1$.

Así que $\frac{1+\rho}{1+\varepsilon} < 1$.

Por el Lema de Kesten se tiene:

$$(1 + \varepsilon)^{-1} \frac{\bar{F}^{n*}(x)}{\bar{F}(x)} \leq (1 + \varepsilon)^{-1} K(1 + \varepsilon)^n = K \left(\frac{1+\rho}{1+\varepsilon} \right)^n, \quad x \geq 0.$$

Además, la serie $\sum_{n=0}^{\infty} \left(\frac{1+\rho}{1+\varepsilon} \right)^n$ es convergente, por ser geométrica con razón

$$\frac{1+\rho}{1+\varepsilon} < 1.$$

Entonces,

$$\lim_{x \rightarrow \infty} \frac{\bar{G}_t(x)}{\bar{F}(x)} = \lim_{x \rightarrow \infty} \left(\sum_{n=0}^{\infty} p_t(n) \frac{\bar{F}^{n*}(x)}{\bar{F}(x)} \right) = \sum_{n=0}^{\infty} \lim_{x \rightarrow \infty} \left(p_t(n) \frac{\bar{F}^{n*}(x)}{\bar{F}(x)} \right)$$

$$= \sum_{n=0}^{\infty} p_t(n) \lim_{x \rightarrow \infty} \left(\frac{\overline{F^{n*}}(x)}{\overline{F}(x)} \right).$$

Pero como F es subexponencial, $\lim_{n \rightarrow \infty} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)} = n$. Reemplazando esta expresión en la anterior se tiene: $\lim_{x \rightarrow \infty} \frac{\overline{G}_t(x)}{\overline{F}(x)} = \sum_{n=0}^{\infty} p_t(n) n = EN(t)$.

Esta última igualdad por definición de valor esperado de $N(t)$.

$$\text{Entonces } \lim_{x \rightarrow \infty} \frac{EN(t)\overline{F}(x)}{\overline{G}_t(x)} = 1.$$

O en forma equivalente: $\overline{G}_t(x) \sim EN(t)\overline{F}(x), x \rightarrow \infty$. \square

2.3.3.4 Teorema de la Fórmula analítica para OpVaR

Considere el modelo LDA estándar para un $t > 0$ fijo y una severidad subexponencial con función de distribución F . Suponga, además, que el estimativo de cola (2.21) se cumple. Entonces:

$$OpVaR_t(\alpha) = F_t^{\leftarrow} \left(1 - \frac{1-\alpha}{EN(t)} (1 + o(1)) \right), \text{ cuando } \alpha \rightarrow 1. \quad (2.22)$$

Demostración

Si $\alpha \rightarrow 1$, entonces $x \rightarrow \infty$.

Además, $o(1)$ corresponde a una función que tiende a 0 si su argumento tiende a un límite, en nuestro caso si $\alpha \rightarrow 1$ o $x \rightarrow \infty$.

Del teorema 2.3.3.3, se tiene que $\overline{G}_t(x) \sim EN(t)\overline{F}(x)$.

Además, $\overline{G}_t(x) \sim EN(t)\overline{F}(x) \Rightarrow 1 - G_t(x) = EN(t)\overline{F}(x)(1 + o(1)), x \rightarrow \infty$.

O en forma equivalente: $G_t(x) = 1 - EN(t)\overline{F}(x)(1 + o(1)), x \rightarrow \infty$.

Tomando el lado derecho igual a α se tiene:

$$1 - EN(t)\overline{F}(x)(1 + o(1)) = \alpha \Rightarrow EN(t)\overline{F}(x)(1 + o(1)) = 1 - \alpha$$

$$\Rightarrow EN(t)\bar{F}(x) = \frac{1-\alpha}{(1+o(1))} \Rightarrow \bar{F}(x) = \frac{1-\alpha}{EN(t)(1+o(1))} \Rightarrow 1 - F(x) = \frac{1-\alpha}{EN(t)(1+o(1))}$$

Despejando $F(x)$ y tomando límite cuando $x \rightarrow \infty$ ($o \alpha \rightarrow 1$), se tiene la solución asintótica $F(x) = 1 - \frac{1-\alpha}{EN(t)}(1 + o(1))$, $x \rightarrow \infty$.

Por lo tanto:

$$x = G_t^{\leftarrow}(\alpha) = F^{\leftarrow}\left(1 - \frac{1-\alpha}{EN(t)}(1 + o(1))\right), \alpha \rightarrow 1. \quad \square$$

Algunas conclusiones fundamentales de esta aproximación son las siguientes:

- a. Para altos niveles de confianza el OpVaR únicamente depende de la cola, y no del cuerpo de la distribución de severidad. Por lo tanto, si el objetivo es calcular el VaR operacional, no es necesario modelar toda la función de distribución F .
- b. La distribución de frecuencias solo interviene en la expresión (2.22) con su valor esperado. Por lo cual, para aplicar este modelo sería suficiente estimar la media muestral de las frecuencias. Esto implica que la sobre-dispersión de un modelo como la distribución binomial negativa, asintóticamente no tendría impacto sobre el OpVaR.
- c. Como la carga de capital está basada en un cuantil muy alto de la distribución de pérdidas agregadas G_t , es natural estimar el OpVaR mediante el comportamiento asintótico de la cola y estimación cuantil. En lugar de considerar la distribución completa, es suficiente concentrarnos sobre la cola derecha $P(S(t)) > x$ para un valor grande de x .
- d. El resultado en (2.22) se cumple para una clase muy general de modelos LDA, y muestra que para obtener una aproximación de primer orden para el OpVaR, mediante un modelo LDA específico, es suficiente combinar (2.22) con la cola de la distribución de severidad subexponencial F .

Aplicando la ecuación (2.22) a las funciones de distribución de la tabla 2.3, se obtienen inmediatamente las soluciones en forma cerrada para un $OpVaR_t(\alpha)$ asintótico ($\alpha \rightarrow 1$), para las correspondientes distribuciones de severidad, como se indica en la Tabla 2.4.

Tabla 2.4. Aproximaciones asintóticas para el $OpVaR_t(\alpha)$, para las distribuciones de severidad más usuales.

DISTRIBUCIÓN	$OpVaR_t(\alpha)$
Lognormal	$\exp\left[\mu - \sigma\Phi^{-1}\left(\frac{1-\alpha}{EN(t)}\right)\right]$
Weibull	$\theta \left[\ln\left(\frac{EN(t)}{1-\alpha}\right) \right]^{1/\tau}$
Pareto	$\theta \left[\left(\frac{EN(t)}{1-\alpha}\right)^{1/a} - 1 \right]$

A partir de múltiples experimentaciones simuladas, Carrillo y Suárez (2006), concluyen que la fórmula de Böcker y Klüppelberg tiene un mejor desempeño cuando el número de eventos es pequeño, y el cuantil requerido de la distribución de pérdidas agregadas está adecuadamente explicado por la suma de un pequeño número de pérdidas, entre las cuales se pueden tener pérdidas muy altas; es decir, cuando se cumple la subexponencialidad. Por el contrario, cuando las pérdidas agregadas son grandes como resultado de sumar muchas pérdidas pequeñas, esta aproximación no será adecuada. Como en situaciones reales se presentan mixturas de ambas situaciones, el problema de encontrar una distribución de severidad que describa fielmente los datos empíricos de pérdidas es determinante, aunque complejo; sin embargo, la tarea de parametrizar distribuciones apropiadas para la frecuencia y la severidad, es decisiva para cualquier modelo de medición avanzada.

CAPÍTULO 3. APLICACIÓN

3.1 ANÁLISIS DE LOS DATOS

Las bases de datos relativas a riesgo operacional son escasas, y las entidades son muy reservadas para permitir acceso a ellas. Por esa razón, para elaborar una muestra de datos de pérdidas se recurrió a un grupo de expertos de una entidad financiera local, que proporcionaron estimativos de dos variables de interés: frecuencia y severidad. Los datos corresponden a un evento particular de riesgo que denominamos Fraude Total (Interno y Externo unificados), en una línea específica de negocios, que denominamos Banca Comercial. Aunque, el horizonte de tiempo para el VaR operacional recomendado por Basilea II es de un año, dada la poca disponibilidad de bases de datos, es necesario aclarar que los datos suministrados son mensuales, y así son incorporados e interpretados en las aplicaciones y en el contexto de los modelos, sin hacer ningún escalamiento.

Para la variable frecuencia se obtuvieron datos desde enero del 2003 hasta diciembre del 2007. Las frecuencias corresponden al número de veces que ocurrió el evento durante el mes. La severidad es pérdida por ocurrencia del evento (expresada en miles de pesos, y con un umbral mínimo de recolección de cero).

En la Tabla 3.1, se tiene un resumen de las estadísticas relevantes de la base de datos:

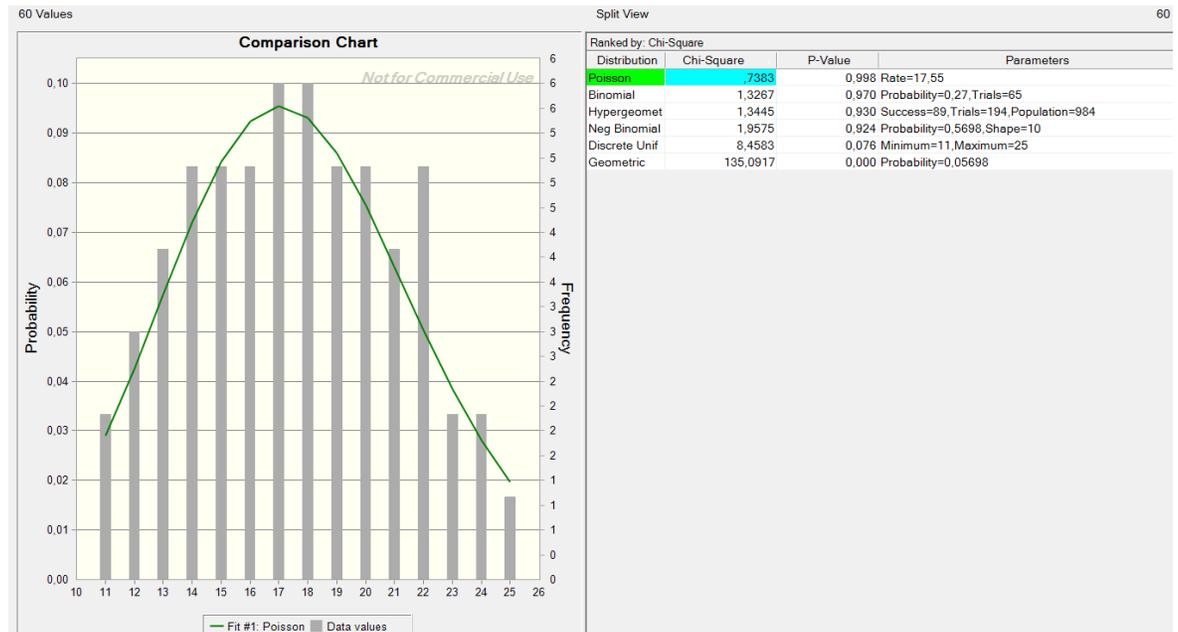
Tabla 3.1. Estadísticas relevantes de los datos.

	FRECUENCIA (Número de ocurrencias/mes)	SEVERIDAD (Pérdida/ocurrencia (miles de \$))
Media	18,93333333	2844,86667
Mediana	19	1675
Moda	18	1900
Desviación estándar	4,55044359	7169,59295
Varianza de la muestra	12,6056497	51403063.1
Curtosis	-0,1149924	2,36519017
Coficiente de asimetría	0,15308704	1,6728824
Rango	15	13784
Mínimo	12	186
Máximo	27	13970
Suma	1136	170692
Cuenta	60	60

Para la aplicación de los modelos y algoritmos se deben utilizar los parámetros de las distribuciones que mejor se ajusten a los datos reales. Por lo tanto, se procedió a hacer los respectivos ajustes utilizando el software Crystal Ball.

Para los datos de frecuencia se utilizó el software mediante el proceso de autoselección de distribuciones discretas, que fueron ranqueadas según la prueba Chi-Cuadrado, como se indica en la tabla 3.2.

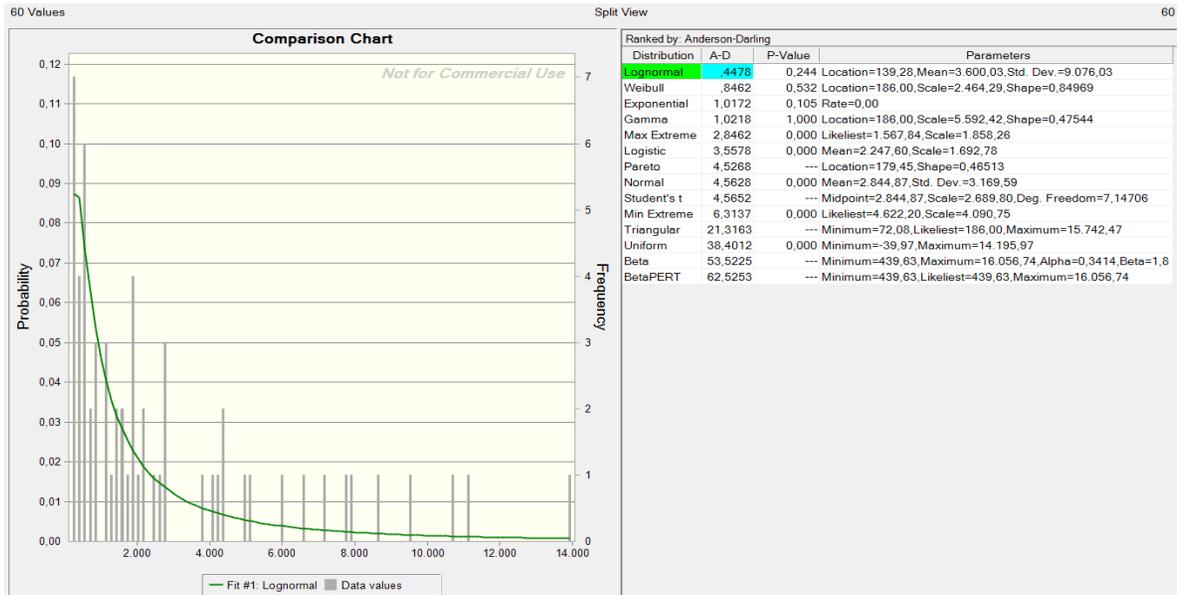
Tabla 3.2. Ajuste de la Distribución de Frecuencias.



Indicando que la distribución de Poisson con $\lambda = 17.55$ es la que mejor se ajusta a los datos de frecuencias. La segunda opción es una binomial con $n = 65$ y $p = 0.27$.

En forma análoga se procedió con los datos de severidad, caso en el cual las distribuciones fueron jerarquizadas como se indica en la Tabla 3.3.

Tabla 3.3. Ajuste de la Distribución de Severidad



Indicando como mejor ajuste una distribución lognormal con media 3600 y desviación estándar 9076. Estos valores corresponden a unos parámetros dados por $\mu = 7.19$ y $\sigma = 1.42$.

Aunque se han implementado los algoritmos para modelar las pérdidas con diversas combinaciones posibles de frecuencia y severidad, para aplicar los algoritmos en este caso, se asume que la distribución de frecuencias es Poisson con $\lambda = 17.55$, y que la distribución de severidad es lognormal con $\mu = 7.19$ y $\sigma = 1.42$.

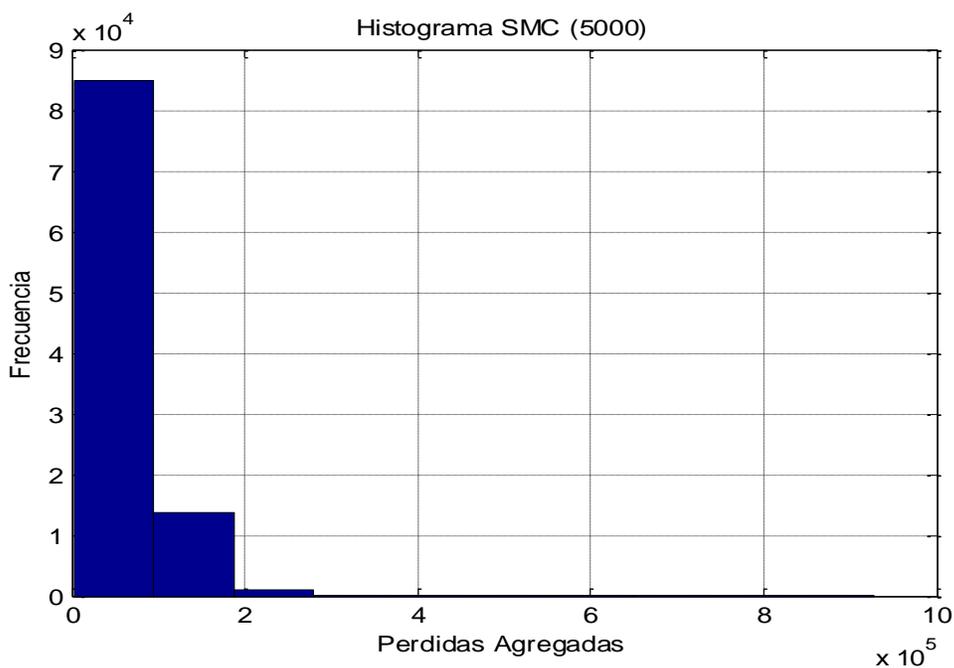
3.2 APLICACIÓN DE SIMULACIÓN MONTECARLO

Aplicando el algoritmo con diferentes números de iteraciones (5000,10000, 100000) y para efectos de comparación, con diferentes niveles de confianza (95%, 99% y 99.9%), se obtuvieron los siguientes resultados:

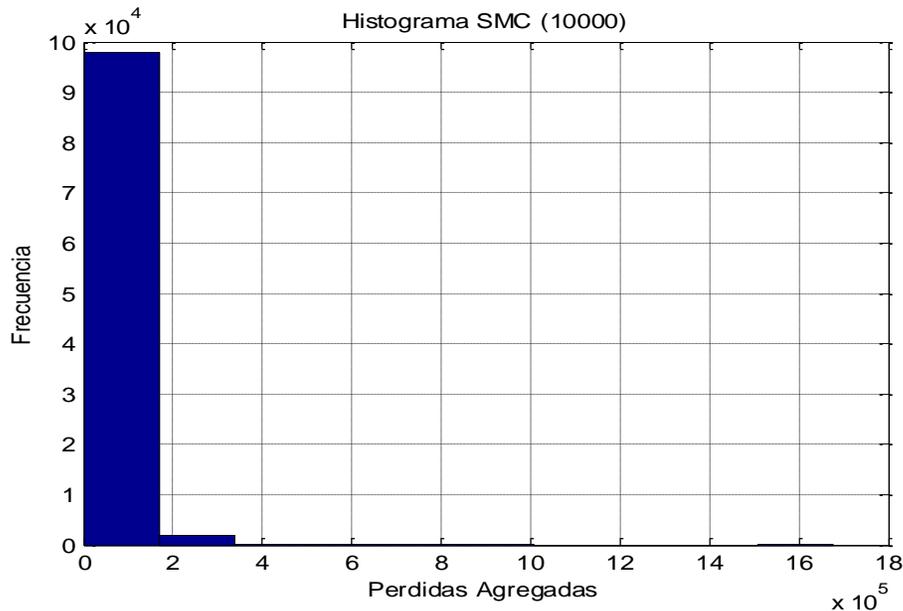
Número iteraciones	5000	10000	100000
MONTECARLO			
95%	137973.9	135520.6	134125.3
99%	218983.6	214036.9	206496.8
99.9%(OPVaR)	360935.4	375486.7	387681.4

Los histogramas obtenidos para las pérdidas agregadas, son los siguientes:

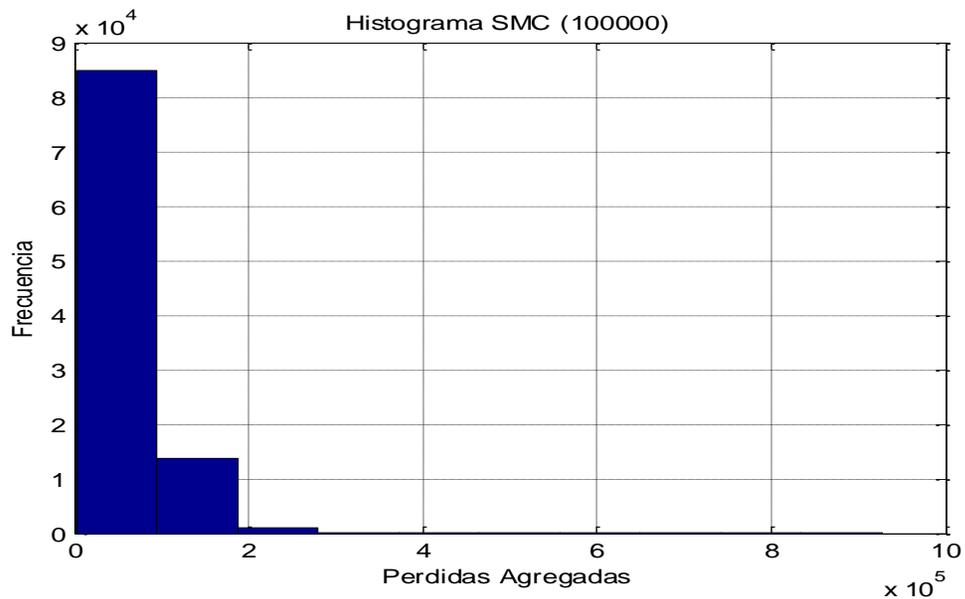
- SMC con 5000 iteraciones



- **SMC con 10000 iteraciones**



- **SMC con 100000 iteraciones**



Con 100000 iteraciones los resultados indican, por ejemplo, que el OpVaR estimado, con un nivel de confianza del 99.9%, es 387.641.4. En el contexto de la aplicación esto significaría que: Con un nivel de confianza del 99.9%, se estima que las pérdidas operacionales por mes, ocasionadas por fraude total en la Banca Comercial, no serán superiores a \$ 387.681.400.

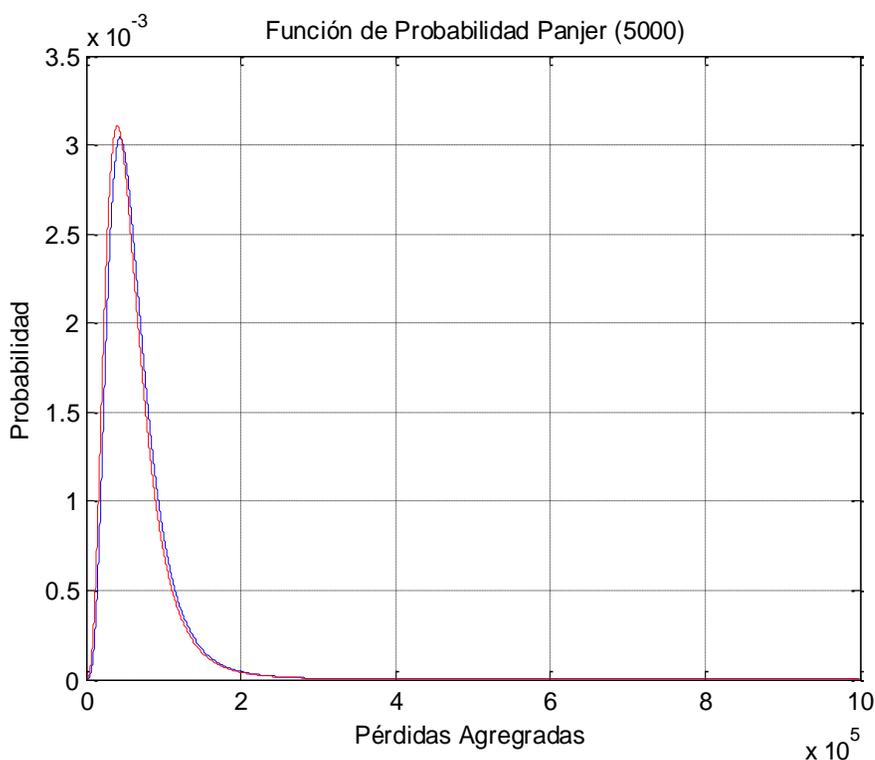
3.3 APLICACIÓN DEL ALGORITMO DE RECURSIÓN DE PANJER

Aplicando el algoritmo con diferentes números de iteraciones (5000, 10000, 100000) y para efectos de comparación, con diferentes niveles de confianza (95%, 99% y 99.9%), se obtuvieron los siguientes resultados:

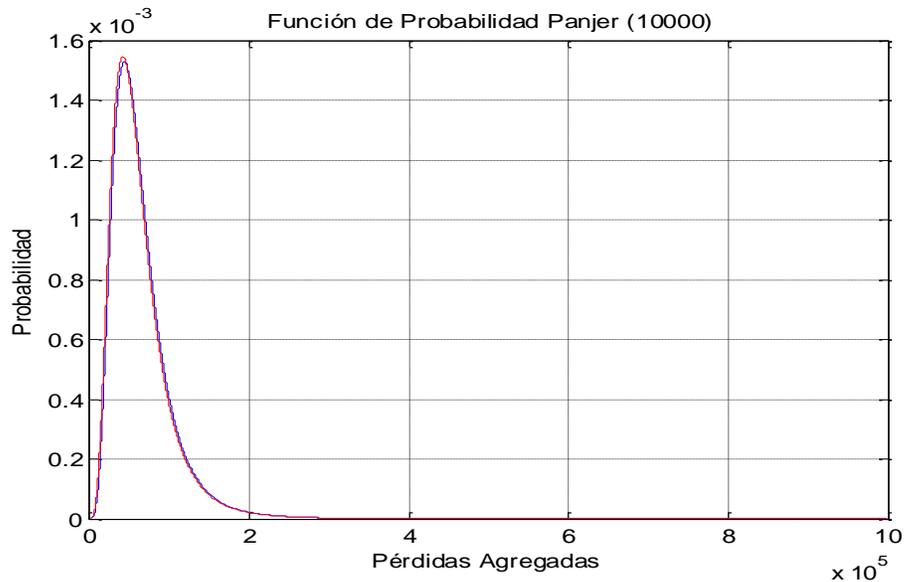
PANJER	5000	10000	100000
95%	140600.0	137500.0	134770.0
99%	215200.0	212100.0	209380.0
99.9%(OPVaR)	394800.0	391800.0	389160.0

Las gráficas obtenidas para las funciones de probabilidad de las pérdidas agregadas son respectivamente, como se muestran a continuación:

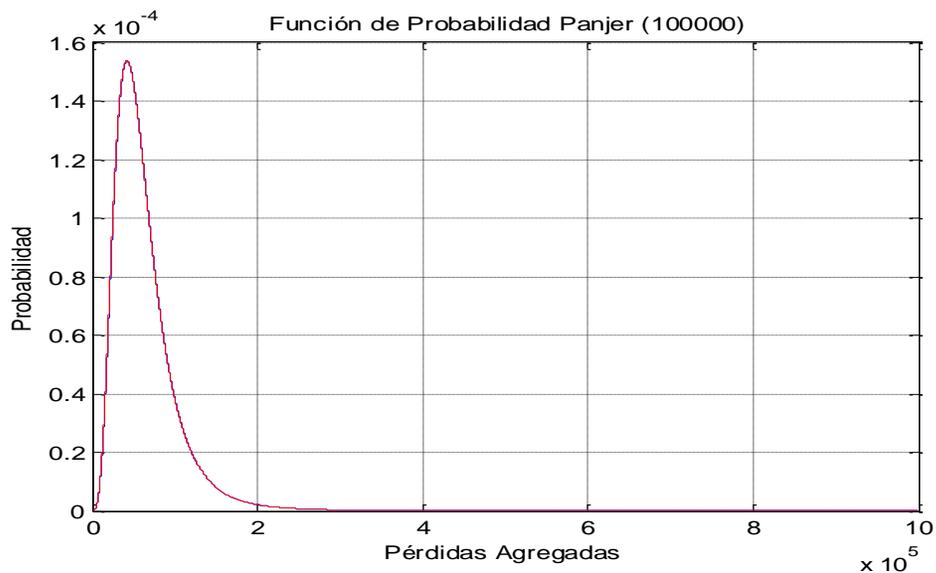
- **Panjer con 5000 iteraciones:**



- **Panjer con 10000 iteraciones:**



- **Panjer con 100000 iteraciones:**



Con el algoritmo de Panjer, con 100000 iteraciones, la interpretación en el contexto de la aplicación sería: Con un nivel de confianza del 99.9%, se estima que las pérdidas operacionales por mes, ocasionadas por fraude total en la Banca Comercial, no serán superiores a \$ 389.160.000.

3.4 APLICACIÓN DE APROXIMACIÓN ANALÍTICA DE BÖCKER Y KLÜPPELBERG

Aplicando la Aproximación Analítica de Böcker y Klüppelberg, correspondiente a la combinación de distribuciones Poisson-Lognormal, con los parámetros $\lambda = 17.55$ para la Poisson y con $\mu = 7.19$ y $\sigma = 1.42$ para la lognormal, y para los niveles de confianza considerados tiene:

Nivel de Confianza(α)	$\phi^{-1}\left(\frac{1-\alpha}{EN}\right)$	$OpVaR = \exp\left(\mu - \sigma\phi^{-1}\left(\frac{1-\alpha}{EN}\right)\right)$
95%	-2,76467201	67227,2703
99%	-3,25358129	134603,061
99.9%=OpVaR	-3,85876787	317886,722

Para visualizar conjuntamente todos los resultados, se presentan unificados en la tabla3.4.

Tabla 3.4. Resumen de Resultados

Número de Iteraciones	5000	10000	100000
MONTECARLO			
95%	137973.9	135520.6	134125.3
99%	218983.6	214036.9	206496.8
99.9%=OpVaR	360935.4	375486.7	387681.4
PANJER			
95%	140600.0	137500.0	134770.0

99%	215200.0	212100.0	209380.0
99.9%=OpVaR	394800.0	391800.0	389160.0
BÖCKER Y KLÜPPELBERG*			
95%	67227.3	67227.3	67227.3
99%	134603.1	134603.1	134603.1
99.9%=OpVaR	317886.7	317886.7	317886.7

*En la aproximación analítica el número de iteraciones no interviene.

Tanto en la simulación Montecarlo como en el algoritmo de Panjer, se observa que a medida que aumenta el número de iteraciones, el estimativo de las pérdidas para niveles de confianza del 95% y 99% está disminuyendo.

Con respecto a los estimativos de pérdidas para niveles de confianza del 99.9%, es decir, para los OpVaR, se encuentra que la simulación Montecarlo genera estimativos crecientes al aumentar el número de iteraciones.

Por el contrario, el algoritmo de Panjer está generando estimativos decrecientes para las pérdidas agregadas, a medida que crece el número de iteraciones, a un nivel de confianza del 99.9%.

Combinando las dos últimas conclusiones se podría afirmar que la simulación Montecarlo genera estimativos que convergen por debajo hacia el verdadero VaR operacional, y el algoritmo de Panjer genera estimativos que convergen por encima hacia el verdadero VaR operacional.

Lo anterior muestra claramente la conveniencia de que los departamentos de gestión del riesgo en las entidades financieras operen constantemente con diversos modelos de riesgo operacional para tener una visión más acertada de las verdaderas pérdidas potenciales.

Observando la convergencia de los resultados generados por Simulación Montecarlo y con el Algoritmo de Panjer, para los diferentes niveles de confianza,

y comparando con los resultados obtenidos por la fórmula de Böcker y Klüppelberg, se puede afirmar, en esta aplicación específica, que la Aproximación Analítica de Böcker y Klüppelberg subestima las pérdidas operacionales totales. Esto puede ocurrir, como afirman los mismos autores, debido a que el OpVaR para altos niveles de confianza, depende dramáticamente de la cola y no del cuerpo de la distribución, y la modelación en este caso está enfatizada sobre el cuerpo de los datos, sin considerar un análisis específico para valores extremos. Así mismo, como muestran Carrillo y Suárez (2004), cuando las pérdidas agregadas llegan a ser relativamente grandes como consecuencia de sumar muchas pérdidas pequeñas, como ocurre en la aplicación, la aproximación de Böcker y Klüppelberg, no genera estimativos adecuados.

CONCLUSIONES

En los últimos años, fundamentalmente desde el surgimiento del Nuevo Acuerdo de Basilea, que planteó el requerimiento de incorporar el riesgo operacional en la relación de solvencia, han surgido diversas alternativas metodológicas para su cuantificación.

La gestión integral del riesgo financiero, y en particular el riesgo operacional, se ha convertido en los últimos años en un gran reto para los investigadores y operadores en finanzas, ante los conocidos grandes desastres financieros de muchas entidades, mucha parte de ellos atribuidos a causas operacionales.

No solo la emergente regulación internacional, y la inminente normatividad nacional, que sólo está en la etapa de identificación, sino también y fundamentalmente, el propósito de mejorar continuamente el proceso de toma de decisiones en las instituciones, han motivado en los últimos años el interés por la investigación en el campo de riesgo financiero y, específicamente en la identificación, valoración, y gestión del riesgo operacional.

Uno de los problemas cruciales para la cuantificación del riesgo operacional ha sido, indudablemente, la escasez de datos confiables, pero ante la convicción creciente de la necesidad de cuantificar ese riesgo, las entidades han iniciado serios procesos de recolección de datos, y se considera que con el transcurrir del tiempo será un proceso superado, y los modelos de medición avanzada AMA, entre los cuales se considera el LDA, tomarán plena vigencia y aplicación.

Entretanto, sigue siendo un gran reto la formulación de un modelo con una complejidad manejable, y con las características de completitud, exactitud, y satisfacción de los estándares generales, cualitativos y cuantitativos planteados por el nuevo acuerdo de Basilea. Eso puede explicar el porqué incluso hasta el año 2005 grandes entidades financieras en el ámbito internacional estaban manejando el método estándar de Basilea II, para cuantificar su riesgo operacional, a pesar de que generaba excesivas cargas de capital.

En este trabajo se han presentado formalmente, junto con siete algoritmos asociados, tres de las metodologías de medición avanzada aplicables para cuantificar el riesgo operacional. Como ninguna de ellas parece ser consistentemente mejor que las otras, es necesario modelar adecuadamente tanto la frecuencia como la severidad, teniendo en cuenta los requerimientos de Basilea, y a partir de ellas utilizar varias metodologías, si se quiere tener una visión más acertada de las pérdidas potenciales, generadas por riesgo operacional.

APÉNDICE

CÓDIGOS DE MATLAB

1. SIMULACIÓN MONTECARLO (POISSON-LOGNORMAL)

```
close;
clear all;
clc;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ENTRADAS

lambda = 17.55;
mu = 7.19;
sigma = 1.42;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PRIMER PASO
niteraciones=5000;
for j1=1:niteraciones
    datop = poissrnd(lambda,1,1);
    L = lognrnd(mu,sigma,[datop 1]); % genera tantas perdidas como
indica la Poisson
    S=sum(L);
    DATOSx(j1,1)=S;
end
%hist(DATOSx);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PASO FINAL
Y = prctile(DATOSx,[95 99 99.9]);
```

2. ALGORITMO DE PANJER (POISSON-LOGNORMAL)

```
close;
clear all;
clc;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ENTRADAS
n = 5e3;
xmax = 1000000;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PRIMER PASO
a=0;
```

```

lambda=17.55;
b=lambda;
mu=7.19;
sigma=1.42;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% SEGUNDO PASO
dx=xmax/n;
x=dx*(0:n-1);
F=@(x) logncdf(x,mu,sigma);
F0=F(0);
dF=diff([F(x),1]);
datosD=[F0,dF(1:end-1)];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% TERCER PASO
E=1-sum(datosD);
datosDcorr=(datosD/sum(datosD))*E;
datosD=datosD+datosDcorr;
sum(datosD);
p0 = exp(-b);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CUARTO PASO
g=panjer2009poisson(lambda,datosD,n,mu,sigma,p0,a,b);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% QUINTO PASO
[valor]=sumak(length(xp),[0.95,0.99,0.999],xp,g)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PASO FINAL
xp=x;
plot(g,'b-');

```

3. ALGORITMO DE PANJER (BINOMIAL-LOGNORMAL)

```

close;
clear all;
clc;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ENTRADAS
n = 1e5;
xmax = 1000000;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PRIMER PASO
lambda=17.55;
mu=7.19;
sigma=1.42;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% SEGUNDO PASO
dx=xmax/n;
x=dx*(0:n-1);

% a<0,
% binomial: p=a/(a-1), n=-b/a-1
p0 = (1-a/(a-1))^( -b/a-1);
F=@(x)binocdf(X,N,P);
%F=@(x)logncdf(x,mu,sigma);

F0=F(0);
dF=diff([F(x),1]);
datosD=[F0,dF(1:end-1)];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% TERCER PASO
E=1-sum(datosD);
datosDcorr=(datosD/sum(datosD))*E;
datosD=datosD+datosDcorr;
sum(datosD);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CUARTO PASO
gb=panjer2009binomial(lambda,datosD,n,mu,sigma,p0,a,b);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% QUINTO PASO
[valor]=sumak(length(xp),[0.95,0.99,0.999],xp,gb)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PASO FINAL
xp=x;
plot(g,'b-')

```

4. ALGORITMO DE PANJER (BINOMIAL NEGATIVA-LOGNORMAL)

```

close;
clear all;
clc;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ENTRADAS
n = 1e5;
xmax = 1000000;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PRIMER PASO
a>0
a=
b=

```

```

lambda=17.55;
mu=7.19;
sigma=1.42;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% SEGUNDO PASO
dx=xmax/n;
x=dx*(0:n-1);
% a<0,
% binomial: p=a/(a-1), n=-b/a-1
p0 = (1-a)^(1+b/a);
p0 = (1-a/(a-1))^(-b/a-1);
F=@(x)wblcdf(X,A,B);
%F=@(x)logncdf(x,mu,sigma);
F0=F(0);
dF=diff([F(x),1]);
datosD=[F0,dF(1:end-1)];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% TERCER PASO
E=1-sum(datosD);
datosDcorr=(datosD/sum(datosD))*E;
datosD=datosD+datosDcorr;
sum(datosD);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CUARTO PASO
gb=panjer2009binomialnegativa(lambda,datosD,n,mu,sigma,p0,a,b);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% QUINTO PASO
[valor]=sumak(length(xp),[0.95,0.99,0.999],xp,gb)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PASO FINAL
xp=x;
plot(g,'b-')

```

5. ALGORITMO DE PANJER (POISSON-WEIBULL)

```

close;
clear all;
clc;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ENTRADAS
n = 5e3;
xmax = 1000000;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PRIMER PASO
a=0;
lambda=17.55;
b=lambda;

```

```

mu=7.19;
sigma=1.42;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% SEGUNDO PASO
dx=xmax/n;
x=dx*(0:n-1);
%P = wblcdf(X,A,B)
F=@(x)wblcdf(X,A,B);
F0=F(0);
dF=diff([F(x),1]);
datosD=[F0,dF(1:end-1)];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% TERCER PASO
E=1-sum(datosD);
datosDcorr=(datosD/sum(datosD))*E;
datosD=datosD+datosDcorr;
sum(datosD);
p0 = exp(-b);
datosD=[F0,dF(1:end-1)];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CUARTO PASO
g=panjer2009poisson(lambda,datosD,n,mu,sigma,p0,a,b);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% QUINTO PASO

[valor]=sumak(length(xp),[0.95,0.99,0.999],xp,g)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PASO FINAL
xp=x;
plot(g,'b-')

```

6. ALGORITMO DE PANJER (BINOMIAL-WEIBULL)

```

close;
clear all;
clc;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ENTRADAS
n = 1e5;
xmax = 1000000;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PRIMER PASO
lambda=17.55;
mu=7.19;
sigma=1.42;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% SEGUNDO PASO
dx=xmax/n;
x=dx*(0:n-1);

% a<0,
% binomial: p=a/(a-1), n=-b/a-1
p0 = (1-a/(a-1))^( -b/a-1);
%P = wblcdf(X,A,B)
F=@(x)wblcdf(X,A,B)
%F=@(x)logncdf(x,mu,sigma);
F0=F(0);
dF=diff([F(x),1]);
datosD=[F0,dF(1:end-1)];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% TERCER PASO
E=1-sum(datosD);
datosDcorr=(datosD/sum(datosD))*E;
datosD=datosD+datosDcorr;
sum(datosD);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CUARTO PASO
gb=panjer2009binomial(lambda,datosD,n,mu,sigma,p0,a,b);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% QUINTO PASO
[valor]=sumak(length(xp),[0.95,0.99,0.999],xp,gb)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PASO FINAL
xp=x;
plot(g,'b-')

```

7. ALGORITMO DE PANJER (BINOMIAL NEGATIVA-WEIBULL)

```

close;
clear all;
clc;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ENTRADAS
n = 1e5;
xmax = 1000000;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PRIMER PASO
a=
b=

```

```

lambda=17.55;
mu=7.19;
sigma=1.42;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% SEGUNDO PASO
dx=xmax/n;
x=dx*(0:n-1);
% a<0,
% binomial: p=a/(a-1), n=-b/a-1
p0 = (1-a/(a-1))^( -b/a-1);
%P = wblcdf(X,A,B)
F=@(x)wblcdf(X,A,B);
%F=@(x)logncdf(x,mu,sigma);
F0=F(0);
dF=diff([F(x),1]);
datosD=[F0,dF(1:end-1)];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% TERCER PASO
E=1-sum(datosD);
datosDcorr=(datosD/sum(datosD))*E;
datosD=datosD+datosDcorr;
sum(datosD);
p0 = (1-a)^(1+b/a);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CUARTO PASO
gb=panjer2009binomialnegativa(lambda,datosD,n,mu,sigma,p0,a,b);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% QUINTO PASO
[valor]=sumak(length(xp),[0.95,0.99,0.999],xp,gb)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% PASO FINAL
xp=x;
plot(g,'b-')

```

8. SUB-RUTINAS DE MATLAB

```

function g=panjer2009poissonW(lambda,f,n,mu,sigma,p0,a,b)

suma(1)=0;
g(1)=p0*exp(b*f(1));
for j=1:n
    for i=2:j+1
        suma(i)=suma(i-1)+(i-1)*f(i-1)*g(j-i+2);
        h(i)=(lambda/j)*suma(i);
    end
    g(j+1)=h(j+1);
end

```

```
function g=panjer2009binomialnegativaW(lambda,f,n,mu,sigma,p0,a,b)

suma(1)=0;

g(1)=p0/(1-a*f(1))^(1+b/a);
for j=1:n
    for i=2:j+1
        suma(i)=suma(i-1)+(i-1)*f(i-1)*g(j-i+2);
        h(i)=(lambda/j)*suma(i);
    end
    g(j+1)=h(j+1);
end
```

```
function g=panjer2009binomialnegativa(lambda,f,n,mu,sigma,p0,a,b)

suma(1)=0;

g(1)=p0/(1-a*f(1))^(1+b/a);
for j=1:n
    for i=2:j+1
        suma(i)=suma(i-1)+(i-1)*f(i-1)*g(j-i+2);
        h(i)=(lambda/j)*suma(i);
    end
    g(j+1)=h(j+1);
end
```

```
function g=panjer2009binomial(lambda,f,n,mu,sigma,p0,a,b)

suma(1)=0;

g(1)=p0/(1-a*f(1))^(1+b/a);
for j=1:n
    for i=2:j+1
        suma(i)=suma(i-1)+(i-1)*f(i-1)*g(j-i+2);
        h(i)=(lambda/j)*suma(i);
    end
    g(j+1)=h(j+1);
end
```

```
function g=panjer2009poisson(lambda,f,n,mu,sigma,p0,a,b)
suma(1)=0;
g(1)=p0*exp(b*f(1));
for j=1:n
    for i=2:j+1
        suma(i)=suma(i-1)+(i-1)*f(i-1)*g(j-i+2);
        h(i)=(lambda/j)*suma(i);
    end
    g(j+1)=h(j+1);
end
```

```
function [valor]=sumak(n,casosn, xp, gl)

for j=1:length(casosn)
    suma1=0;
    k=1;
    while suma1<=casosn(j) & k<=length(xp)
        suma1=gl(k)+suma1;
        valor(j)=xp(k);
        k=k+1;
    end

end

suma_media=0;
for t=1:length(xp)
    suma_media=xp(t)*gl(1)+suma_media;
end
media=suma_media;
valor=[valor media];
```

BIBLIOGRAFÍA

1. Adelson, R.M. (1966) Compound Poisson Distributions, *Operations Research Quarterly*, **17**, 73-75.
2. Akkizidis, I. y Bouchereau, V. (2006). *Guide to Optimal Operational Risk & Basel II*. Boca Raton: Auerbach Publications.
3. Aue, F., y Kalkbrener, M. (2007). *LDA at Work*. (Documento de trabajo).
4. Böcker, K. y Klüppelberg, C. (2005) Operational VaR: a closed-form solution. *Risk Magazine*, December, 90-93.
5. Böcker, K. (2006). *Closed-form approximation for OpVaR when high frequency losses are parameterized by a generalized Pareto distribution (GPD)*. Munich: Hypo Vereinsbank.
6. Bühlmann, H. (1970). *Mathematical Methods in Risk Theory*. Springer-Verlag.
7. Carrillo, S. y Alberto Suárez. (2004). Medición Efectiva del Riesgo Operacional. En: *Estabilidad Financiera*, No. 11, 61-90.
8. Comité de Supervisión Bancaria de Basilea. (2001). *Operational Risk*, Basilea: BIS.
9. Comité de Supervisión Bancaria de Basilea. (2003). *Supervisory Guidance on Operational Risk Advanced Measurement Approaches for Regulatory Capital*, Basilea: BIS
10. Comité de Supervisión Bancaria de Basilea. (2006). *Convergencia Internacional de Medidas y Normas de Capital*. Basilea: BIS.
11. Cruz, M. (2002). *Modeling, Measuring and Hedging Operational Risk*. New York: John Wiley & Sons.
12. Chernobai, A. y Rachev, S. (2007). *Applying Robust Methods to Operational Risk Modeling*. (Documento de trabajo).
13. De Fontnouvelle, P., Jordan, J. y Rosengren, E. (2005). *Implication of Alternative Operational Risk Modeling Techniques*. (Documento de trabajo).

14. Degen, M., Embrechts, P. y Lambrigger D.(2007). *The quantitative modeling of operational risk: between g-and-h and EVT*. (Documento de trabajo).
15. Dutta, K., Perry, J. (2006) *A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital*. (Documento de trabajo).
16. Embrechts, P., Klüppelberg, C. and Mikosch, T.(1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlín.
17. Fera, J. M. y Jiménez, E. (2005). *El OpVaR como medida del Riesgo Operacional*. (Documento de trabajo).
18. Frachot, A., Georges, P. y Roncalli, T. (2001). *Loss Distribution Approach for Operational Risk*. (Documento de trabajo, Credit Lyonnais).
19. Frachot, A., Moudoulaud, O., Olivier y Roncalli, T. (2003). *Loss Distribution Approach in Practice*. Paris: Credit Lyonnais.
20. Frachot, A., Roncalli, T. y Salomon, E. (2004). *The Correlation Problem in Operational Risk*. (Documento de trabajo).
21. Franco, L. C., Murillo J. G. (2006). Riesgo Operacional: Reto actual de las entidades financieras. *Revista Ingenierías Universidad de Medellin*, 5(9), 97-110.
22. González, M. (2004). *Análisis del Nuevo Acuerdo de Capitales de Basilea (BIS-II): Pyme-Risk, Country-Risk y Operational-Risk*. (Documento de trabajo).
23. Hess, K. Th., A. Liewald, y K. D. Schmidt (2002). *An extension of Panjer's recursion*, *Astin Bull.* **32**, no. 2, 283–297.
24. Jobst, A. (2007). Operational Risk – The Sting is Still in the Tail But the Poison Depends on the Dose. *Journal of Operational Risk* Vol. 2, No. 2.
25. Klugman, S., Panjer, H. and Willmot, G. (2004). *Loss Models - From Data to Decisions*. Wiley, Hoboken, New Jersey.
26. Lee, C. M. (2001). *Measuring and Managing Operational Risk in Financial Institutions*. Singapore: John Wiley & Sons.

27. Nešlehová, J., Embrechts, P. y Chavez-Demoulin, V. (2005). *Quantitative Models for Operational Risk: Extremes, Dependence and Aggregation*. (Documento de trabajo).
28. Panjer, H. H. (1980) The Aggregate Claims distribution and Stop-Loss Reinsurance, *Trans of the Society of Actuaries*, XXXII.
29. Panjer, Harry H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin (International Actuarial Association)* 12 (1): 22–26.
30. Panjer, H. H. y S. Wang. (1993). *On the stability of recursive formulas*, *Astin Bull.* **23**, no. 2, 227–258.
31. Peters, G. W., P. V. Shevchenko and M. V. Wüthrich (2008). Dynamic operational risk: Modelling dependence and combining different sources of information. Preprint, CSIRO call number CMIS 2741.
32. Peters, G. W. P. V. Shevchenko and M. V. Wüthrich (2009). Model uncertainty in claims reserving within Tweedie's compound Poisson models. To appear in *ASTIN Bulletin* 39, CSIRO call number CMIS 2609.
33. Shevchenko, P. y Donnelly, J. (2005). *Validation of the Operational Risk LDA Model for Capital Allocation and AMA Accreditation under Basel II*. CMIS Confidential report prepared for Basel II programme ANZ bank.
34. Shevchenko, P. y G. Temnov (2008). Modelling operational risk data reported above varying threshold. Preprint, CSIRO call number CMIS 2828.
35. Shevchenko, P.V. (2008). Implementing Basel II Loss Distribution Approach for Operational Risk. Preprint, CSIRO call number CMIS 2794.
36. Sundt, B. y W. S. Jewell. (1981). *Further results on recursive evaluation of compound distributions*, *Astin Bull.* **12**, no. 1, 27–39.
37. Superintendencia Financiera de Colombia. (2007). *Circular Básica Contable y Financiera, Capítulo XXI*. Reglas relativas a la administración del riesgo operativo.

38. Tripp, M. H., Bradley, H. L., Devitt, R., Orros, G. C., Overton, G. L., Pryor, L. M. y Shaw, R. A. (2004). *Quantifying operational Risk in General Insurance Companies*. (Documento de trabajo).
39. Willmot, G. E. (1986). *Mixed compound Poisson distributions*, *Astin Bull.* **16**, no. S, 59–76.
40. X. Luo and P.V. Shevchenko (2008). Computing tails of compound loss distributions using direct numerical integration, Preprint, CSIRO call number CMIS 2688.